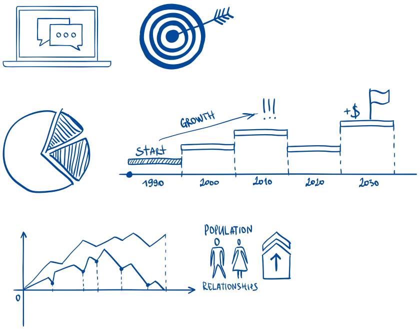
DATA RULEZZZ!



DATA-DRIVEN DECISION-MAKING IN MEDICAL EDUCATION AND HEALTHCARE

18 case studies based on good practice

Martin Komenda et al.

MASARYK
UNIVERSITY
PRESS

MUNI
MED

ÚZIS ČR

MUNI
MED

ÚZIS ČR

MASARYK
UNIVERSITY
PRESS

DATA RULEZZZ!

DATA-DRIVEN DECISION-MAKING IN MEDICAL EDUCATION AND HEALTHCARE

18 case studies based on good practice

Martin Komenda et al.

Masaryk University
Institute of Health Information and Statistics of the Czech Republic
Brno 2023

MOTTO OF THE BOOK

"Without data, you are just another person with an opinion."

William Edwards Deming (1900–1993)
An American engineer, statistician, professor, author,
lecturer and management consultant

"All good work is done the way ants do things: Little by little."

Lafcadio Hearn (1850–1904)
An Irish-Greek-Japanese writer,
translator and teacher

REVIEWERS:

MEDICAL AND HEALTHCARE EDUCATION
Assoc. Prof. Josef Hanuš, PhD
Charles University, Faculty of Medicine in Hradec Králové, Hradec Králové, Czech Republic
Prof. Elena Zaitseva, PhD
University of Zilina, Faculty of Management Science and Informatics, Zilina, Slovakia

HEALTH INFORMATION AND STATISTICS
Assoc. Prof. Lenka Lhotská, PhD
Czech Technical University in Prague, Department of Cognitive Systems and Neurosciences, Prague, Czech Republic
Czech Institute of Informatics, Robotics and Cybernetics, Prague, Czech Republic
Assoc. Prof. Martin Nečaský, PhD
Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic

ISBN 978-80-280-0392-0 (Masaryk University)
ISBN 978-80-7472-198-4 (Institute of Health Information and Statistics of the Czech Republic)

MAIN EDITOR

Martin Komenda

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Masaryk University, Faculty of Medicine, Department of Simulation Medicine, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Martin Komenda holds a master's and PhD degree in Informatics and a doctoral degree in Applied Informatics from Masaryk University. He has acquired a solid academic background in computer science as a senior university teacher and researcher, especially in the following domains: medical informatics, technology enhanced learning, educational data / text mining, data processing and visualisation, web application development, information and health literacy, health information systems, open data, supervision of bachelor and diploma theses. He is the senior leader of the development and research team at the joint workplace of the Institute of Biostatistics and Analyses at the Faculty of Medicine of Masaryk University and the Institute of Health Information and Statistics of the Czech Republic, coordinator of healthcare open data, member of boards and councils at Masaryk University (Working group for data and visual presentation strategy, Working group for simulation in medicine), member of the Coordinating Council of the MEFANET network (an educational network of all Czech and Slovak medical faculties), member of the Coordinating Council of the National Health Information Portal, member of the European Open Science Cloud Working Group (domain of sensitive data) and member of the MedBiquitous Curriculum Inventory board. He has also been involved in the management of dozens of projects at national and international levels.

CO-AUTHORS

Albreht Tit

National Institute of Public Health of Slovenia, Ljubljana, Slovenia

Section C:

|15| Innovations in patient-centred cancer care: Online benchmarking tool

Antol Matej

Masaryk University, Institute of Computer Science, Brno, Czech Republic

Section A:

Public (health) sector and academia

Bamidis Panos

Aristotle University of Thessaloniki, School of Medicine, Lab of Medical Physics and Digital Innovation, Thessaloniki, Greece

Section B:

|05| Medical curriculum innovations using technological standards

Bareš Martin

St. Anne's University Hospital in Brno, First Clinic of Neurology, Brno, Czech Republic
Masaryk University, Faculty of Medicine, Brno, Czech Republic

Section B:

|01| Outcome-based curriculum development and overview using an innovative online platform

Bartůněk Vladimír

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|10| Data and analytical basis for a modern mental health care system in the Czech Republic

Barvík Daniel

University Hospital Brno, Department of Anaesthesiology and Intensive Care Medicine, Brno, Czech Republic
Masaryk University, Faculty of Medicine, Department of Simulation Medicine, Brno, Czech Republic

Section B:

|06| Processing, analysis, and visualisation of objective structured clinical examination data

Benáček Petr

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|09| Cancer screening programmes: Open data and visualisations as a support tool for monitoring and evaluation

|10| Data and analytical basis for a modern mental health care system in the Czech Republic

Benešová Klára

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

Blaha Milan

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|14| National health information portal: Text-based open data

Bulhart Vojtěch

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section B:

|06| Processing, analysis, and visualisation of objective structured clinical examination data

Section C:

|14| National health information portal: Text-based open data

Búřilová Petra

Masaryk University, Faculty of Medicine, Department of Health Sciences, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|13| Adverse Events Reporting System: Development of the platform for healthcare providers and lay public

Ciureanu Adrian

Grigore T. Popa University of Medicine and Pharmacy Iasi, Department of Medical Informatics, Biostatistics, Computer Science, Mathematics and Modelling Simulation, Iasi, Romania

Section B:

|07| Building Curriculum Infrastructure in Medical Education

Číhal Jaroslav

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|16| Interactive tool for crisis management in times of epidemic in the Czech Republic

Dobrovolná Julie

Masaryk University, Faculty of Medicine, Department of Pathophysiology, Brno, Czech Republic
Masaryk University, Faculty of Sports Studies, Department of Physical Activities and Health Sciences, Brno, Czech Republic

Section B:

|01| Outcome-based curriculum development and overview using an innovative online platform

Dolanová Dana

Masaryk University, Faculty of Medicine, Department of Health Sciences, Brno, Czech Republic
Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|13| Adverse Events Reporting System: Development of the platform for healthcare providers and lay public

Dušek Ladislav

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section A:

Health data opening

Section B:

|01| Outcome-based curriculum development and overview using an innovative online platform

Section C:

|09| Cancer screening programmes: Open data and visualisations as a support tool for monitoring and evaluation

|10| Data and analytical basis for a modern mental health care system in the Czech Republic

|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service

|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

|13| Adverse Events Reporting System: Development of the platform for healthcare providers and lay public

|14| National health information portal: Text-based open data

|15| Innovations in patient-centred cancer care: Online benchmarking tool

|16| Interactive tool for crisis management in times of epidemic in the Czech Republic

|17| Management tool for epidemic monitoring, analysis, and visualisation

|18| A pilot toolbox against health misinformation in the Czech Republic in the COVID-19 age

Dvořák Vladimír

Centre for Outpatient Gynaecology and Primary Care, Ltd., Brno, Czech Republic

Board of the National Register of Reproductive Health

Section C:

|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service

Eclerová Veronika

Masaryk University, Faculty of Science, Department of Mathematics and Statistics, Brno, Czech Republic

Section C:

|17| Management tool for epidemic monitoring, analysis, and visualisation

Gregor Jakub

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

|15| Innovations in patient-centred cancer care: Online benchmarking tool

|18| A pilot toolbox against health misinformation in the Czech Republic in the COVID-19 age

Hege Inga

University Augsburg, School of Medicine, Augsburg, Germany

Section B:

|07| Building Curriculum Infrastructure in Medical Education

Hejduk Karel

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

National Screening Centre, Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|09| Cancer screening programmes: Open data and visualisations as a support tool for monitoring and evaluation

Hohensinger Doris

Software Competence Center Hagenberg, Hagenberg, Austria

Section B:

|08| Visualisation of text-based data describing best practices in software development

Chloupková Renata

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

National Screening Centre, Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|09| Cancer screening programmes: Open data and visualisations as a support tool for monitoring and evaluation

Janků Petr

University Hospital Brno, Department of Gynaecology and Obstetrics, Brno, Czech Republic
Masaryk University, Faculty of Medicine, Brno, Czech Republic
Board of the National Register of Reproductive Health

Section C:

|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service

Jarkovský Jiří

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic
Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|10| Data and analytical basis for a modern mental health care system in the Czech Republic
|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service
|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

Jírová Jitka

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service

Kacerovský Marian

University Hospital Hradec Králové, Clinic of Obstetrics and Gynaecology, Hradec Králové, Czech Republic
Charles University, Faculty of Medicine in Hradec Králové, Hradec Králové, Czech Republic
Board of the National Registry of Reproductive Health

Section C:

|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service

Karolyi Matěj

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic
Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section B:

|01| Outcome-based curriculum development and overview using an innovative online platform
|03| User behaviour analysis and interactive visualisation in a curriculum management system

Kerberová Michaela

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section B:

|08| Visualisation of text-based data describing best practices in software development

Klimeš Daniel

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|16| Interactive tool for crisis management in times of epidemic in the Czech Republic

Kononowicz Andrzej

Jagiellonian University, Medical College, Department of Bioinformatics and Telemedicine, Kraków, Poland

Section B:

|07| Building curriculum infrastructure in medical education

Kosinová Martina

University Hospital Brno and Masaryk University, Faculty of Medicine, Department of Paediatric Anaesthesiology and Intensive Care Medicine, Brno, Czech Republic

Masaryk University, Faculty of Medicine, Department of Simulation Medicine, Brno, Czech Republic

Section B:

|06| Processing, analysis, and visualisation of objective structured clinical examination data

Kraus Andrea

Masaryk University, Faculty of Science, Department of Mathematics and Statistics, Brno, Czech Republic

Section C:

|17| Management tool for epidemic monitoring, analysis, and visualisation

Kraus David

Masaryk University, Faculty of Science, Department of Mathematics and Statistics, Brno, Czech Republic

Section C:

|17| Management tool for epidemic monitoring, analysis, and visualisation

Linhart Aleš

General University Hospital in Prague, Second Clinic of Internal Medicine – Clinic of Cardiology and Angiology, Prague, Czech Republic
Charles University, First Faculty of Medicine, Prague, Czech Republic

Section C:

|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

Ľubušký Marek

University Hospital Olomouc, Clinic of Obstetrics and Gynaecology, Olomouc, Czech Republic
Palacký University in Olomouc, Faculty of Medicine, Olomouc, Czech Republic
Board of the National Registry of Reproductive Health

Section C:

|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service

Macková Barbora

National Institute of Public Health, Prague, Czech Republic

Section C:

|14| National health information portal: Text-based open data

Macková Denisa

Masaryk University, Faculty of Medicine, Department of Health Sciences, Brno, Czech Republic
Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|13| Adverse Events Reporting System: Development of the platform for healthcare providers and lay public

Májek Ondřej

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic
National Screening Centre, Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|09| Cancer screening programmes: Open data and visualisations as a support tool for monitoring and evaluation
|15| Innovations in patient-centred cancer care: Online benchmarking tool
|17| Management tool for epidemic monitoring, analysis, and visualisation

Majerník Jaroslav

Pavol Jozef Šafárik University in Košice, Faculty of Medicine, Department of Medical Informatics and Simulator Medicine, Košice, Slovakia

Section B:

|07| Building Curriculum Infrastructure in Medical Education

Marek Jiří

Masaryk University, Institute of Computer Science, Brno, Czech Republic

Section A:

The big picture

Public (health) sector and academia

Mařík Radek

Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Telecommunications Engineering, Prague, Czech Republic

Section C:

|18| A pilot toolbox against health misinformation in the Czech Republic in the COVID-19 age

Matyska Luděk

Masaryk University, Institute of Computer Science, Brno, Czech Republic

Section A:

Public (health) sector and academia

Mayer Jiří

University Hospital Brno, Clinic of Internal Medicine, Haematology and Oncology, Brno, Czech Republic

Masaryk University, Faculty of Medicine, Brno, Czech Republic

Section B:

|01| Outcome-based curriculum development and overview using an innovative online platform

Melicharová Hana

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|10| Data and analytical basis for a modern mental health care system in the Czech Republic

Moravec Václav

Charles University, Faculty of Social Sciences, Institute of Communication Studies and Journalism, Prague, Czech Republic

Section C:

|18| A pilot toolbox against health misinformation in the Czech Republic in the COVID-19 age

Mužík Jan

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|13| Adverse Events Reporting System: Development of the platform for healthcare providers and lay public

Ngo Ondřej

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

National Screening Centre, Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|09| Cancer screening programmes: Open data and visualisations as a support tool for monitoring and evaluation

Pařenica Jiří

University Hospital Brno, Clinic of Internal Medicine and Cardiology, Brno, Czech Republic

Masaryk University, Faculty of Medicine, Brno, Czech Republic

Section C:

|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

Pařízek Antonín

First Faculty of Medicine and General University Hospital in Prague, Department of Gynaecology, Obstetrics and Neonatology, Prague, Czech Republic

Charles University, First Faculty of Medicine, Prague, Czech Republic

Board of the National Registry of Reproductive Health

Section C:

|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service

Pavlík Tomáš

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|17| Management tool for epidemic monitoring, analysis, and visualisation

Pichler Mario

Software Competence Center Hagenberg, Hagenberg, Austria

Section B:

|08| Visualisation of text-based data describing best practices in software development

Pokorná Andrea

Masaryk University, Faculty of Medicine, Department of Health Sciences, Brno, Czech Republic
Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section B:

|02| Complex structured evaluation of selected parts of the medical curriculum

Section C:

|13| Adverse Events Reporting System: Development of the platform for healthcare providers and lay public

Pospíšil Michal

Masaryk University, Faculty of Medicine, Department of Health Sciences, Brno, Czech Republic
Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|13| Adverse Events Reporting System: Development of the platform for healthcare providers and lay public

Prokopová Tereza

University Hospital Brno, Department of Anaesthesiology and Intensive Care Medicine, Brno, Czech Republic
Masaryk University, Faculty of Medicine, Department of Simulation Medicine, Brno, Czech Republic

Section B:

|06| Processing, analysis, and visualisation of objective structured clinical examination data

Přibylová Lenka

Masaryk University, Faculty of Science, Department of Mathematics and Statistics, Brno, Czech Republic

Section C:

|17| Management tool for epidemic monitoring, analysis, and visualisation

Ramler Rudolf

Software Competence Center Hagenberg, Hagenberg, Austria

Section B:

|08| Visualisation of text-based data describing best practices in software development

Růžička Michal

Masaryk University, Institute of Computer Science, Brno, Czech Republic

Section A:

Public (health) sector and academia

Růžičková Petra

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section B:

|06| Processing, analysis, and visualisation of objective structured clinical examination data

|08| Visualisation of text-based data describing best practices in software development

Ryšavý Jan

Czech Armed Forces, General Staff, Prague, Czech Republic

Section C:

|16| Interactive tool for crisis management in times of epidemic in the Czech Republic

Seifert Bohumil

Charles University, First Faculty of Medicine, Institute of General Practice, Prague, Czech Republic

Czech Medical Association of J. E. Purkyně, Society of General Practice

Section C:

|14| National health information portal: Text-based open data

Schwarz Daniel

Masaryk University, Faculty of Medicine, Department of Simulation Medicine, Brno, Czech Republic

Institute of Biostatistics and Analyses Ltd., Brno, Czech Republic

Section B:

|02| Complex structured evaluation of selected parts of the medical curriculum

|04| Detection of intersections between curriculum mapping and virtual patient information systems

Soukupová Jitka

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|10| Data and analytical basis for a modern mental health care system in the Czech Republic

Spachos Dimitris

Aristotle University of Thessaloniki, School of Medicine, Medical Education Informatics Group, Thessaloniki, Greece

Section B:

|05| Medical curriculum innovations using technological standards

Svačina Štěpán

General University Hospital in Prague, Third Clinic of Internal Medicine – Clinic of Endocrinology and Metabolism, Prague, Czech Republic

Charles University, First Faculty of Medicine, Prague, Czech Republic

Czech Medical Association of J. E. Purkyně

Section C:

|14| National health information portal: Text-based open data

Svoboda Marek

Masaryk Memorial Cancer Institute, Department of Comprehensive Cancer Care, Brno, Czech Republic

Section C:

|15| Innovations in patient-centred cancer care: Online benchmarking tool

Šanca Ondřej

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|10| Data and analytical basis for a modern mental health care system in the Czech Republic

Šnajdárek Petr

Czech Armed Forces, General Staff, Prague, Czech Republic

Section C:

|16| Interactive tool for crisis management in times of epidemic in the Czech Republic

Šnajdrová Lenka

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|09| Cancer screening programmes: Open data and visualisations as a support tool for monitoring and evaluation

|10| Data and analytical basis for a modern mental health care system in the Czech Republic

|11| National Registry of Reproductive Health: Data Infrastructure and Comprehensive Information Service

|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

|16| Interactive tool for crisis management in times of epidemic in the Czech Republic

Šteflová Alena

Charles University, First Faculty of Medicine, Institute of Public Health and Medical Law, Prague, Czech Republic

Czech Medical Association of J. E. Purkyně

Section C:

|14| National health information portal: Text-based open data

Štěrba Jaroslav

University Hospital Brno, Clinic of Childhood Cancer Care, Brno, Czech Republic

Masaryk University, Faculty of Medicine, Brno, Czech Republic

Section B:

|01| Outcome-based curriculum development and overview using an innovative online platform

Štourač Petr

University Hospital Brno and Masaryk University, Faculty of Medicine, Department of Paediatric Anaesthesiology and Intensive Care Medicine, Brno, Czech Republic

Masaryk University, Faculty of Medicine, Department of Simulation Medicine, Brno, Czech Republic

Section B:

|02| Complex structured evaluation of selected parts of the medical curriculum

|04| Detection of intersections between curriculum mapping and virtual patient information systems

|06| Processing, analysis, and visualisation of objective structured clinical examination data

Štrombachová Veronika

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section C:

|13| Adverse Events Reporting System: Development of the platform for healthcare providers and lay public

Táborský Miloš

University Hospital Olomouc, First Clinic of Internal Medicine – Cardiology, Olomouc, Czech Republic

Palacký University in Olomouc, Faculty of Medicine, Olomouc, Czech Republic

Section C:

|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

Turek Boris

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section A:

Health data opening

Vafek Václav

University Hospital Brno and Masaryk University, Faculty of Medicine, Department of Paediatric Anaesthesiology and Intensive Care Medicine, Brno, Czech Republic

Masaryk University, Faculty of Medicine, Department of Simulation Medicine, Brno, Czech Republic

Section B:

|06| Processing, analysis, and visualisation of objective structured clinical examination data

Vafková Tereza

Masaryk Memorial Cancer Institute, Department of Comprehensive Cancer Care, Brno, Czech Republic

Masaryk University, Faculty of Medicine, Department of Simulation Medicine, Brno, Czech Republic

Section B:

|06| Processing, analysis, and visualisation of objective structured clinical examination data

Vaitsis Christos

Karolinska Institute, Department of Learning, Informatics Management and Ethics, Solna, Sweden

Section B:

|05| Medical curriculum innovations using technological standards

Vičar Michal

Masaryk University, Faculty of Medicine, Institute of Biostatistics and Analyses, Brno, Czech Republic

Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic

Section A:

Health data opening

Section C:

|16| Interactive tool for crisis management in times of epidemic in the Czech Republic

Vodochodský Ivan

Newton Media Inc., Prague, Czech Republic

Section C:

|18| A pilot toolbox against health misinformation in the Czech Republic in the COVID-19 age

Vrablík Michal

General University Hospital in Prague, Third Clinic of Internal Medicine – Clinic of Endocrinology and Metabolism, Prague, Czech Republic

Charles University, First Faculty of Medicine, Prague, Czech Republic

Section C:

|12| National Cardiology Information System: Data Infrastructure and Comprehensive Information Service

Woodham Luke

St George's, University of London, Centre for Technology in Education, London, United Kingdom

Section B:

|05| Medical curriculum innovations using technological standards

ACKNOWLEDGEMENT OF THE EDITOR

My sincere thanks go especially to two exceptional persons from the Faculty of Medicine at Masaryk University, who gave me the opportunity to become interested and actively involved in the academic sphere, namely Ladislav Dušek and Daniel Schwarz. They allowed me to gain valuable experience from the beginning, provided me with background, patience and support for my ideas and visions, and above all, they have become my close friends for more than 15 years.

Many thanks are also due to all my colleagues at the Faculty of Medicine of Masaryk University and the Institute of Health Information and Statistics, who have been involved in the proposal, design and implementation of the projects on which the individual case studies of this book are based. The active, close and intensive collaboration between the institutional management, the project office, the analytical division and the development team led to the creation of a number of valuable and innovative outputs, for which, among other things, countless open datasets were created.

Last but not least, I would like to take this opportunity to thank my loved ones, family and close friends; without their support and tolerance, I would not have found the mental strength and time to write this book.

This book is specially dedicated to the memory of Ing. Hana Komendová (+ March 2023).

Martin Komenda

RELATED PROJECTS AND COMMUTIES

The three institutions listed below provide auspices for this book and are also part of globally significant consortia focusing on health data in various contexts.

- Faculty of Medicine of Masaryk University (www.med.muni.cz/)
- Institute of Computer Science of Masaryk University (www.ics.muni.cz/)
- Institute of Health Information and Statistics of the Czech Republic (www.uzis.cz/)

The content of this book is very closely linked to the implementation of selected national and international projects and activities across academia and the public sector.

PROJECTS AND ACTIVITIES (IN ALPHABETICAL ORDER):

- Adverse Event Reporting System (shnu.uzis.cz)
- Akutne.cz (www.akutne.cz/en)
- Building Curriculum Infrastructure in Medical Education (www.muni.cz/en/research/projects/43286)
- Evaluation in Medical Education (www.muni.cz/vyzkum/projekty/29268)
- Infomore.cz (www.infomore.cz)
- INTENT (programme2014-20.interreg-central.eu/Content.Node/INTENT.html)
- Medical Curriculum Innovations (medcin-platform.iba.muni.cz)
- MERGER (www.muni.cz/en/research/projects/36097)
- Monitoring, Analysis and Management of Epidemic Situations (mames.iba.muni.cz)
- National Health Information Portal (www.nzip.cz)
- National Portal of Mental Health Care (psychiatrie.uzis.cz)
- National Screening Centre (nsc.uzis.cz)
- Optimized Medical Education (www.muni.cz/en/research/projects/16523)
- Sharing Experience in Scientific Software and Applications Development (www.med.muni.cz/sessad)
- Simulation Centre of MED MUNI (www.med.muni.cz/simu)

INTERNATIONAL ASSOCIATIONS, COMMUNITIES AND NETWORKS (IN ALPHABETICAL ORDER):

- European Open Science Cloud (www.eosc.cz)
- Medical Faculties Network (www.mefanet.cz)
- MUNI Open Science (openscience.muni.cz)

ABBREVIATIONS

5A: Assess, Access, Analyse, Act, Automate
ACL: Access Control List
AE: Adverse Event
AERS: Adverse Event Reporting System
AI: Artificial Intelligence
API: Application Programming Interface
BCIME: Building Curriculum Infrastructure in Medical Education
CRISP-DM: Cross-Industry Standard Process for Data Mining
CRR: Central Reporting Repository
CSV: Comma-Separated Values
CurrMS: Curriculum Management System
CVD: Cardiovascular Disease
CZSO: Czech Statistical Office
DAX: Data Analysis eXpression
DIA: Digital Information Agency
DM: Data Mining
DMP: Data Management Plan
DOI: Digital Object Identifier
DSW: Data Stewardship Wizard
EBHC: Evidence-Based Healthcare
EBP: Evidence-Based Practice
ECMO: Extracorporeal Membrane Oxygenation
EGI: European Grid Infrastructure
EHR: Electronic Health Record
EOSC: European Open Science Cloud
ESFRI: European Strategy Forum on Research Infrastructures
ETL: Extract, Transform, Load
EU: European Union
EVAMED: Evaluation in Medical Education
EVLT: Endovenous Laser Therapy
FAIR: Findability, Accessibility, Interoperability, Reuse
FOBT: Faecal Occult Blood Test
GA: Google Analytics
GDPR: General Data Protection Regulation
HAI: Healthcare-Associated Infections
HCP: Healthcare Provider
HEI: Higher Education Institutions
ICD: International Classification Diseases
ICU: Intensive Care Unit
IHIS: Institute of Health Information and Statistics
ISID: Information System of Infectious Diseases
IVF: In-Vitro Fertilisation
KDD: Knowledge Discovery Databases
KM: Knowledge Model
LTS: Long-Term Support
MAMES: Monitoring, Analysis and Management of Epidemic Situations
MED MUNI: Faculty of Medicine of Masaryk University
MEDCIN: Medical Curriculum Innovations
MEFANET: Medical Faculties Network
MeSH: Medical Subject Headings
MZ CR: Ministry of Health of the Czech Republic
NAERS: National Adverse Events Reporting System
NCIS: National Cardiology Information System
NCHP: National Cardiovascular Health Plan
NDI: National Data Infrastructure
NHIS: National Health Information System
NREN: National Research and Education Network
NRHZS: National Registry of Reimbursed Health Services
NRP: National Repository Platform
NSC: National Screening Centre
NZIP: National Health Information Portal
OBE: Outcome-Based Education
OECD: Organisation for Economic Co-operation and Development
OECI: Organisation of European Cancer Institutes
OPTIMED: Optimized Medical Education
OSCE: Objective Structured Clinical Examination
OSINT: Open-Source Intelligence
PCCM: Patient-centred Cancer Care Model
PCRS: Physician Competency Reference Set
RFA: Radiofrequency Ablation
RFO: Research Funding Organisation
RIs: Research Infrastructures
SAML: Security Assertion Markup Language
SCCH: Software Competence Center Hagenberg
SEMMA: Sample, Explore, Modify, Model, Assess
SESSAD: Sharing Experience in Scientific Software and Applications Development
SIMU: Simulation Centre of MED MUNI
SRIA: Strategic Research and Innovation Agenda
UML: Unified Modeling Language
UMLS: Unified Medical Language System
VKHC: Virtual Know-How Centre
WHO: World Health Organization
XML: eXtendable Markup Language

CONTENT

INTRODUCTION BY EDITOR

Martin Komenda

At a time when we are surrounded by the everyday use of communication and information technologies very closely linked to the Internet, it is challenging to discern the accuracy, truthfulness and objectivity of the information presented and published. This book and its chapters aim to provide an overview of selected projects and activities across the academic and governmental domains focused on data processing and visualisation. It is crucial to recognise that, given the volume of data of varying quality that we now have at our disposal, we need to focus much more on understanding, identifying, and distributing correct information and inferences directly from the data. This is the main reason why this book was written. The individual case studies focus on examples of both good and bad practices, drawing on experiences from real-life projects. Data should always serve as a basis for decision-making processes and mechanisms, but only if they are correctly processed, understood, and, above all, interpreted. There are various ways to present results over descriptive statistics and data analysis, from summary tables to static graphs, to interactive web visualisations. It is only possible to say which type and presentation format is best with additional information (such as the target audience or primary purpose of use). The selected chapters in this book highlight the complete lifecycle of understanding, processing, visualising and validating data, so that all of the critical components of this process are remembered.

Every reader will likely ask for whom this book is intended, who is the target group, and who should and could be interested in this issue. The primary aim is to present the experience gained across real-world projects in medical education and health information comprehensively and transparently. The groups of readers for whom the book is intended undoubtedly include: representatives of middle and senior management of governmental and academic institutions involved in projects where data processing and presentation play a significant role; users dealing with open data, not only in the healthcare domain; enthusiasts who want to be inspired in their projects; and, last but not least, students of selected degree programs where this book will serve as one of the recommended study materials. We must remember the MEFANET educational network community, which brings together all Czech and Slovak medical and health sciences faculties, where the book is sure to be of use to management members, supervisors and teachers of selected courses, or even students. It is clear from this list that the total potential readership is quite broad and is not, nor should it be, limited by a direct focus on healthcare or medical education. The actual concept of individual chapters is certainly not a limiting factor in applying the basic principles to other domains of human interest and knowledge. Sharing experiences and recommending proven approaches to working with data are the underlying motivations of this book. Thus, it distinguishes between subjective and objective perceptions, feelings and impressions, and conclusions directly derived from data. This book highlights the importance of the difference between subjective opinion (the client's point of view supported by impressions and feelings) and objective opinion (measurable and challenging – quantitative – data). Especially today, when it is tough to distinguish truth from half-truths and misinformation, it is necessary to focus on and highlight the role of valid and correctly processed data.

The Faculty of Medicine of Masaryk University in Brno, one of the founding members of the MEFANET educational network, was at the beginning of the activities related to the outputs of this book. Since 2007, this community has focused on building and strengthening the cooperation among Czech and Slovak medical and non-medical faculties in the development of teaching using modern information and communication technologies. Over time, this effort has grown to European and then global dimensions. Today, dozens of international projects, conferences and workshops are proof of this. This collaboration aims to create a network of horizontally connected teaching sites and to facilitate interaction between teams from different faculties in a sustainable way. This enables students and teachers to effectively share electronic teaching materials, innovative teaching methods and technological solutions. The Institute of Health Information and Statistics of the Czech Republic, directly connected to the Ministry of Health of the Czech Republic, is the second institution that has contributed significantly to the outputs of this book. The National Health Information System (NHIS) is a unified nationwide public administration information system that collects and processes data from the basic registries of public administration bodies, ministries and health service providers. Data from NHIS were used for selected chapters in this book, highlighting the importance of the correct release and publication of datasets in open data format.

STRUCTURE OF THE BOOK

This book is divided into three main sections:

i. Introduction (general background, EU context, methodologies)

ii. Medical and healthcare education in selected case studies

iii. Health information and statistics in selected case studies

Each chapter, except the introduction, has the same format describing a particular project result as a case study, which is always based on a well-proven interdisciplinary methodology (specifically CRISP DM – Cross-Industry Standard Process for Data Mining – the structured approach to planning and running data mining projects.

— As a methodology, it includes descriptions of individual project phases, the tasks involved with each stage, and the relationships between them.

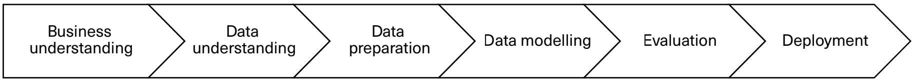
— As a process model, it provides an overview of the complete data mining life cycle.

The case studies in this book are always based on the achieved results of a specific project from academia or the public sector. In the case of the second section (eight chapters dealing with medical and healthcare education), the data originally comes mainly from the internal systems of Masaryk University, which are designed to support teaching and learning activities (the OPTIMED portal platform, the SIMUportfolio integration platform, Information System). The input data sets in the third section (ten chapters dealing with health information and statistics) are mainly from the health open data catalogue or data obtained in a national or international collaboration framework among healthcare institutions. In neither case, however, was primary patient data ever handled and the sensitive issue of data protection during data processing and publication processes was always strictly observed.

The usage of statistical and analytical methods in this book is based on long-term experience, often solving particular needs and tasks over available data. The basis is always detailed descriptive statistics supported by business intelligence tools, and a specific set of analytical procedures and algorithms is designed as a superstructure, considering the characteristics and challenges of the particular project. Machine learning methods, especially the nowadays highly emphasised artificial intelligence (AI), are not yet among the approaches tried and tested by our teams. Moreover, many of the mentioned projects were carried out several years ago when the results of AI methods were much less accurate and precise. It should be added that so far, the role of a human as a guarantor is irreplaceable. However, from the long-term perspective, the selected AI software methods will likely be applicable to solve similar tasks described in the case studies.

STRUCTURE OF CASE STUDIES[1]

- Title
- Authors
- Affiliations
- Dedication section
  - expresses gratitude or acknowledges
- Keywords
  - main characteristics describing each case study
- Main research question
- Year of result achievement
- CRISP-DM: One or more of the most crucial phases



- Type of result
  - Complex web application
  - Interactive visualisation
  - Static analytical report
- Level of data processing
  - Advanced analyses
  - Descriptive statistics
  - Open datasets

[1] Based on IMRAD structure: SOLLACI, Luciana B.; PEREIRA, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. Journal of the medical library association, 2004, 92.3: 364.

- Data to download
- Introduction
  - introduces the case report, including the background and any previous studies of the given task/issue
- Aims
  - describes the purpose of a particular study using specific research questions using action verbs (to describe/identify/analyse/recommend…)
- Methods
  - explains how the study was carried out from a methodological and technical perspective
- Results
  - describes what the authors found through solving the given problem using CRISP-DM
    - Business understanding
    - Data understanding
    - Data Preparation
    - Modelling
    - Evaluation
    - Deployment
- Discussion
  - explains the significance of the study and what can be learnt from it
  - evaluation of the aims of the chapter
- Lessons learned
  - provides proposals for future action to solve the problem or improve the situation and lessons learned
  - highlights the most crucial CRISP-DM phases, including a short justification

FOREWORD

ACADEMIA LEVEL

Martin Repko
Dean
Masaryk University, Faculty of Medicine, Brno, Czech Republic

Modern medicine has made extraordinary progress in recent decades, and especially in recent years. In this respect, excellent medical data and analyses have made an extraordinary contribution to correct decision-making in diagnostic and therapeutic practice. On the one hand, the massive development of the digital world has brought a considerable amount of possible information; on the other hand, it has presented us with the problem of handling and using these data rationally. In particular, advances in openly shared and correct medical data create a rational basis for valid diagnostic and therapeutic reasoning in clinical practice.

The present publication combines excellent statistical-analytical background with case-based elements to demonstrate valid practice in the daily routine of clinical and preclinical biomedical data professionals. The many years of systematic work in this field by Dr. Komenda and his co-authors have resulted in a unique publication that presents first-hand experience in many areas of medicine, the use of biomedical data, and their correct use. These data not only serve to further progress in patient care but also give a new dimension to modern education in biomedical fields. In this publication, the authors draw, among other things, on their own experience with the unique MEFANET project, which has been running for more than 16 years and has significantly contributed to the networking and development of teaching in medical faculties using modern information and communication technologies. I firmly believe that the present publication will substantially enrich all those who wish to deepen the rational practices and proper use of modern technologies and contribute to the use of exact data in medical research and treatment settings.

GOVERNMENT LEVEL

Vlastimil Válek
Deputy Prime Minister and Minister of Health
Ministry of Health of the Czech Republic, Prague, Czech Republic
Masaryk University, Faculty of Medicine, Brno, Czech Republic

As the Minister of Health of the Czech Republic and Deputy Prime Minister of the Government of the Czech Republic, I am very pleased to welcome this publication, which demonstrates that the Czech health sector has taken a big step towards using data for diagnostic, clinical and managerial decisions. We have many strategic tasks ahead of us that cannot be accomplished without data; for example, optimising reimbursement in all segments of care, setting up an accurate system for assessing the availability and quality of care or planning the necessary staff capacity and the associated support for training. In all of these areas, we need not only input data, which point to the system's weaknesses, but also continuous data, which will allow us to evaluate the effect of the measures taken.

The National Health Information System and some of its outputs, which are already available, are an excellent example of how the data and information basis of modern healthcare should be developed. However, I do not intend to overlook the second essential part of the book, i.e. the standardisation and parameterisation of the content of education in clinical and medical disciplines. These outputs and methodological materials are also of considerable international importance, as they prepare our healthcare sector for the inevitable advent of the era of working with open data, i.e. the European Health Data Space (EHDS).

First of all, however, I would like to emphasise in this introduction the essential contribution of this publication to the digitisation of healthcare. Building a modern eHealth system is not just about digitising processes and sharing e-documentation but especially about standardising the content of documentation. Without a standardised collection of diagnostic and clinical data, it is impossible to create meaningful analyses or automate information services, and therefore not even to implement data-driven decision-making.

Therefore, as Minister of Health, I have initiated major steps in the field of standardisation of clinical information systems, which I will summarise in three levels:

1. It is necessary to define interoperability standards for health information systems, and to ensure that these standards are the same as in other parts of the EU. Although the transition period will last until the end of 2026, new installations will already have to take place now in these standards. We refer to this step, in working terms, as Standard A.

2. It must be possible to send requests for examinations and treatments directly to the information systems of healthcare institutions or to transmit them using the citizens' electronic health cards. I will give an example that is close to my heart as a radiologist. If a doctor issues a request for a CT scan, there are basically three options for getting that request to the relevant department: either the patient brings it in paper form (as is often the case today), or it is sent electronically (similar to how we send CT scan images today), or the patient uploads it to their electronic health card (a special mobile app) and brings it to the relevant department. Electronic appointment scheduling (online calendars) must be a condition of approval for new equipment, including renewals. The electronic transmission of examination findings and final discharge reports (or patient summaries) from healthcare facilities must be linked to apps accessible to every citizen. We refer to the standardisation of examination requests as Standard B.

3. The next necessary step is to standardise the structure of the examination description and discharge report. The degree of standardisation (including the structured description and final report) may vary from facility to facility; it will also depend on the needs of the care segment or specialty. We refer to this step as Standard C.

At this point, I would like to thank the professional societies that have already made significant progress in designing the parameterisation (standardisation) of the content of their medical records, as is evident in many chapters of this publication. For example, very sophisticated proposals can be seen in cancer reports, including the results of screening tests, acute treatment of strokes, treatment of cardiovascular diseases, vaccination records by general practitioners, etc. I firmly believe that this publication is the first step towards further standardisation and digitisation of our healthcare system.

EXPERT IN DATA ANALYSIS

Ladislav Dušek
Director
Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic
Masaryk University, Faculty of Medicine, Brno, Czech Republic

Data-driven decision-making, particularly assessing the availability and quality of health services, has recently become an integral part of healthcare. This applies to all developed countries worldwide, which, apart from lifestyle diseases, must cope with a noticeably ageing population. However, very comprehensive data are needed to evaluate healthcare [1]. The data basis of any medical evaluation can be defined as a complex set of parameters that describe the input characteristics of the evaluated subjects, their subsequent development and the results obtained. For the data basis to be functional, the data obtained must be relevant to the evaluated process (i.e., they must carry relevant information value) and obtained in a clearly defined format (i.e., they must have usable information value). In clinical practice, both requirements are far from being always met, so reduced data availability becomes a limiting factor for outcome assessment. This is a paradoxical situation, as the cost of data collection and evaluation represents only a small part of the total investment in the treatment itself, yet it can substantially increase the whole system's efficiency. The desire to remedy this situation is the primary motive for the progressive development of the National Health Information System. The case studies published in this book provide clear evidence of its successful implementation.

Studies focusing on reproductive health, cancer care or population-based screening programmes demonstrate the successful utilisation of already available data sources and the subsequent use of the information obtained to standardise the actual data collection. This is crucial, as the most significant limitation to development in this area is the lack of standardisation of routine clinical data collection in healthcare facilities. Electronic and fully parametric patient documentation is not sufficiently optimised in many medical disciplines. This problem is far from limited to the Czech Republic [2,3], and the publication of successful solutions can have a considerable international impact.

The importance of guaranteed data quality increases with the growing volume of multidimensional clinical data sets. The so-called “molecularisation” of contemporary medicine generates data sets where the number of evaluated markers exceeds the number of patients by several orders of magnitude. In parallel, the volume of available data is increasing: according to international literature, the volume of archived clinical data approximately doubles every year. These trends naturally increase the pressure for structured and well-documented storage of data sets. Only standardised and well-described data can serve as the basis for interoperability of information systems. The progressive computerisation of healthcare has been shown to increase the usability of administrative data and reduce burdensome repetitive reporting to different systems. In the Czech Republic, the new Act No. 325/2021 Coll., on the Digitisation of Healthcare [4], is in force. This Act and other regulations have defined different levels of clinical data registration and established procedures for the maintenance and standardisation of electronic medical records.

In addition to their informative content, the case studies presented in this book have a considerable methodological benefit. They illustrate the changing position of evidence-based medicine (EBM) in various clinical applications. Indeed, the term EBM is often simplified to the methodology of interventional prospective clinical trials, which, of course, provide the highest level of confidence in the evidence. However, the spectrum of EBM methodologies is much broader, and there is now a growing need for exact analyses of data from real clinical practice. The so-called real-world evidence (RWE) must not be perceived as a counterpart or even a competitor to EBM: the two approaches are complementary and mutually supportive.

Modern clinical research faces many methodological challenges in segments of medicine where the rapid pace of innovation does not allow for multi-year follow-up of large patient cohorts. There are also complications in areas where the heterogeneity of patient characteristics is so high that a sufficiently large yet homogeneous sample of subjects cannot be obtained. Typical examples of the latter situation include palliative care or systems that monitor the evolution of infectious disease epidemics in real time [5]. High-quality registries can provide an almost unbiased picture of reality in these areas. In particular, protocol-equipped observational studies, based on the platform of clinical registries, should be perceived as an essential research tool that extends – and often corrects – the conclusions of clinical trials but does not replace them [6].

The published case studies are also textbook examples of sharing and making data available for secondary use, which is undoubtedly one of the priority goals of modern health information systems. Sharing and publishing open data is a systematic way to make even large data sets available for further manual or machine processing in a uniform and technically well-defined form. At the beginning of this development, there was a particular drive towards transparency and making data available for management purposes; however, as the digitisation of health services progresses and the volume of centralised data increases, the objectives are expanding substantially. Open data systems determine the development of progressive areas of medical research, such as medical bioinformatics or personalised medicine. In addition, the planned real-time access to data is one of the conditions for the effective digitisation of health services, and de facto forms the basis for the development of entirely new segments of health and social care, telemedicine and systems for remote patient monitoring. I believe that this publication is a step towards further development in this direction.

[1] Topol EJ. Transforming medicine via digital innovation. Sci Transl Med 2010; 2 (16): 1 – 3.

[2] Kawamoto, K., Houlihan, C.A., Balas, E.A., Lobach, D.F.: Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Brit. Med. J. 330 (2005), pp. 765-768.

[3] Dick, R.S., Sheen, R.B.: The Computer-Based Patient Record: An Essential Technology For Health Care, National Academy Press, Washington DC, (1991)

[4] Těšitelová V, Blaha M, Klimeš D, Policar R, Dušek L. Elektronizace zdravotnictví řečí paragrafů: VERZE 1.1. Praha: Ústav zdravotnických informací a statistiky ČR, 2021, 345 s. ISBN 978-80-7472-189-2.

[5] Nattinger, A.B., McAuliffe, T.L., Schapira, M.M.: Generalizability of the Surveillance, Epidemiology, and End Results registry population: factors relevant to epidemiologic and health care research. J. Clin. Epidemiol. 50 (1997), pp. 939–945.

[6] Ahern S, Hopper I, Evans SM. Clinical quality registries for clinician-level reporting: strengths and limitations. Med J Aust 2017; 206 (10)

SECTION A

INTRODUCTION

THE BIG PICTURE

Martin Komenda, Jiří Marek

TERMINOLOGY

This book focuses on data analytics and different approaches to visualise the obtained results, which are directly used in the decision-making processes. Before briefly describing the methodological background that underpins each of the chapters, let us take a detailed look at the key concepts and their explanation within the context of this publication. In today's world, which is full of information, it is possible to find multiple definitions and descriptions of the same concept. For the purposes of this publication, a published and verified source of information was always chosen that fit well within the context of the editor's point of view on the subject.

LITERACY

Information literacy is the adoption of appropriate information behaviour to obtain, through whatever channel or medium, information well fitted to information needs, together with critical awareness of the importance of wise and ethical use of information in society [1].

Data literacy is the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle and ethically use data [2].

Statistical literacy is envisaged as the component of data literacy involved in the critical appraisal, interpretation, processing and statistical analysis of data [2].

Health literacy refers to the ability to find, understand and correctly use information on health and healthcare. This includes, for example, information on disease or lifestyle risk factors, invitations to appointments, package leaflets, instructions from health professionals, basic orientation in the healthcare delivery system, knowledge of the symptoms of common diseases, knowledge of the basic functions of the human body, knowledge of the basic steps to take care of oneself or to ensure self-sufficiency in the event of illness [3]. Improving health literacy in populations provides the foundation on which citizens are enabled to play an active role in improving their own health, engage successfully with community action for health, and push governments to meet their responsibilities in addressing health and health equity. Meeting the health literacy needs of the most disadvantaged and marginalised societies will particularly accelerate progress in reducing inequities in health and beyond [4].

KNOWLEDGE DISCOVERY

Knowledge discovery in databases (KDD) is often used in data mining. It is a non-trivial process of discovering novel and potentially useful information from large amounts of data and aims to identify new understandable patterns in the data convincingly [5].

Data mining is often a synonym for extracting useful information from databases. It refers to the application of algorithms for extracting patterns from data without the additional steps of the knowledge discovery process [6].

There needs to be more consistency in data mining (DM) and knowledge discovery in databases (KDD) in the literature and web resources. Some authors use these terms synonymously. Thus, both terms more or less mean the same thing; with KDD, the actual preparation of the data is also considered essential (Figure 1). This figure shows the complexity of data processing to extract valid and correct information or possibly knowledge. Below, the main steps are shortly described [5].

1. Data: Building a data domain based on a detailed understanding will then be worked with.

2. Selection: Choosing and creating a data set on which the discovery process will be performed.

3. Preprocessing: Application of basic operations such as handling the missing values, noise removal, data format unification, etc.

4. Transformation: Finding functional characteristics of data, dimensionality reduction, discretisation of numerical attributes, etc.

5. Data mining: Matching the given aims (step 1) to a particular data mining method (i.e. summarisation or classification), choosing the proper data mining algorithm and application of the selected algorithm to search the patterns (i.e. classification rules or trees).

6. Interpretation / Evaluation: Assessment and correct interpretation of the mined results, possibly return to any of the above-mentioned steps (i.e. 1 to 5) for further iteration.

7. Knowledge: Incorporating achieved knowledge for further action.

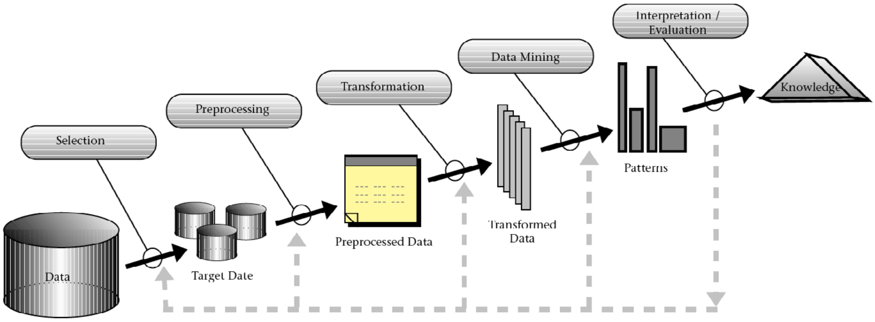


Figure 1: The overview of tile steps comprising the KDD process

Generally, the main goal of proven process methodologies (see section Methodologies for data processing) is to provide users with a unified framework for solving various knowledge discovery tasks. These methodologies allow for sharing and transferring experiences from successful projects. This book contains individual case studies based on the proven KDD methodology. The separate stages, which are appropriately linked to each other, make it easy for the user to understand each stage of the lifecycle while giving it the attention it deserves.

When used correctly, the methodology enables the correct and, above all, complete implementation of a process that leads from the successful mining of **data** (objective facts with no further information regarding patterns or context), through the acquisition of hitherto unknown **information** (contextualised, categorised, calculated, and condensed data with a specific meaning painting a bigger picture), to **knowledge**, which is closely linked to doing and implies know-how and understanding (information with added value, including a specific context) [7,8].

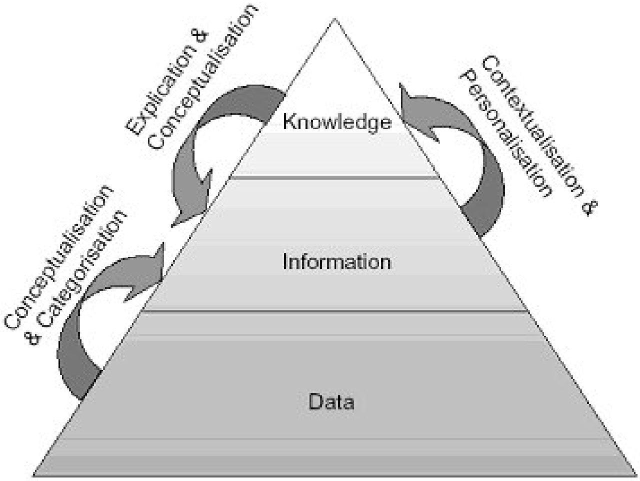

Explication & Conceptualisation
Contextualisation & Personalisation
Knowledge
Conceptualisation & Categorisation
Information
Data

Figure 2: Data-Information-Knowledge pyramid

For a better understanding, a real example from the field of curriculum description and mapping is given below:

Floor of pyramid	Example
Data	A learning unit (typically one lecture of practice) is defined by several descriptive characteristics, i.e. title, section, range (in hours), type, annotation, keywords, medical discipline, etc. A total of 1,342 learning units were described in the General Medicine study programme at the Faculty of Medicine of Masaryk University.
Information	The Internal Medicine section contains the most learning units (446), and Diagnostic Sciences and Neurosciences (226) the least. Ratio[1] = 1.97
Knowledge	The average number of learning outcomes () listed per learning unit does not correspond to the proportion of learning units represented in each learning section. Internal Medicine 5.5 learning outcomes per learning unit Diagnostic Sciences and Neurosciences 7.7 learning outcomes per learning unit Ratio[1] = 0.73 Moreover, the total teaching range (in hours) does correspond to the number of learning units in each section. Internal Medicine 2,492.1 learning outcomes per learning unit Diagnostic Sciences and Neurosciences 1,225.5 learning outcomes per learning unit Ratio[1] = 2.03

[1] Internal Medicine / Diagnostic Sciences and Neurosciences

Although the topic of **artificial intelligence** (AI) is not the main focus of this publication, given the current rapid development of this approach, it is appropriate to mention it here in passing. The application of AI in medicine has two main branches: (i) virtual component represented by **machine learning** techniques also called deep learning (i.e., electronic medical records where specific algorithms are used to identify subjects with a family history of a hereditary disease or an augmented risk of a chronic disease), (ii) physical component includes physical objects, medical devices and increasingly sophisticated robots (i.e., robot companion for the aging population with cognitive decline or limited mobility) [9].

Especially machine learning techniques, among all fields of human interest, represent a promising approach for extracting knowledge and high added value of the vast amount of high-granularity data. For the illustration, (Figure 3) demonstrates the most relevant data-driven related areas used for power systems data processing and analysis [10]. Generally speaking, the proper usage of AI tools that follow human behaviour based on collected data and are under human supervision can be the starting point for developing data-driven services.

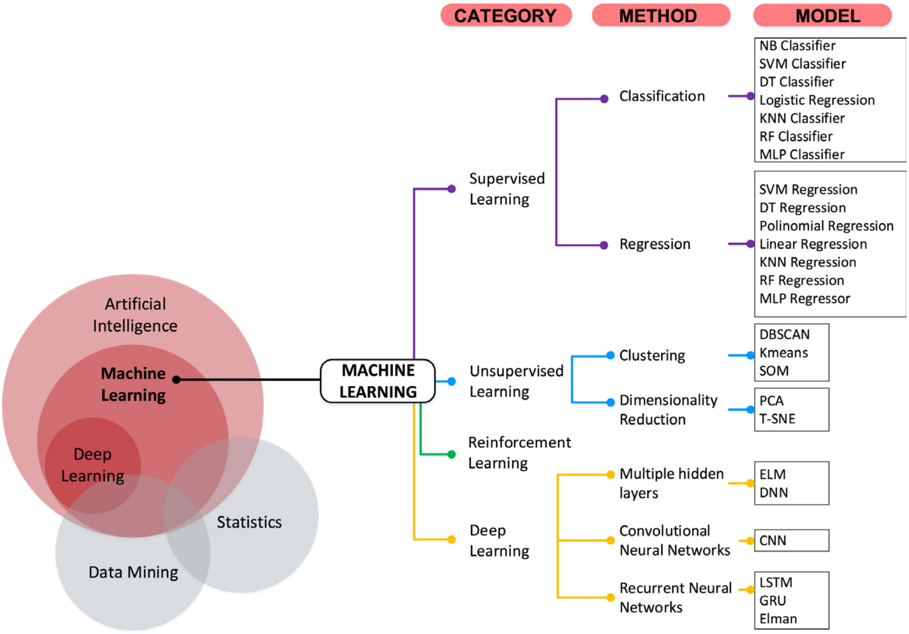


Figure 3: Data-driven techniques classification in the context of machine learning categories for power systems analysis

OPEN COMMUNITY

Open science is a collection of actions designed to make scientific processes more transparent and results more accessible. Its goal is to build a more replicable and robust science; it does so using new technologies, altering incentives, and changing attitudes. The most well-known initiative from the Open Science agenda is Open Access to scientific publications or Open/FAIR Data approach towards managing research data (in Europe, mostly connected for now with the initiative of European Open Science Cloud – EOSC) [11,12].

Open data, according to § 3 par. 5 of Act No. 106/1999 Coll. on Free Access to Information (Czech Freedom of Information Act) and in line with the directive (EU) 2019/1024 of the European Parliament and of the Council on open data and the re-use of public sector information, is "… information published in a manner that allows remote access in an open and machine-readable format, the manner and purpose of subsequent use of which is not restricted by the obliged entity publishing it and which is registered in the national catalogue of open data". In accordance with this definition and the recommended practices, each open dataset must comply with the rules set out below, independently on the field of human interest:

1. It is accessible as a data file in a machine-readable and open format with complete and up-to-date database content or aggregated statistics.

2. It is provided with non-restrictive terms of use, with the terms of use being set as follows, depending on the nature of the content of the dataset:

3. A CC BY 4.0 public license by which the copyright holder allows free use of his works provided that the user of the work credits him as the author.

4. The CC0 public license, which serves as a means of waiving the database rights of the database creator.

5. It is registered in the National Catalogue of Open Data as a direct link to the dataset.

6. It is accompanied by clear documentation.

7. It is available for download without technical barriers (registration, access restrictions, CAPTCHA, etc.).

8. It is prepared with the aim of making it as easy as possible for programmers, etc., to be machine-processed.

9. It is provided with a curator contact for feedback (bugs, extension requests, etc.).

FAIR data refers to a set of principles, focused on ensuring that research objects are reusable, and actually will be reused, and so become as valuable as is possible [13]. Making data available according to FAIR data principles require being in compliance with these attributes, which are composed of several recommendations clustered around the acronym FAIR:

- **F**indable – metadata, registration, global persistent identifiers
  - Data should be easily discoverable, allowing researchers to locate and identify the data of interest. This is achieved through persistent identifiers, standardised metadata, and comprehensive data descriptions.
- **A**ccessible – standards for machine-readability, authentication and authorisation infrastructure
  - Data should be readily accessible to both humans and machines. It should be available through well-defined access protocols, with minimal barriers to access, such as login requirements or subscription fees.
- **I**nteroperable – semantic description of data and metadata, ontologies, standards
  - Data should be structured and organised to facilitate integration and interoperability with other data sources. This involves using standardised data formats, adopting common vocabularies and ontologies, and providing clear data and metadata specifications.
- **R**eusable – clear licensing, data provenance (reproducibility) [14]
  - Data should be designed and documented in a manner that enables its reuse for different purposes. This includes providing detailed information about the data's provenance, methodology, and context, as well as clear licensing and usage terms.

FAIR data are not in conflict with open data, these two terms are usually linked together. To simplify it, all data can be FAIR, but only some data can be open (personal data handling restrictions, commercialisation aspects, etc.). Therefore, health data will usually be more FAIR than open data due to the extensive need to handle sensitive data.

DATA ANALYSIS AND DELIVERY

Data analytics is the application of computer systems to analyse large data sets to support decisions. This interdisciplinary field has adopted aspects from many other scientific disciplines, such as statistics, machine learning, pattern recognition, system theory, operations research, and artificial intelligence. Such an approach allows us to find relevant information, structures, and patterns, gain new insights, identify causes and effects, predict future developments, or suggest optimal decisions. We need models and algorithms to collect, preprocess, analyse, and evaluate data [15].

BUSINESS INTELLIGENCE

- **as a process** can be defined as the process of turning data into information and then into knowledge. Knowledge is typically obtained about customer (i.e. representative of the Ministry of Health of the Czech Republic, health insurance company, open data community, etc.) needs, customer decision-making processes, the competition, conditions in the industry, and general economic, technological, and cultural trends [16].
- **as a process and a product** can be used to refer to an organised and systematic process by which organisations acquire, analyse and disseminate information from both internal and external information sources significant for their business activities and for decision-making [17].
- **as a process, a product and technologies** encompass a set of tools, techniques, and processes to help harness this wide array of data and allow decision-makers to convert it to useful information and knowledge [18].
- **in general**, it allows managers to make informed and intelligent decisions regarding the functioning of their organisation [19].

DATA-DRIVEN APPROACH

People in academia, government and business sectors often make their decisions based on their subjective opinions and habits. A data-driven approach helps to bring objectivity and facts into decision-making. It must be emphasised, however, that data alone do not have the necessary and required informative and telling value. Context and correct interpretation must always be considered. The context and assumptions represent external aspects out of the control of any decision-maker, but the premises and the knowledge of the company depend on available data [20]. One of the definitions introduces **data-driven decision-making** as the practice of basing decisions on data analysis rather than purely intuition [21]. On the government or the public sector level, **evidence-based policymaking** is similar in that decisions are based on factual data [22]. In both cases, the following applies: the more data are available, the more people, stakeholders and institutions can construct their perceptions and decisions of reality.

METHODOLOGIES FOR DATA PROCESSING

Setting standards in data mining primarily results in methodological guidelines on how to achieve this goal. It has become essential because of the increased demand for methodologies and tools to help analyse and understand data and, last but not least, to make data more interoperable for further easier exchange or processing. Many related standards and recommendations in this area have already been set up and published (Figure 4) [23]. SEMMA (Sample, Explore, Modify, Model, Assess), 5A (Assess, Access, Analyse, Act, and Automate), and CRISP-DM (CRoss-Industry Standard Process for Data Mining) are considered to be the most frequently used methodologies [24]. Each phase is crucial: it is not just about processing and visualising the data. It is good to remember that the lifecycle of long-standing and field-tested methodologies has its meaning and importance, and needs to be thoroughly addressed. Nowadays, we are inundated with data on a daily basis due to the routine operation of Internet applications and various information systems. All of them involve the generation of data, backups and archives, be it telecommunications, banking transactions or scientific research. The different phases of the chosen methodology will provide the space for a proper understanding and addressing of the differences that characterise the individual domains. In general, the process guides the interpreter of a given problem through theoretically very well-described steps, which always contain a set of detailed actions by the selected algorithm.

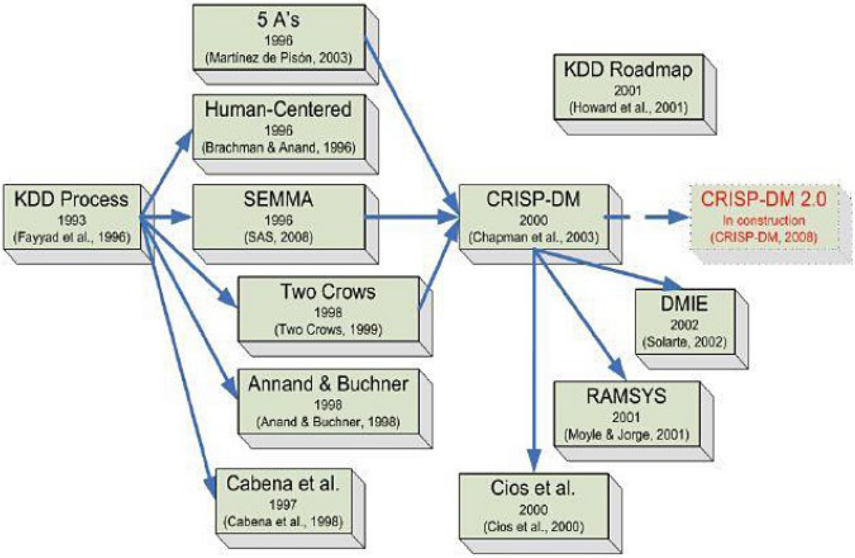


Figure 4: Evolution of DM & KDD process models and methodologies

Each chapter of this book is based on the most proven CRISP-DM methodology in practice, which consists of six loosely linked steps (Figure 5). Although these steps may seem trivial and simple, one or even more of them are often forgotten – and this is one of the main reasons why the complete process of each methodology needs to be very thoroughly adopted, addressed, and then correctly applied in real life:

1. what to solve (Business understanding) – understanding the problem, formulating the task,
2. where to get data (Data understanding),
3. how to prepare data (Data preparation),
4. how to analyse data (Data modelling),
5. what we found (Evaluation) – understanding the results,
6. how to use the results (Deployment).

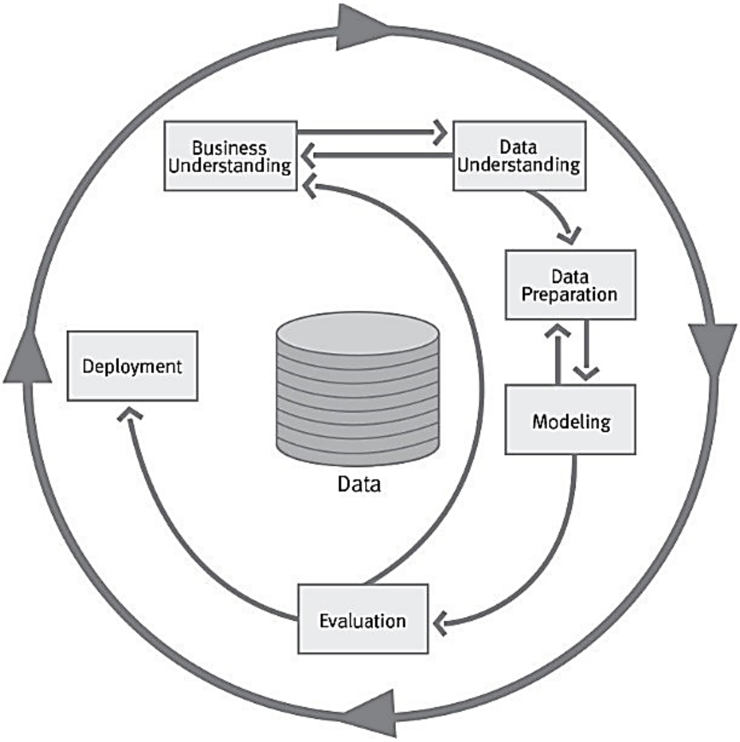


Figure 5: Diagram of CRISP-DM

1. **Business understanding** is a step towards understanding of particular domain that always hides the specifics of a given expertise. In the case of this book, this is the domain of healthcare, medicine, education, health literacy, and health informatics. Examples include working with thesauri or vocabularies closely related to, for example, selected nomenclature (International Classification of Diseases), a specific vocabulary for an individual medical domain (MeSH, SNOMED). The legislation is also a very important agenda, especially the nowadays very accentuated and sensitive issue of data protection (GDPR, General Data Protection Regulation). When working with data, it is necessary to think about protecting an individual's data and preventing their identification at all times. Finally, it is essential to recognise whether descriptive attributes (columns characterising a given row like an instance) are present. If the solver has thoroughly mastered this methodology step, he may need to pay more attention to it in subsequent tasks than in the beginning. In any case, it is crucial to highlight that, due to legal, privacy and sensitivity considerations, not all health data can be open; more often,

however, it should be FAIR. Last but not least, an understanding of the data lifecycle and how data are created must be emphasised. Who enters it into the system and how, what automatic checks and validation rules are implemented in the information system, and how long is the update interval? These processes are not always set up; therefore, the selected case studies focus on a thorough understanding of the data lifecycle.

2. Data understanding is directly related to the first of the steps. An example would be correctly selecting, understanding and using the above-mentioned vocabularies. Business understanding logically translates into a subsequent data understanding and its proper processing. This is not only in terms of content correctness but also economics (how much will it cost?) or personnel (do I have enough experts?). With a detailed check and subsequent knowledge of all the dataset's characteristics, it is possible to process and correctly present the data (regardless of the form - table, graph, or interactive visualisation). The quality of the input data determines the possibilities of descriptive statistics and the use of advanced analytical methods. Unfortunately, the input data may need to be completed or burdened with noise. This is a random error, but it affects the quality of the data. This is where the power of the Open and FAIR Data concept comes into full play, which always strictly requires a complete metadata description and, therefore, complete input information for anyone working with the data. In addition, of course, the technical treatment of the dataset is also essential, e.g. the data types or codebook structure used. This apparent small and technical detail can often cause considerable processing, mapping, and transformation complications.

3. The next step is to **prepare the data** in the form of pre-processing. Here we are talking about a significantly individualised approach directly related to the quality of the input data. Data cleaning, selecting relevant attributes, filling in missing values (e.g. using suitable open datasets), conversion or unification of data types, mapping multiple datasets to one, or transformations are standard and not entirely trivial techniques.

4. Modelling is used in basic descriptive statistics, analytical or machine learning methods (e.g. decision trees, clustering or association rules) etc. A combination of multiple ways is typically used to provide a comprehensive view of the available dataset. Undoubtedly, it also includes visualisation of the results obtained through basic or advanced statistical and analytical techniques. It must be emphasised here that the final form, i.e., how the results of the solution to the problem are also presented, can profoundly affect how the outputs are understood and subsequently used in practice. Simplicity, clarity, and comprehensibility are essential from a long-term

perspective and experience, of course always related to the named target groups for whom the outputs are primarily intended.

5. Despite its crucial importance, the **evaluation** phase is often neglected and should be addressed more thoroughly. Here it is not only the researcher's detailed manual or machine validation which should be automatically included in the entire CRISP-DM process. Feedback should also be provided by an expert who is ideally not part of the research team but belongs to the target group. Such experts are specialists in a particular field and can provide a relevant and valid review. This process should also include evaluating the results achieved against the exploratory/research questions and an overall performance assessment against the set brief. It is logical that, as in other phases, the outcome of the evaluation may be the need for more or less intervention in the previous steps. This may even mean a redesign or a more fundamental change in the solution of the whole task.
6. The **deployment** or application of the output in practice is the last phase of this methodology. It is not just a technical matter of, for example, releasing a new version of a web application or an interactive infographic. Compared to the results obtained in the modelling phase, there may be slight modifications resulting from the possibility of implementing the results in practice. For example, we can talk about publishing an open dataset, a single summary data table in the given context, a more detailed static presentation with graphs, explanations, and conclusions, or a comprehensive interactive data visualisation, including selections of custom views of the data using filters.

Moreover, a complex research data lifecycle model [25] can significantly help re-use data in a different context, such as research and development or policymaking, and it also provides more interoperability in various case studies on these data. This model consists of the following six stageTable 1 1. This concept also follows the research data management toolkit for life sciences[2], which introduces best practices and guidelines to help make data more manageable in accordance with the Horizon Europe Programme Guide recommendations.

[2] https://rdmkit.elixir-europe.org/data_life_cycle

Table 1: The research data lifecycle model.

Stage	Activities
Creating data	• design research • plan data management (formats, storage etc.) • plan consent for sharing • locate existing data • collect data (experiment, observe, measure, simulate) • capture and create metadata
Processing data	• enter data, digitise, transcribe, translate • check, validate and clean data • anonymise data where necessary • describe data • manage and store data
Analysing data	• interpret data • derive data • produce research outputs • author publications • prepare data for preservation
Preserving data	• migrate data to best format • migrate data to suitable medium • back-up and store data • create metadata and documentation • archive data
Giving access to data	• distribute data • share data • control access • establish copyright • promote data
Re-using data	• follow-up research • new research • undertake research reviews • scrutinise findings • teach and learn

REFERENCES

[1] Weber S. Getting the knowledge. Library and Information Update. 2002;1:52–3.

[2] Calzada Prado J, Marzal MA. Incorporating data literacy into information literacy programs: Core competencies and contents. Libri. 2013;63(2):123–34.

[3] Institute of Health Literacy. Health literacy [Internet]. Prague: Ministry of Health of the Czech Republic; 2023 [cited 20 Jul 2023]. Available from: https://www.nzip.cz/clanek/226-zdravotni-gramotnost.

[4] World Health Organization. Improving health literacy [Internet]. [cited 16 Jul 2023]. Available from: https://www.who.int/activities/improving-health-literacy.

[5] Fayyad UM, Piatetsky-Shapiro G, Smyth P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: Simoudis E, Han J, Fayyad UM (eds). KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Washington: AAAI Press; 1996. p. 82–8.

[6] Fayyad UM, Stolorz P. Data mining and KDD: Promise and challenges. Future generation computer systems 1997;13(2–3):99–115.

[7] Davenport TH, Prusak L. Working Knowledge: How Organizations Manage What They Know. Brighton: Harvard Business Press; 1998.

[8] Zimmermann A, Lorenz A, Specht M. The Use of an Information Brokering Tool in an Electronic Museum Environment [Internet]. 2003 [cited 16 Jul 2023]. Available from:http://www.archimuse.com/mw2003/papers/zimmermann/zimmermann.html.

[9] Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69S:S36–S40.

[10] Barja-Martinez S, Aragüés-Peñalba M, Munné-Collado I, et al. Artificial intelligence techniques for enabling Big Data services in distribution networks: A review. Renew Sust Energ Rev. 2021;150:111459.

[11] Spellman B, Gilbert E, Corker KS. Open science: What, why, and how [Internet]. PsyArXiv; 2017. Available from: https://psyarxiv.com/ak6jr/.

[12] European Commission. Open Science [Internet]. European Commission; 2020 [cited 16 Jul 2023]. Available from: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en.

[13] Mons B, Neylon C, Velterop J, et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Inform Serv Use. 2017;37(1):49–56.

[14] Wilkinson M, Dumontier M, Aalbersberg I, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018.

[15] Runkler TA. Data analytics: Models and Algorithms for Intelligent Data Analysis. Wiesbaden: Springer Fachmedien Wiesbaden; 2020.

[16] Shollo A, Kautz K. Towards an understanding of business intelligence. In: ACIS 2010 Proceedings; 2010. 86.

[17] Lönnqvist A, Pirttimäki V. The measurement of business intelligence. Inf Syst Manag. 2006;23(1):32–40.

[18] Clark TD, Jones MC, Armstrong CP. The dynamic structure of management support systems: theory development, research focus, and direction. MIS Q. 2007;31(3):579–615.

[19] Foley E, Guillemett MG. What is business intelligence? Int J Bus Intell Res. 2010;1(4):1–28.

[20] Diván MJ. Data-driven decision making. In: Khatri SK, Kapur RK, Rana AS, Sanjay PK (eds). 2017 International Conference on Infocom Technologies and Unmanned Systems (ICTUS). Los Alamitos: IEEE; 2017. p. 50–6.

[21] Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making. Big Data. 2013;1(1):51–9.

[22] Luthfi A, Janssen M. Open data for evidence-based decision-making: Data-driven government resulting in uncertainty and polarization. Int J Adv Sci Eng Inform Technol. 2019;9(3):1071–8.

[23] Marbán O, Mariscal G, Segovia J. A data mining & knowledge discovery process model. In: Ponce J, Karahoca A (eds). Data Mining and Knowledge Discovery in Real Life Applications. London: IntechOpen; 2009. p. 1–16.

[24] Komenda M. Towards a Framework for Medical Curriculum Mapping [Ph.D. Thesis]. Brno: Masaryk University, Faculty of Informatics; 2016. Available from: https://is.muni.cz/th/hcl4g/.

[25] Ball A. Review of Data Management Lifecycle Models. Bath: University of Bath; 2012.

OPENING HEALTH DATA

Martin Komenda, Boris Turek, Michal Vičar, Andrea Pokorná, Ladislav Dušek

METHODOLOGICAL BACKGROUND IN HEALTH DATA SPACE

Data processing and subsequent visualisation are integral parts of this publication. The field of health data is fundamentally complicated concerning the sensitivity of published information and personal data protection. It is essential to conceptually address and highlight what data will be published (the content), how (the appropriate format and interface), and to whom (target groups). A key role is played by the electronic health record (EHR), which has not only made patients' medical information easier to read and available from almost any location in the world, but also changed the format of health records, and thus changed health care [1]. Individual datasets must be adopted as an official and guaranteed source for outputs of third parties, including public authorities, non-governmental organisations, scientists, and online news portals [2]. Thus, most chapters of this book are working with the concept of open data. From a general perspective of working with data, it is elementary to realise that there are several modes or approaches to working with health data. In this specific domain of healthcare, there are logically very specific cases that need to be systematically dealt with in accordance with current legislation and data protection. It is also necessary to mention the importance and significance of the data's origin (sourcing or collection) in an ethical context. In academic settings, there are often various guidelines or internal regulations in faculties that describe the moral code of a scientist or researcher. The main objective is to define clear and transparent rules for handling, publishing and archiving data for the retrospective validation of the results obtained. This chapter introduces this issue with the aim of introducing the concept of open data and the fact that historically there has been, and often will continue to be, a completely individual consideration of each dataset. This is from all relevant perspectives, such as methodological, technical, analytical and legal.

Sharing datasets, preferably in open data format, provides a systematic way to make selected datasets available for further manual or machine processing in a uniform and technically well-defined form (= dataset ready for "re-use"). The development of the open data domain, together with the operation of the

National Catalogue of Open Data[1], was for many years coordinated by the Ministry of the Interior of the Czech Republic; today, this role is secured by the newly established Digital and Information Agency (DIA). In the health information domain, a new working group under the leadership of the Ministry of Health of the Czech Republic is being set up to systematically coordinate activities related to the opening of health data (approval, creation, validation, and publication of datasets from the National Health Information System and other departmental information systems). The target group involves all stakeholders who aim to work with health data on a one-off or continuous basis (business, scientific and research infrastructures, academia, media and news, working groups of public authorities, regional governments, professional and lay public and others). Securing all the necessary input in the form of management and executive-level staff, a robust information technology infrastructure and methodological leadership are essential aspects of the systematic development of health data opening.

All regimes of the data provision from National Health Information System (NHIS) strictly require a certain degree of legislative regulation and must fully meet the criteria set for NHIS by the Czech legislation, particularly Act No. 372/2011 Coll., on Health Services and Conditions of Their Provision. In other words, publishing open data cannot be misinterpreted as the publication of primary records without any regulation and standardisation; the term "open data" does not necessarily describe prior database records (the data may be aggregated, statistically processed, etc.).

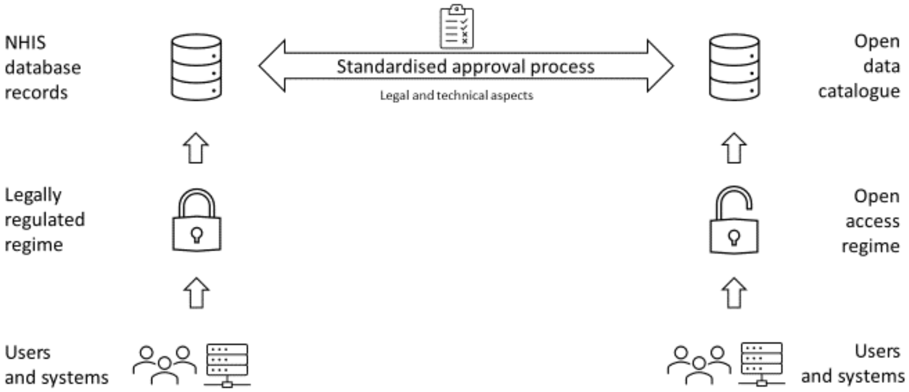


Figure 6: National Health Information System: Data provision schema

[1] https://data.gov.cz/datasets

The dataset design, preparation, and publishing should respect the algorithm of dataset preparation shown below, which always respects several principal rules:

1. Both direct and indirect identification of individual person cannot be possible.

2. Explicit identification of providers and other healthcare-related subjects must not be identifiable unless expressly stated by the law.

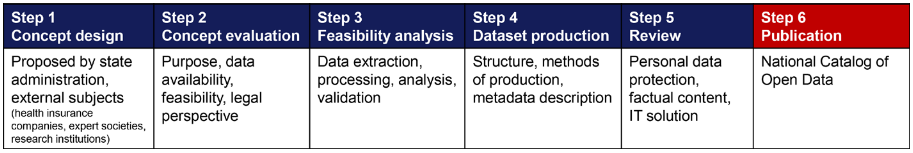
3. Secondary processing must lead to the pseudonymisation of the dataset.

4. The purpose of the dataset publication must correspond to the NHIS purpose.

5. The standardised approval process and publishing must be adhered to (Figure 7).

6. Each dataset has clearly defined authors and associated licensing is strongly recommended (if applicable).

The essential requirement of a comprehensive process, which this approach meets, is to ensure the necessary completeness, validity, and overall quality of data [2].



Step 1 Concept design	Step 2 Concept evaluation	Step 3 Feasibility analysis	Step 4 Dataset production	Step 5 Review	Step 6 Publication
Proposed by state administration, external subjects (health insurance companies, expert societies, research institutions)	Purpose, data availability, feasibility, legal perspective	Data extraction, processing, analysis, validation	Structure, methods of production, metadata description	Personal data protection, factual content, IT solution	National Catalog of Open Data

Figure 7: A chart of dataset production and publication

A universal and comprehensive methodological description of designing, preparing, validating, and publishing datasets is essential information with which the wider open data community in healthcare must be familiar. As the case of NHIS open data sets, the final output is always a work of authorship dataset (a value of the main idea introducing dataset structure, meaning and purpose), including complete disclaimers and a clear definition of licensing rules (if applicable) for appropriate citation in professional publications and other results.

An integral part of this is the categorisation of the individual datasets of the National Health Information System (NHIS). This categorisation, together with the explicit statutory purpose and level of aggregation, determines the possible mode of “openness” of particular subsystems of the NHIS as follows:

- systems not public by law,
- systems accessible only to legally defined readers/editors,
- sources of reference statistical data for the identified purpose and for the identified applicants,
- sources of published statistical data in the form of open datasets,
- sources open in primary “open data” mode.

When preparing datasets, it is recommended to follow the dataset creation scheme (Figure 7), where it is always crucial to respect the following rules: to comprehensively grasp the complex issue of data opening in Czech healthcare system, following three categories for data handling are proposed, defining different approaches according to the content of the information to be published.

1. Freely available primary data (very rarely used); examples may include service providers as defined in the respective acts, machines, swimming pools, chemical substances or drugs, and their primary characteristics.

2. Primary data publishable after necessary processing (used most frequently).

3. Data requiring reference interpretation – reference statistics, which means “a comment or summary to data output given by an expert in the field, usually given to data and values that have a more sophisticated background and require careful interpretation concerning objective uncertainties” [2].

The individual case studies mentioned in this book mainly refer to either open datasets or datasets that meet all technical aspects of open data except publication in a local or national catalogue (those datasets are not open but somehow FAIR). The book is intended to serve, among other things, as study material. However, this does not automatically mean that many samples or pilot data must be automatically published in an open data catalogue. However, the openness and availability of the data without access restrictions remain, and any user can freely work/re-use (with) the data referenced in the book.

TARGET GROUPS AND COMMUNICATION

An important part of the processing of open data in healthcare is connected with a proper definition of human resources and stakeholders dealing with healthcare open data. A universal and comprehensive methodological description of the process of design, preparation, validation and publication of data is essential information with which the broad community (composed of several general target groups) dealing with open data in healthcare must be familiar to be able to reuse the datasets. Ensuring clear and consistent communication about open data in healthcare is essential to a functional system. User profiling should consider not only communication goals but also information, health and data literacy. The outcome of such thinking is a division into three basic levels, which may overlap in preference of output format (Figure 8). Clarity, level of detail and, above all, a guarantee of validity and correct interpretation are essential attributes in terms of minimum requirements and demands. Such a concept requires very close collaboration between experts across the necessary expertise (theoretical physician or clinician, member of an expert medical society, healthcare expert, computer scientist, analytical guarantor, guarantor of data visualisation, open data guarantor). Depending on the format of the output, these people are actively involved in the actual process of communicating with the selected target group. In particular, data explorers and data experts have an excellent opportunity to disseminate the results correctly among stakeholders who have the mandate to reach the general lay public, i.e. data novices.

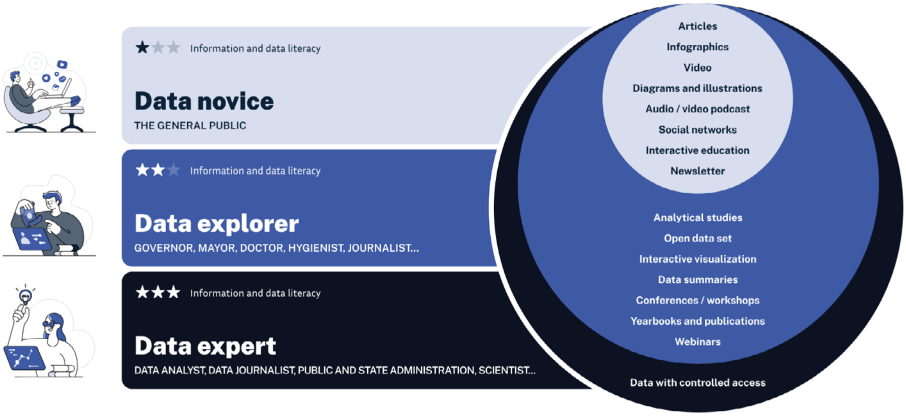

★☆☆ Information and data literacy
Data novice
THE GENERAL PUBLIC
★★☆ Information and data literacy
Data explorer
GOVERNOR, MAYOR, DOCTOR, HYGIENIST, JOURNALIST...
★★★ Information and data literacy
Data expert
DATA ANALYST, DATA JOURNALIST, PUBLIC AND STATE ADMINISTRATION, SCIENTIST...
Articles
Infographics
Video
Diagrams and illustrations
Audio / video podcast
Social networks
Interactive education
Newsletter
Analytical studies
Open data set
Interactive visualization
Data summaries
Conferences / workshops
Yearbooks and publications
Webinars
Data with controlled access

Figure 8: Communication of health data

Similarly important is the communication aspects towards the identified target groups. The lay public, together with the informed and professional public, must be able to be informed about the current status, the publication plan and the overall vision for open data within the specific healthcare domain, if needed. Communication tools that will be continuously used not only to communicate the current status but also to gather suggestions, requests and feedback, include:

- National and regional conferences and webinars
- Discussion panels and educational seminars
- Individual consultations
- National Health Information Portal[2] as a source of guaranteed and verified information for the general public
- Online platform data.nzis.cz[3] for up-to-date information and news
- National Catalogue of Open Data[4]
- Social media (Facebook, Twitter, Instagram)
- Press conferences and official opinions of the Ministry of Health and the Institute of Health Information and Statistics of the Czech Republic or other public sector authorities
- Specialist publications and dedicated web portals

Systematic communication across the professional community at the international level is also an integral part of this effort; the aim of this communication is to share and transfer experiences, especially directly in the specific domain of open data in healthcare. Table 2 below describes the idea defining the three focus/target groups of (open) data professionals, their level of information processing, data knowledge and skills, and the main objectives underpinned by key motivating factors for each group.

[2] https://nzip.cz

[3] https://data.nzis.cz

[4] https://data.gov.cz/datov%C3%A9-sady

Table 2: Description of target groups in terms of health data communication

Domain	Data novice	Data explorer	Data expert
Information and data literacy	The individual has basic knowledge of using information technologies. Can recognise, collect and share information in a digital environment. Understands the basic principles of assessing the credibility of sources. He/she occasionally searches for information in areas of his own interest.	The individual has more advanced knowledge in the use of information technologies. He/she is able to work with data, analyse it and draw conclusions from it. Understands the basic principles of data processing, such as collecting, organising, analysing and visualising data. Has a basic understanding of data analysis tools and understands the importance of critical thinking when interpreting results.	The individual has more advanced or expert knowledge in the use of information technologies. He/she is an expert in the field of information and data literacy. He/she is able to evaluate sources, process and analyse data, recognise sophisticated forms of misleading and disinformation. He/she has a deep understanding of data rights, ethical aspects of data processing and is able to comprehensively and critically evaluate information in the digital world.
Goals and motivation	The individual searches for information and data from the health sector on an ad hoc basis. He/she is more interested in the results than the journey. Articles and data overviews are accessed primarily via a search engine or via social networks. Charts and articles are shared via URL, in image form, video, or as PDF exports. To fulfil the need for information, he/she looks for mostly unstructured data. He/she approaches information more emotionally.	The individual often searches for information and data from the health sector for his/her work. Articles and data reports are accessed directly. Knows the basic sources of data reports and recognises their relevance. He/she approaches information critically and can draw conclusions from it and interpret it. He/she looks for mostly secondary sources of data. He/she wants to have all reports at hand as quickly as possible. Searches for structured and unstructured data to fulfill its information needs. He/she is intrinsically and explicitly motivated to education in data issues. He/she approaches information critically and ethically.	He/she searches for information and data from the health sector every day for his work. The data are accessed directly. Knows most sources of data reports and recognises their relevance. He /she approaches information critically and is able to evaluate, process, analyse, identify trends and subsequently interpret them. He/she looks for mostly primary data sources. He/she wants all reports to be complete and easily navigated (filtering). To fulfil the need for information, he/she mostly searches for structured data. There is a need for as much input data as possible to work with data. He/she is intrinsically and explicitly motivated to educate himself/herself in data issues to actively contribute to his/her know-how (professional publications, annual reports). He/she approaches information critically, ethically and legally.
Device usage	20 % laptop 80 % mobile	50 % laptop 50 % mobile	80 % laptop 20 % mobile
Characteristics	Fragile Low dissemination rate Risk of unintentional misinterpretation	Influenceable High dissemination rate Risk of intentional misinterpretation	Resistant and stable Medium dissemination rate Low risk of misinterpretation High level of distrust

The key attributes listed in Table 2 were essential characteristics in designing the open data communication matrix in healthcare. They helped define three main target groups based on information literacy, general goals and overall motivation for delivery and distributing the information, typical device usage habits, and ability to form conclusions and further dissemination.

REFERENCES

[1] Evans RS. Electronic Health Records: Then, Now, and in the Future. Yearb Med Inform. 2016;Suppl 1(Suppl 1):S48–S61.

[2] Komenda M, Jarkovský J, Klimeš D, et al. Sharing datasets of the COVID-19 epidemic in the Czech Republic. PLoS One. 2022;17(4):e0267397.

PUBLIC (HEALTH) SECTOR AND ACADEMIA

Matej Antol, Michal Růžička, Luděk Matyska, Jiří Marek

OPEN SCIENCE AND EOSC: RESEARCH DATA IN EVERYDAY ACADEMIC LIFE

INFORMATION SOCIETY IN ACADEMIA[1]

"Open Science is a system change allowing for better science through open and collaborative ways of producing and sharing knowledge and data, as early as possible in the research process, and for communicating and sharing results." European Commission [1]

Following many years of preparation and evolution, Open Science is becoming a standard of scientific life[2]. Researchers worldwide experience the effects of opening research and accessing its outcomes in their day-to-day work. The current topics include Open Access[3], increased sharing of research data (FAIR Data = Findability, Accessibility, Interoperability, and Reuse [2,3]), collaboration and other components of the scientific process – from open methodologies, peer review, software and tools to new methods of research quality assessment [4]. At the same time, the findings of publicly funded research are becoming interconnected with society (Citizen Science). Open Science represents a new, modern way to implement research and open access to scientific knowledge through digital technologies and tools enabling advanced cooperation.

To be "an internationally recognised research university with excellent higher education approach, there is a need for academic institutions to set trends in fulfilling all the roles of a university."[4] As such, these progressive institutions cannot stay aloof from the trends transforming global science; they must reflect and actively engage with them. It is no longer sufficient to see Open Science as an external development that the universities merely monitor as a modern trend. Open Science is a reality and provides an advanced research framework, where

[1] Text adopted from the Preamble of Open Science Strategy MU 2022–2028

[2] For more details see e.g. https://council.science/current/news/unesco-science-commission-adopts-open-science-recommendation/

[3] An open access to scientific publications in the electronic form without limitations on their use.

[4] One of the visions in the Masaryk University Strategic Plan 2021–2028.

a research institution behaves by the Open Science principles. Open Science is an advanced environment for managing and disclosing the university's findings, an instrument for scientific diplomacy and a synergistic complement to communication, popularisation and transfer of knowledge and technologies. It is also important for social acceptance and perception of science and research as an integral part of societal responsibility and development.

As mentioned in the Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC): *"The digital age, the most recent stage in an evolving continuum of ways in which technology has supported science, presents an opportunity to improve the conduct of research in multiple directions, including with regard to openness, speed of access to scientific results, reproducibility and multi-disciplinarity. This should result in better science, increased trust in science, and an improved ability to meet global challenges. However, this potential will only be realised if research infrastructures evolve to allow scientists to exploit, in an easy-to-use and integrated environment, the (vast amounts of) relevant data being produced."* [5]

The focus should be on science as such and its excellence; Open Science and research infrastructures [6] have to be understood as indispensable support tools for its academic environment, and academic institutions should advocate that the Open Science principles to be a standard not only in the Czech Republic but also in the broader European and social context.

FROM OPEN ACCESS TO OPEN/FAIR DATA

The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanitiesp [7] is currently the basic document for implementing Open Access. It builds on two previous initiatives: *Budapest Open Access Initiative* [8] and *Bethesda Statement on Open Access Publishing* [9]. These three documents are collectively referred to as the *Open Access BBB initiatives*.

The Berlin Declaration was adopted in October 2003 at a conference organised by the Max Planck Society in Berlin, was signed by 497 institutions worldwide, and has been widely adopted and matured to the de facto standard requirement of the European project calls [1].

However, the scientific publications alone are insufficient in data-driven 21st century science. There is an increased need to make available also the data that stay behind the publication – they are critical for reproducibility and thus trust in the published findings. But even that is not sufficient, as the data themselves become a valuable resource, opening opportunities for more research. Strong motivation for making research data available as a proof of the soundness of the published finding is the so-called *Replication crisis* [10]. Another is the principle of making the results of public money-funded research freely available. In

Europe, the Open/FAIR data requirements [3] started to emerge in the project calls [11] and currently are the standard requirement in projects [12].

OPEN/FAIR DATA, EOSC AND OPEN SCIENCE TODAY

Research data usually refers to the factual information or evidence generated during research (research project). Another definition emphasises the use, considering it as "data used for research" [13]. With the emphasis on data sharing and open access to the research data, the scientific publishers widely adopted the concept of data supporting a journal article. The *"data, including associated metadata, needed to validate the results presented in scientific publications"* are nowadays a standard component of journal articles, giving readers at least a theoretical opportunity to check the data used. However, this should be considered as the first step only, as this degrades the research data to a secondary position of "supporting the publication" only.

The interest is shifting towards **individual data sets** (*"other data, including associated metadata connected to a particular research project, as specified in the data management plan"*) [14], made available directly, without direct association to any specific publication. The data are becoming first-level citizen in the research space, having identity independent from any other scientific artifact. The datasets are assigned permanent identifiers and can be cited as any other scientific publication.

The **Open Research Data** are governed by the same principles as open data of the public sector explained in the previous chapter. We can consider the data supporting open-access articles as a classical representative of this concept data. However, the Open Research Data concept is too generic and cannot be applied to all research data (e.g. because they are protected by laws or represent valuable industrial property). Therefore, the FAIR data concept has emerged as a usable compromise between full openness and the restrictions that can accompany research data. In general, *"the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals"* [2,14].

By adhering to these features, FAIR data aims to enhance the value and impact of research data by enabling its widespread discovery, access, integration, and reuse, thereby fostering collaboration, reproducibility, and innovation in the scientific community.

European Open Science Cloud (EOSC) is an ambitious initiative by the European Union (EU) to create a digital platform that enables seamless access to research data, services, and infrastructure across Europe. The main goal of EOSC is to promote open and collaborative research, making scientific knowledge more accessible, transparent, and reproducible. The EOSC origins can be traced back to the realisation that scientific research has become increasingly data-intensive and requires advanced computational tools and infrastructure. The European Commission launched the European Cloud Initiative in 2016. It led to the establishment of EOSC as a long-term vision for the digital transformation of science and research in Europe, making it a unique horizontal partnership within the current HEU framework. By providing a federated and trusted environment, EOSC aims to enable researchers to access and share data, tools, and services across borders, disciplines, and sectors. The European Open Science Cloud (EOSC) and FAIR data share a close relationship, as both initiatives align their goals and principles regarding open and accessible data in the research community.

EOSC and FAIR data share a very close relationship, as both concepts are aligned in their goals and principles regarding open and accessible data availability in the research community. EOSC is a key instrument to deliver Open/FAIR Data in everyday practice in academia. It is supposed to "deliver Europe's contribution to the realisation of scientists', and science's, potential in the digital age, enhancing Europe's leadership position in exploiting digital capabilities at the service of science." 30 The three main objectives of EOSC are visualised in Figure 9. They are framing the current research data sharing landscape within Europe and together with the European Data Strategy and other "data spaces" they create the new generation of European digital infrastructure [5].

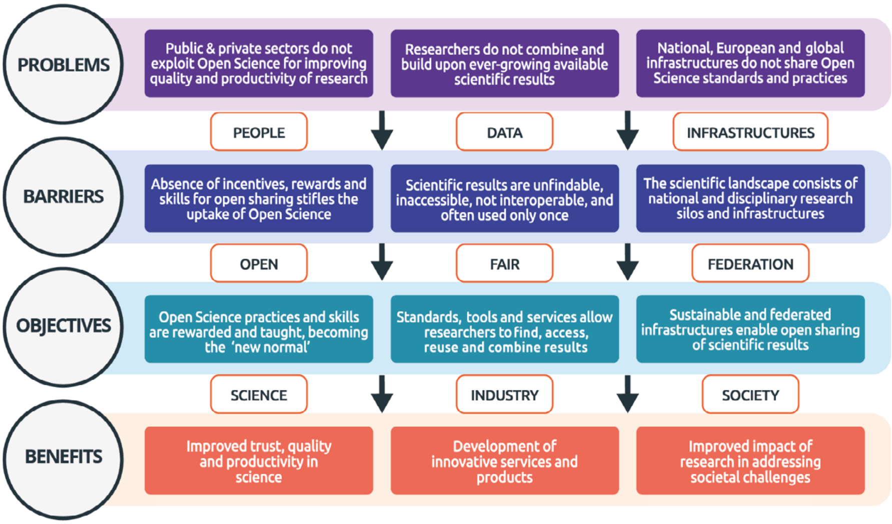

PROBLEMS
Public & private sectors do not exploit Open Science for improving quality and productivity of research
Researchers do not combine and build upon ever-growing available scientific results
National, European and global infrastructures do not share Open Science standards and practices
PEOPLE
DATA
INFRASTRUCTURES
BARRIERS
Absence of incentives, rewards and skills for open sharing stifles the uptake of Open Science
Scientific results are unfindable, inaccessible, not interoperable, and often used only once
The scientific landscape consists of national and disciplinary research silos and infrastructures
OPEN
FAIR
FEDERATION
OBJECTIVES
Open Science practices and skills are rewarded and taught, becoming the 'new normal'
Standards, tools and services allow researchers to find, access, reuse and combine results
Sustainable and federated infrastructures enable open sharing of scientific results
SCIENCE
INDUSTRY
SOCIETY
BENEFITS
Improved trust, quality and productivity in science
Development of innovative services and products
Improved impact of research in addressing societal challenges

Figure 9: European Open Science Cloud Objectives Tree

The EOSC objectives show EOSC as an initiative that aims to federate and make available (although not necessarily fully open) research data across disciplines, sectors, and countries. To achieve this, EOSC promotes the adoption of open and FAIR data principles to enhance the accessibility and reusability of research data. While creating federated infrastructures for seamless access and collaboration, EOSC also seeks to support the development and provision of a wide range of services and tools for effective data management, analysis and sharing. It focuses on ensuring interoperability and seamless access to research data and services while respecting privacy, security, and ethical considerations. Last but not least, EOSC fosters collaboration and networking among researchers and institutions across Europe and beyond. EOSC can be considered the primary tool to foster an open science culture, supporting the transition towards openness and broad multidisciplinary research collaboration.

Research data are used for more than just research itself. The use of research data for initiatives like Citizen Science or Science Diplomacy is getting more and more attention, and the Policymakers are looking towards research and objective scientific information as one of the tools in the new post-factum era as was seen during the COVID-19 pandemic. The following chapters will dwell more in detail on concrete sectors of open science/research data management, that are key in today's world of data analytics, interactive visualisations, and web applications as a basic infrastructure to effectively support decision-making on various policy levels.

E-INFRASTRUCTURES AS A KEY STAKEHOLDER IN THE RESEARCH DIGITAL AGE

RESEARCH INFRASTRUCTURES IN THE EUROPEAN CONTEXT

Scientific communities thrive within a dynamic ecosystem centred around universities and research infrastructures. Universities provide an ideal setting for scientists to concentrate on their research and teaching. In contrast, research infrastructures establish networks of scientists within specific fields across countries, fostering cooperation and enabling domain-specific innovation. These infrastructures facilitate collaboration and distribution of services, irrespective of national or university affiliations.

International collaboration is one of the crucial aspects of research infrastructure operations. In Europe, the European Strategy Forum on Research

Infrastructures (ESFRI) identifies over 60 pan-European research infrastructures[5]. Additionally, European National Research and Education Networks (NRENs) have emerged as another supportive environment for research, education, and innovation, united under an association called GÉANT. These infrastructures, along with other institutions that bring together researchers from various scientific domains, continually develop and maintain robust environments for data storage, processing, and analysis—fundamental to modern research endeavours.

The European Open Science Cloud (EOSC) initiative has recently strengthened the cooperation among these infrastructures and institutions. EOSC aims to provide a federated and open multidisciplinary environment, enabling European researchers, innovators, companies, and citizens to publish, discover, and reuse data, tools, and services for research, innovation, and educational purposes. The data within EOSC adhere to the FAIR principles, ensuring they are Findable, Accessible, Interoperable, and Reproducible. Beyond data management, EOSC also encompasses data processing, as the value of scientific data lies in their manipulation, analysis, and application to real-world problems.

The history of the development of research infrastructures, with their robust technological background for storing and processing scientific data, together with new, data-centric initiatives such as EOSC, and challenges stemming from globalised society relying more on data collection and analysis, together represent a fertile environment for new, ground-breaking discoveries across scientific domains.

RELEVANT INFRASTRUCTURES FOR THE MEDICAL SECTOR

The above-mentioned European Infrastructures recognised by ESFRI can be categorised into six thematic areas[6], two of which are most relevant to this book.

First, **data, computing & digital research infrastructures**, commonly known as e-infrastructures. The only already established European e-Infrastructure recognised by the ESFRI is PRACE (Partnership for Advanced Computing in Europe), with three newly emerging infrastructures (EBRAINS, SLICES, SoBigData) in the preparation phase. However, the field of digital support of research concentrated in RIs is complemented by European Grid Infrastructure (EGI), European High Performance Computing Joint Undertaking (EuroHPC JU), EUDAT, OpenAIRE and GÉANT, which itself coordinates the operation of National Research and Education Networks (NERNs). All these infrastructures and collaborations seek to develop and maintain digital services in terms of

[5] For more details see https://roadmap2021.esfri.eu

[6] See ibid

networking, computing, and data management and foster the emergence of Open Science practices.

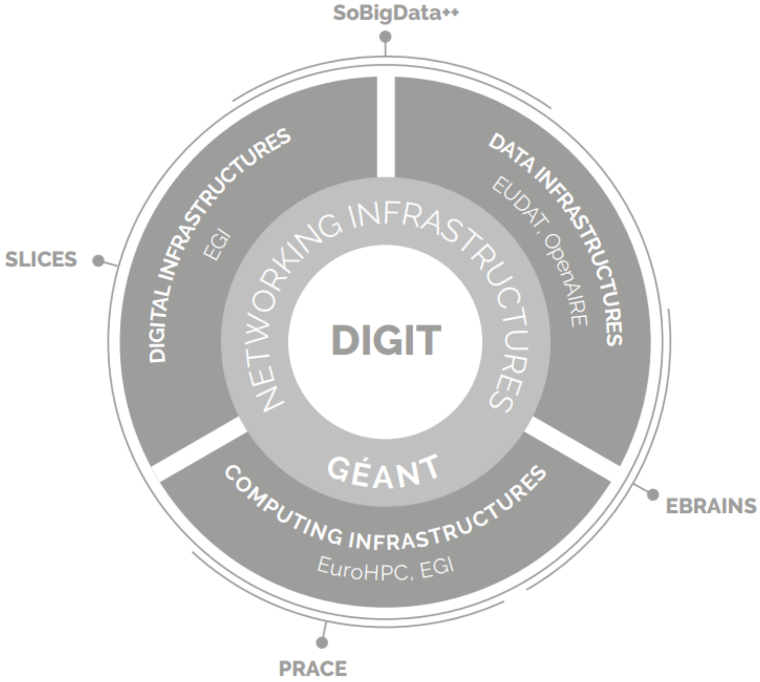


Figure 10: The landscape of the data, computing & digital research infrastructure domain

Second are infrastructures falling into groups labelled as **health & food**, specifically the health subdomain. Already existing infrastructures falling within this category are AnaEE, BBMRI ERIC, EATRIS ERIC, ECRIN ERIC, ELIXIR, EMBRC ERIC, ERINHA, EU-OPENSCREEN ERIC, Euro-BioImaging ERIC, INFRAFRONTIER, INSTRUCT ERIC and MIRRI, with four newly emerging RIs: EIRENE RI, EMPHASIS, IBISBA, METROFOOD-RI [15].

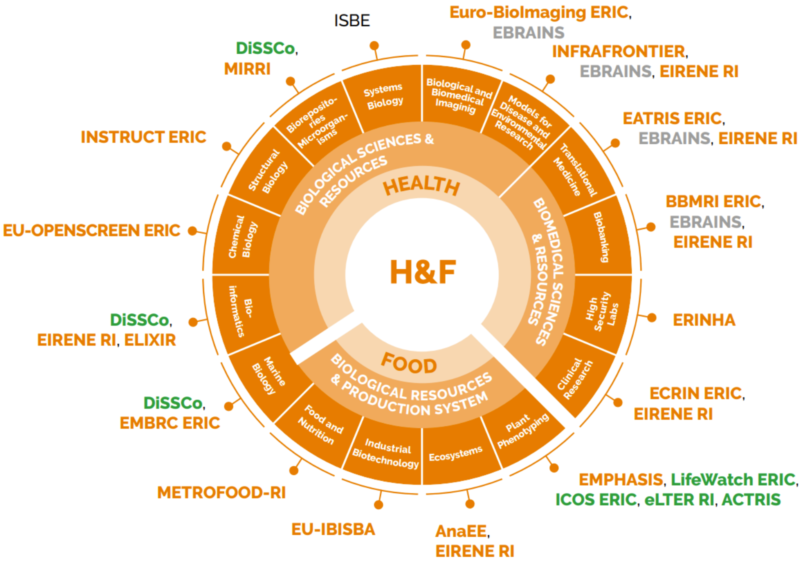


Figure 11: The landscape of the health & food domain

Within the health and food thematic area, the health subdomain encompasses many research areas, portfolio technologies, and related services. The dynamic nature of this domain is particularly noteworthy, especially considering the lessons learned during the SARS-CoV-2 pandemic. The response of research infrastructures (RIs) to the COVID-19 crisis is a remarkable demonstration of coordinated efforts in addressing an urgent problem.

The most prominent mission of RIs falling within this category is dealing with cancer and cardiovascular diseases, as they are responsible for an alarming number of total deaths in Europe. These RIs also seek to protect human health by increasing preparedness for and capability to respond to highly pathogenic infectious threats. Other strategic goals of these infrastructures are related to understanding plant performance in the context of health and famine challenges, increasing understanding of oceans, seas and inland waters, and many other related areas.

European research infrastructures continuously seek to support research via environments of services, tools and competencies required for domain-specific conduct of research. These requirements are increasingly more concentrated around data and various digital artefacts. With their inherent cross-sectional function, research infrastructures play a vital role in data management, from data storage, access, analyses, sharing and cooperation to data visualisation and publication for scientific communities and the general public.

SITUATION IN THE CZECH REPUBLIC

Following the clustering established at the European level, Roadmap of Large Research Infrastructures of the Czech Republic recognises six clusters of research infrastructures operated at and funded on the national level [16].

Ten research infrastructures fall within the health & food category on the national level, most of which also constitute a local node of European and pan-European infrastructures. These are the Bank of Clinical Specimens (BBMRI-CZ), the Czech Centre for Phenogenomics – (CO), the Czech Infrastructure for Integrative Structural Biology – (CIISB), the Czech National Node to the European Clinical Research Infrastructure Network – (CZECRIN), the National Research Infrastructure for Biological and Medical Imaging – (Czech-BioImaging), the National Infrastructure for Chemical Biology – (CZ-OPENSCREEN), the Czech National Node to the European Infrastructure for Translational Medicine – (EATRIS-CZ), the Czech National Infrastructure for Biological Data – (ELIXIR-CZ), the Infrastructure for Promoting Metrology in Food and Nutrition in the Czech Republic – (METROFOOD-CZ) and the National Center for Medical Genomic – (NCMG).

In contrast to the European landscape, where several e-infrastructures (e.g. GEANT, EGI.eu, EUDAT, PRACE, EuroHPC) co-exist, there is only one e-Infrastructure on the national level. e-INFRA CZ was created by merging three formerly independent infrastructures: CESNET (Czech NREN), CERIT-SC at Masaryk University, and supercomputing centre IT4Innovations at Technical University Ostrava. This merger provides a fertile ground for integrating the national infrastructure of resources and services, which has already proved to bring fruit in the Czech response to the European Open Science Cloud (EOSC) initiative on a national level. The Czech approach to the EOSC implementation is based on developing the National Data Infrastructure (NDI) with the National Repository Platform (NRP) in its heart. The e-INFRA CZ coordinates key research institutions across the national landscape to join forces in implementing EOSC and NDI.

The existence of a single national e-infrastructure guarantees an integration of the NDI and NRP into the already established environments focused on computing, networks, cybersecurity, and other topics related to IT in the research domain. Moreover, a single national e-infrastructure represents the best chance for the future continuation and development of the developed environment, services and practices, including those established within the Czech response to the EOSC initiative.

HOW E-INFRASTRUCTURES ARE CONNECTED TO THE DATA PROCESSING TOPIC

All of the latest progress in science and our understanding now relies on a steady increase of computational power dedicated to processing vast amounts of scientific data. Most recent and dramatic discoveries in various science fields were made due to advancements in the computing area. Optimisations in numerical methods and the remarkable growth of available computational power reflect many exceptional advances in research, such as climate modelling, protein folding, drug discovery, and energy research. It is fair to say that furthering our (human) theoretical understanding has transformed from "pen-and-paper" to designing simulations and modelling applying mathematical concepts using technologies offered by accomplishments within the computer science area.

Research infrastructures at both national and international levels aim to offer baseline data-related services and appliances to the scientific community. This is especially true for e-infrastructures, which, in contrast with domain-specific infrastructures, are focused on administering the raw power and generally applicable IT environments. Research infrastructures represent an indispensable conduit for modern research and play a crucial role in the national and international ecosystem related to the management and processing of research data.

DATA MANAGEMENT PLANNING AT RESEARCH INSTITUTIONS

DATA MANAGEMENT PLANNING AS A SKILL FOR EVERYDAY RESEARCH, TECHNICIAN AND STUDENT

Having standard requirements on projects to produce FAIR data, the researchers and project-support teams in their institutions are naturally motivated to consider their implementation. At the same time, there is a self-evident lack of these competencies across the academic institutions in the Czech Republic and the general research community.

In reaction to the lack of skills on the one hand and requirements from project funders on the other, Open Science support teams started to emerge in large universities in the Czech Republic[7]. At the centre of their focus lies aggregating

[7] Masaryk University (https://openscience.muni.cz/en), Charles University (https://openscience.cuni.cz/OSCIEN-1.html)

knowledge regarding project call requirements, best practices in data management, repositories available for various scientific disciplines for publishing data, and, very importantly, the transfer of relevant information and knowledge to the research community.

Their members started self-organising in so-called expert groups to improve skills and share knowledge among the Open-Science-support teams. Just to name the most prominent of these initiatives, informal Discord groups *AKVŠ-PS Open Science, datasteward.cz*, and *Data Stewardship Wizard* already successfully interconnected many data experts across the Czech Republic.

TOOLS FOR DATA MANAGEMENT PLANNING

One of the first concrete requirements encountered by the researchers while interacting with the research funding organisations is creating and keeping an up-to-date Data Management Plan (DMP). This requirement logically follows the intention of research funding organisations (RFOs) on handling research data as first-class-citizen research outputs: DMP should certify that the project team has a clear idea of how to handle the data from creating to processing, preserving, and finally sharing.

To support DMP preparations, a variety of tools was created. DMPonline[8] represents a tool following document templates from RFOs but transforming them to electronic form with guidance and helping texts from Open-Science-support teams of various institutions. More advanced tools were developed later. For example, ARGOS[9] follows a similar concept as DMPonline in transforming DMP templates into electronic forms in specialised information systems. However, it divides the description of a research project from the description of a dataset to allow common practices of using one dataset in multiple projects and multiple datasets in one project. ARGOS also benefits from integration with other relevant systems like OpenAIRE Research Graph[10] or Zenodo[11] to transfer data in standard forms. The possibility of systems and data integration is one of the motivation factors for using specialised information systems for DMP support instead of creating and sharing them as standard documents in common office formats.

The most advanced DMP-supporting tool, from my point of view, is Data Stewardship Wizard (DSW)[12]. In contrast to the previous tools, its primary intention is to support data management *planning* as a process, not only creating data management plans as documents. DSW builds on an expert system, defining

[8] https://dmponline.dcc.ac.uk/

[9] https://argos.openaire.eu/

[10] https://graph.openaire.eu/

[11] https://www.zenodo.org/

[12] https://ds-wizard.org/

the so-called knowledge model (KM) for a particular scientific discipline. The KM should describe data handling in the area, regardless of the specific DMP template. DMP documents are generated according to a particular DMP template on request for the information provided according to KM.

DSW is focused on modern implementation, supporting real-time collaborations of multiple persons and integration with other information systems providing rich API covering the tool's full functionality. The principle of the wizard highly improves the user experience: The user is asked only the necessary questions. More details are required by the system only if relevant to the current answers. Where suitable, the system is connected to vocabularies and ontologies to support machine-actionability on the DMP data. Metrics of fulfilling FAIR and best-practice requirements are defined in the system and indicated for every DMP to provide users feedback on the quality of the data management in the described project.

LESSONS LEARNED FROM DMP AT CZECH ACADEMIC INSTITUTIONS

Recent experience from Czech academic institutions shows a clear need for training research-data-management support staff like data stewards. There is a lack of mature description of the required competencies. Still, experience from the Open Science support departments of the Czech academic institutions shows the need for knowledge in data management planning, data handling and their specifics in various disciplines, specific IT skills (handling of large datasets and sensitive data, real-time data processing, support of AI computations and accelerations, design of modern applications using containerisation, ...) and expertise in data storage, discipline-specific repositories and metadata creation and management.

The first contact with the Open Science support team is usually initiated by the researchers not knowing how to handle FAIR data / DMP requirements in the project they are participating in or not knowing how to handle sensitive data in their project following the current best practice and project and law requirements.

Handling of sensitive data usually requires the use of dedicated infrastructure. Some of them are available or under construction in the Czech Republic (for example, CERIT-SC SensitiveCloud[13]). Handling DMP is usually easier – Open Science support teams can provide consultation by filling in the RFO DMP template and providing support tools like DSW instance.

[13] https://www.cerit-sc.cz/infrastructure-services/sensitivecloud

For example, at Masaryk University, we have our own instance of DSW, called DSW MUNI[14], that is integrated with an institutional information system that, besides other responsibilities, is used to manage research projects. This integration, made possible via full-featured DSW API, allows MU researchers and management to keep precise track of data management of every solved project by binding DSW DMPs with current research projects. On the other hand, our experience also shows that the researchers often make DMP preparations shortly before the deadline for their release for RFO. And in this case, they sometimes prefer filling in the RFO DMP template directly than filling in the full description of data management in their research in DSW. That indicates that the researchers are still thinking about DMPs as 'bureaucratic' documents for RFO that the process that can help them improve, especially in long-term, overall data handling in their projects.

Another problem often presented by the researchers is the lack of suitable data repositories to store research data for fulfilling the FAIR data requirements of RFO. Researchers from MU CEITEC MAFIL[15], specialising in human magnetic resonance imaging, are currently implementing support for producing FAIR data as the standard output of their laboratory for all their measurements. The investigation of suitable repositories for publishing their data indicates OpenNeuro as the best appropriate existing repository. However, it still does not fit their needs fully.

For example, experts from the Masaryk University CSIRT team[16] were even less successful when looking for a suitable repository for publishing their dataset Encrypted Web Traffic: Event Logs and Packet Traces [17]. As there was a lack of an appropriate topic repository, they tried to use Zenodo. However, the size of the dataset – 270 GiB – was far beyond the available limits of Zenodo. We decided to take advantage of Masaryk University's ability to assign Digital Object Identifier (DOI) [18] persistent identifiers for datasets on our own: The dataset was assigned DOI and stored in the university's cloud storage, making DOI resolve to read-only share link. This is considered a short-term solution; in the long term, we expect to move the dataset to a national data repository (a pilot version of the Czech data repository is currently under development[17]) or use a newly emerged suitable topic repository. As DOI is a persistent identifier, the users will not be affected by the change of the dataset's location as we will guarantee the DOI will always resolve to the current dataset's location. This demonstrates the importance of one of the FAIR data requirements – persistent identifiers hiding users from the complexity of physical data storage and providing them with the

[14] https://dsw.muni.cz/

[15] https://mafil.ceitec.cz/en/

[16] https://csirt.muni.cz/?lang=en

[17] https://data.narodni-repozitar.cz/.

functionality identifier of the dataset itself, guaranteeing access to the current dataset location.

Yet another possibility how to solve the lack of a suitable data repository is to build your topical repository. Even though the best practice is to use an established repository, it is still a valid approach if none exists, for example, in the case of emerging new disciplines. We also have experience with this approach: Our colleagues from the Faculty of Science of Masaryk University operate the World Spider Trait database[18]. To enhance the FAIRness of the contents of this repository, their operators integrated with our institutional DataCite account for automated assigning of DOIs for records in this database[19].

REFERENCES

[1] European Commission. Open Science [Internet]. European Commission; 2020 [cited 16 Jul 2023]. Available from: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en.

[2] Wilkinson M, Dumontier M, Aalbersberg I, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018.

[3] GO FAIR Initiative. FAIR Principles [Internet]. GO FAIR Initiative; 2023 [cited 21 Jul 2023]. Available from: https://www.go-fair.org/fair-principles/.

[4] European Commission, Directorate-General for Research and Innovation. Towards a Reform of the Research Assessment System: Scoping Report [Internet]. Luxembourg: Publications Office of the European Union; 2021. Available from: https://data.europa.eu/doi/10.2777/707440.

[5] European Open Science Cloud. Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud [Internet]. 2022 [cited 21 Jul 2023]. Available from: https://eosc.eu/sites/default/files/SRIA%201.1%20final.pdf.

[6] Ministry of Education and Research of Sweden. Lund Declaration on Maximising the Benefits of Research Data [Internet]. 20 Jun 2023 [cited 21 Jul 2023]. Available from: https://www.government.se/contentassets/69ab5c1102d5435ea24ce18da837b17d/lund-declaration-on-maximising-the-benefits--of-research-data.pdf.

[18] https://spidertraits.sci.muni.cz/

[19] https://commons.datacite.org/repositories/eqjcw9y

[7] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Open Access Initiatives of the Max Planck Society [Internet]. Berlin: Max-Planck-Gesellschaft; 22 Oct 2003 [cited 21 Jul 2023]. Available from: https://openaccess.mpg.de/Berlin-Declaration.

[8] Budapest Open Access Initiative: The original Declaration and guidelines to make research free and available to anyone with internet access and promote advances in the sciences, medicine, and health [Internet]. Budapest: Budapest Open Access Initiative; 2023 [cited 21 Jul 2023]. Available from: https://www.budapestopenaccessinitiative.org/.

[9] Bethesda Statement on Open Access Publishing. Digital Access to Scholarship at Harvard [Internet]. Harvard: Harvard University; 20 Jun 2003 [cited 21 Jul 2023]. Available from: https://dash.harvard.edu/bitstream/handle/1/4725199/Suber_bethesda.htm.

[10] Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016;533:452–4.

[11] Open Research Data (ORD) – the uptake in Horizon 2020 [Internet]. European Commission, Directorate-General for Research and Innovation; 2018 [cited 21 Jul 2023]. Available from: http://data.europa.eu/88u/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020.

[12] Horizon Europe (HORIZON): Programme Guide [Internet]. Version 3.0. Europe: European Commission; 1 April 2023 [cited 16 Jun 2023]. Available from: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf.

[13] Koščík M, Polčák R, Myška M, Harašta J. Research data and their legal context. In: Koščík M, Polčák R, Myška M, Harašta J (eds). Research Data and Research Databases. Legal Framework for the Processing and Sharing of Scientific Findings. Prague: Wolters Kluwer; 2018. p. 17–24.

[14] H2020 Programme AGA – Annotated Model Grant Agreement. Version 5.2 [Internet]. European Commission; 26 June 2019 [cited 14 Jul 2023]. Available from: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf.

[15] ESFRI Roadmap 2021 – Strategy Report on Research Infrastructures [Internet]. European Strategy Forum on Research Infrastructures; 2021 [cited 14 Jul 2023]. Available from: https://roadmap2021.esfri.eu/media/1295/esfri-roadmap-2021.pdf.

[16] Roadmap of Large Research Infrastructures of the Czech Republic for the years 2016–2022 [Internet]. Ministry of Education, Youth and Sports of the Czech Republic; 2019 [cited 14 Jul 2023]. Available from: https://www.vyzkumne-infrastruktury.cz/wp-content/uploads/2019/11/Aktualizace-Cestovn%C3%AD-mapy-2019_en.pdf.

[17] Špaček S, Velan P, Čeleda P, Tovarňák D. Encrypted Web traffic dataset: Event logs and packet traces. Data Brief. 2022;42:108188.

[18] What is a DOI. DOI Foundation [Internet]. DOI Foundation; 2022 [cited 1 Jul 2023]. Available from: https://www.doi.org/the-identifier/what-is-a-doi/.

SECTION B

MEDICAL AND HEALTHCARE EDUCATION

STAKEHOLDER OPINION

Andrea Pokorná
VICE-DEAN FOR HEALTHCARE STUDY PROGRAMMES AND INFORMATION TECHNOLOGY
Masaryk University, Faculty of Medicine, Brno, Czech Republic

Data and data sources (i.e. database systems of various types) are the basis of effective management across all the disciplines in our lives. In essence, data in healthcare education provide a scientific and systematic approach to enhancing the training of future healthcare professionals. It enables educators to adapt their methods, optimise curricula, and ensure that students are well prepared to provide safe and effective patient care in a rapidly evolving healthcare landscape. In the context of healthcare education, data refer to information that is collected, recorded, and analysed to improve the quality of education within the healthcare sector. These data can come from various sources, such as student performance assessments, course evaluations, educational technologies, research studies, etc.

The primary goal of using data in healthcare education is to make informed decisions, enhance teaching and learning processes, and ultimately improve the overall quality of healthcare professionals' education and training. From the position of a person currently working as part of the management of the Faculty of Medicine, I perceive the importance of the use and usability of data as significant for several reasons. First, from the point of view of the need to use the data to manage study agendas, to make them accessible and transparent (for assessment and feedback, curriculum development, personalised learning, quality improvement, identifying trends in evidence-based education as well as ethical consideration) - here from the consumer and creator of database models or at least the requester of the data structure and further usability requirements. Second, it is also essential to convey information about the necessity of adequate data input and processing in the professional undergraduate training of students. In particular, students who undertake clinical placements and are involved in patient care must build high accountability for correctly recorded and monitored data, subsequently enabling appropriate patient care and treatment. Students must be able to manage data records in clinical practice; modern healthcare systems often rely on specialised software and databases to handle electronic health records (EHRs) due to their complexity and security requirements (including data structures, memory management, sorting and filtering and error handling). And third, it is equally important for students and future medical and non-medical graduates to understand the importance of data processing, presentation, visualisation, and subsequent interpretation. They will be the ones who subsequently use the data not only in their professional life and careers but can also influence the general public's views on the data provided. In this way, they help strengthen care recipients' health literacy.

Data in health literacy for the lay population are about presenting health-related information in a way that is easy to understand and use. Clear, accurate, and accessible data empower individuals to take charge of their health, make informed decisions, and engage effectively with healthcare providers and systems. All the above-mentioned components are essential; thus, I am happy that this book includes consistent and comprehensive material in this field.

Daniel Schwarz

HEAD OF THE MEFANET COORDINATING COUNCIL
Masaryk University, Faculty of Medicine, Brno, Czech Republic

I'm delighted to introduce you to this engaging book: "Data-driven decision-making in medical education and healthcare: Data rulezzz!" This book is a game-changer, coming from smart folks at Masaryk University's Faculty of Medicine and the Institute of Health Information and Statistics.

So, why should you read it? Simple: it shows just how important data is in healthcare and education. The book is divided into 18 real-life stories, each one showing the awesome things that happen when you mix healthcare, education and data science. These aren't just random thoughts; they follow a proven method known as CRISP-DM. It's like a roadmap that data scientists use, and it makes sure all these different stories fit together like pieces in a puzzle.

One cool part of the book is how it focuses on visual analytics—like charts and graphs—to make complicated data easy to understand. Let's face it: a bunch of numbers on a page can be hard to wrap your head around. Visuals help us see the bigger picture and make smarter choices.

A big shoutout goes to MEFANET. It's a network of medical schools in the Czech Republic and Slovakia. They're a big deal, not just locally but across Europe. Many stories in this book were born out of projects that MEFANET had a hand in. So, they've played a huge role in pushing healthcare education into the future.

To sum it up, if you're into medicine, education, or data—and especially if you're into all three—this book is for you. It's full of lessons, challenges, and inspiration. Don't miss it!

Petr Štourač

VICE-DEAN FOR DEVELOPMENT AND STUDIES IN GENERAL MEDICINE

Masaryk University, Faculty of Medicine, Brno, Czech Republic

The book "Data-driven decision-making in medical education and healthcare" by the author team led by Dr. Martin Komenda is a unique act not only in the Czech environment. Using the examples of eighteen case studies, it presents real-world problems of using data processing for decision-making in healthcare or education. It must be noticed that this is an insight into more than ten years of systematic activity of the main author and his department at the Faculty of Medicine of Masaryk University, on both national and international levels. In the section dedicated to medical education, there are eight case studies, which by their nature form a unique, temporally and logically connected whole, which ultimately led to a tangible qualitative shift in the field of medical education at the Faculty of Medicine of Masaryk University; this was achieved by the currently valid "SIMU" accreditation focused on the implementation of simulations methods and techniques in which the outputs of the projects were primarily used. The OPTIMED project, which mapped and made available the curriculum of the General Medicine study programme for systematic evaluation, was the cornerstone of further development. The EVAMED project, which evaluated the outputs of this project in two phases, was logically and temporally connected and brought additional value to this endeavour. The evaluation of the textual similarity between the medical curriculum and the virtual patients of AKUTNE.CZ was interesting.

The great news is that other projects with significant international overlap came from the above-mentioned projects, such as MEDCIN, BCIME and others. Furthermore, the outputs of these projects provided a robust basis for strategic change in the educational curriculum that would not have been possible otherwise. The Medical Simulation Centre at the Faculty of Medicine of Masaryk University project, which is proof of the above fact, benefited in many areas precisely from the data and evaluation base provided by the individual projects described in the book. One of the chapters is also devoted to implementing and developing the method of objective structured clinical examination (OSCE), which brought an essential objective element to medical education, especially to evaluating its effectiveness. Together with the tools for collecting and evaluating the obtained data within the integration platform SIMUportfolio, it creates a powerful tool for deciding on further direction in the curriculum development in the clinical part of the study at the Faculty of Medicine of Masaryk University.

I wish the book, which is highly inspiring in many ways, to be a successful and valuable teaching aid in both study programmes. It has the potential and quality for it.

SECTION B

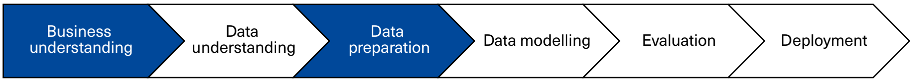
01

OUTCOME-BASED CURRICULUM DEVELOPMENT AND OVERVIEW USING AN INNOVATIVE ONLINE PLATFORM

Martin Komenda, Julie Dobrovolná, Jaroslav Štěrba, Martin Bareš, Jiří Mayer, Matěj Karolyi, Ladislav Dušek

The case study was supported the following grant project: OPTIMED – optimised medical education: horizontal and vertical connections, innovations and efficiency in practice – reg. no. CZ.1.07/2.2.00/28.0042, which was funded by the European Social Fund and the national (state) budget of the Czech Republic.

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2012–2014
Keywords	Outcome-based approach, curriculum management system, interactive visualisation, medical and healthcare education
Research question	How have parametric and structured database approaches deal with curriculum development, harmonisation, and reporting?
Type of result	Complex web application
Level of data processing	Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

The merit of this case study meets up-to-date research issues. It addresses the adoption of agile software development principles and the design, development and implementation of an original web-based solution supporting decision-making in a very comprehensive process of curriculum harmonisation. In 2012, OPTIMED (an acronym for Optimization of Medical Education) was a new project at the Faculty of Medicine of Masaryk University (MED MUNI), which focused mainly on medical curriculum management in the global context of curriculum development, overview and mapping in higher education institutions (HEI) domain. There was no universally suitable solution based on an approved pedagogical approach which would parametrically describe, effectively manage, and visualise an HEI's curriculum and all related information within one system. This was the primary motivation for designing an original web-based platform for the innovation and harmonisation of medical and healthcare curricula while supporting outcome-based education (OBE) as one of the proven approaches to education in medical and healthcare programmes [1]. These systems determine the curriculum content and its organisation, the teaching methods and the strategies, the offered courses, the assessment process, the educational environment and the curriculum timetable.

We have previously presented several studies at local and international academic conferences, where the partial results describing methodology, theoretical background, and pilot instance of the OPTIMED system were demonstrated [2–5]. Moreover, selected research questions regarding curriculum mapping, which covered in-depth analytical reports based on medical curricular data, were also released over the last few years [6–8]. This chapter systematically maps the entire development life cycle of this curriculum management system (CurrMS), where all fundamental aspects of actual deployment are entirely described. A thorough literature search was performed to see whether Ilios is the most widespread solution for curriculum management and innovations. This robust open-source CurrMS has been frequently adopted by medical, pharmacy, and dentistry schools and colleges. It includes functions such as programme, course, calendar and workflow management, data utilisation and reports using technical standards. The primary aim of both systems, Ilios and OPTIMED, is to fully satisfy regional medical and healthcare education requirements and easily integrate new modules and features per institutional needs.

The innovations, if performed in teaching domains formalised with the use of a detailed parametric description and entities adopted from the outcome-based concept (LUs: learning units, LOs: learning outcomes), will enhance the transparency and continuity of the environment in which the teachers, guarantors, faculty management representatives, as well as students, work daily [9]. Our approach to the design, development and implementation was based on the agile software engineering process. A set of functional (describe what the system must do, i.e. interactive reporting on available data) and non-functional (explain how the system works, i.e. user-friendly layout and navigation for end users) requirements were specified across a board of all involved stakeholders covering academics and students at MED MUNI. The result of the OPTIMED project is a robust web application that helps to integrate the well-established concept of a standardised definition of learning outcomes with a clear and transparent overview of immense metadata descriptions of all courses, learning units and outcomes related to the General Medicine study programme.

AIMS

- To define all relevant requirements on a web application and its module/components.
- To model a proper database structure for curriculum description based on a proven methodological background.
- To design and implement a complex reporting system on available data.

METHODS

ENGINEERING BACKGROUND

Today, there are various approaches to constructing software products and web applications. Software development methods are continually improved, driven by the need to ensure the final product's best quality and save as many costs as possible during the development process [10]. Engineers, managers, and other stakeholders face these challenges on different levels. We have adopted the Extreme Programming methodology [11], a popular agile software development framework. Its features are described in the Agile Manifesto of 2001 [12]. Here we list the three most important aspects of our objectives because the medical and healthcare faculty management and IT developers must communicate together intensively throughout the whole project: (i) easy adoption of new and frequently changing requirements helps to satisfy various needs of the academic community during the complicated in-house development of own CurrMS; (ii) productivity improvement, cost reduction, and checkpoint introduction by frequent releases in short periods provide delivery of working software with a preference for a shorter timescale; (iii) clear, simple, and concise coding and careful testing provide the required technical excellence and good software design.

There are also many variants of generic software process models (for example, the waterfall model, component-based software engineering, etc.). We have adopted the evolutionary development technique (see Figure 1), representing an iterative and incremental approach to software development [13]. This model is applicable for small- and medium-sized interactive systems (up to 500,000 lines of code), where fundamental activities such as formal specification, development and validation are interleaved. We have started with well-understood requirements and progressively added new features as proposed by the client. Then we have been developing the system through iterative cycles containing the actual increment as a slice of new functionality delivered from the requirements on the system deployment.

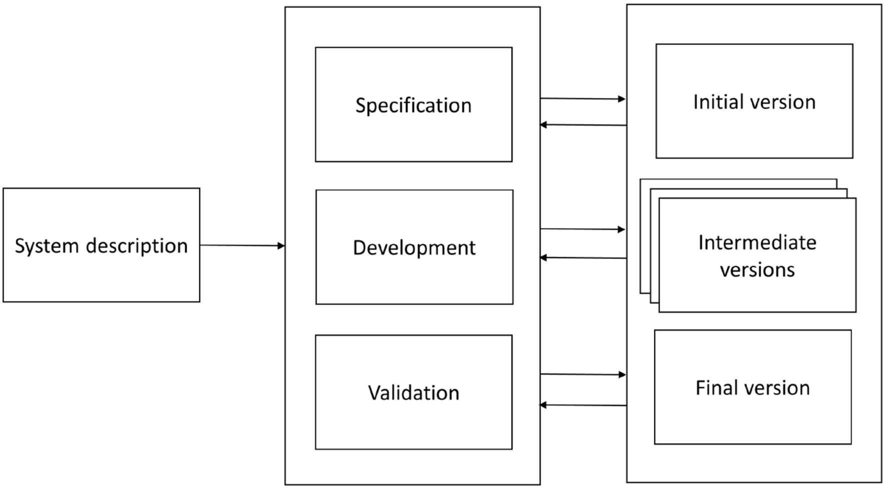


Figure 1: Schema of evolutionary development

A requirements analysis, which allows us to understand and define what essential web application services, functionalities and features, adopts ideas from object-oriented programming and blends them with ideas from semantic data modelling and knowledge representation into a modelling framework that is more powerful than traditional techniques such as data flow diagrams or structured analysis [14]. Moreover, we have applied the Unified Modeling Language (UML) [15] notation and carefully identified all essential features of the proposed system, transforming them into a set of static and dynamic models, such as use case diagrams or activity diagrams. Data management covering the design of a general conceptual data model and a formal database arrangement is based on our previous work [2,3,16] and utilises a legal description of a curriculum; this concept still needs to be implemented in similar solutions.

TECHNOLOGIES

Concerning technological background, the web-based platform runs on the most widely used web servers, including an Apache server or a Microsoft Internet Information Server. We use an Apache server running on Ubuntu Long Term Support (LTS), a Linux distribution of an operating system, concerning its stability and proven performance. The server-oriented scripting language PHP [17] was used to build the application layer of the platform. It represents a hybrid semantic programming language, according to Malkov [18]. Our application is built on a Nette framework. We decided to use this framework because it supports object-oriented design and event-driven development and provides a set of tools for debugging (for example, the Nette debugger). Thanks to mechanisms for securing the application, Nette ensures the smooth running of our system. This framework is released under an open-source license and can be extended with plenty of plug-ins and add-ons. PostgreSQL technology – an open-source Object-Relational Database Management System – was used for the data layer. This model is based on objects, classes and inheritance, which are directly supported in database schemas and the query language. In addition, just as with pure relational systems, it supports the extension of the data model with custom data types and methods.

Based on our previous experience with the development of the MEFANET (Medical FAculties NETwork) educational web portal platform [19,20], the following toolboxes were used: Nette framework and debugger (a tool for PHP web development, including a diagnostic error tool), Smarty (a template engine for PHP), and front-end interactive functionalities such as the jQuery JavaScript library (a simplification of the client-side scripting), jQuery Validation plugin (a client-side form validation), the jQuery Colorbox (lightbox effects), and DHTMLX (JavaScript library Ajax-powered user interface components). The presented system is based on Model-View-Controller, which is the most widespread software architecture used today for web applications [16]. Its main idea is to divide any application into three parts representing core functionalities: model, view, and controller.

The deployment diagram (see Figure 2) shows the existing hardware components, running software components, and how the different pieces of the developed system are interconnected. The central core of the application is placed at Apache HTTP Server [21] on the Ubuntu LTS. The set of scripts, which allows running every part of the OPTIMED CurrMS, is divided according to individual modules. Modules for data transfer communicate with applications outside the central server through clearly defined channels and standardised protocols. For data storage, we use the PostgreSQL database [22], one of the world's most advanced open-source database technology. Connection to the database is established by a connection string containing the following attributes: host, port, database name, username and password. All queries to the database are then implemented using SQL. Because of the sensitive exchange of user data, communication with eduID.cz is secured through an encrypted transfer (Secure Sockets Layer cryptographic protocol). The authentication data is passed using the Security Assertion Markup Language (SAML). Finally, the user accesses the application interface through a web browser. The stateless protocol HTTP maintains the connection.

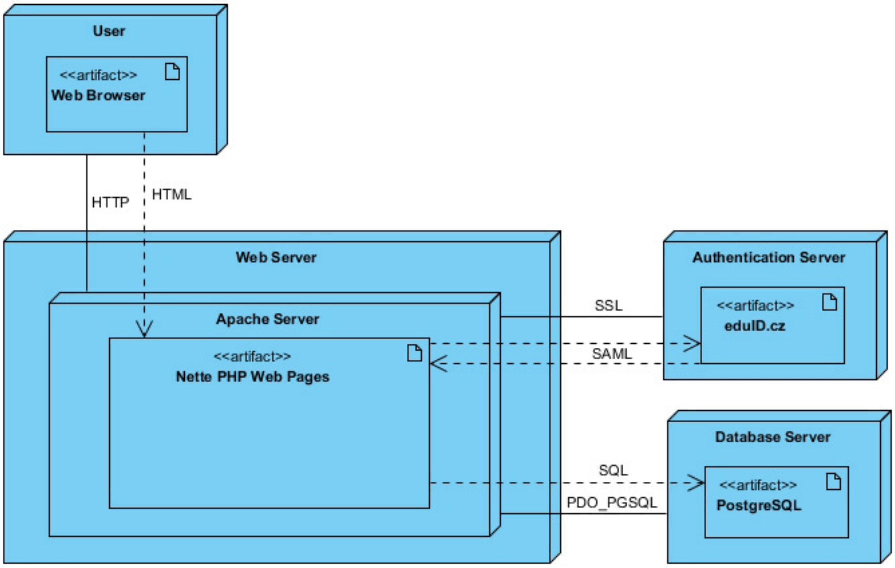


Figure 2: Deployment diagram of the OPTIMED platform

THIRD-PARTY SERVICES

The following external services significantly improve the curriculum management system's efficiency and quality. Google Analytics (GA) is a third-party service that measures and generates up-to-date statistics, reports and analyses based on website traffic and the behaviour of its visitors [23]. GA tracks the visitors' activity, collects real-time statistical data, and stores them for later analysis. Using this tool, we can determine the number of visitors over any period, which pages they viewed, and how long their visit was. An important factor is the flow of visitors, which shows the user transitions between pages and the rate of abandonment of every page. We use a unique monitoring code for each instance of the OPTIMED CurrMS – that is, we collect separate data sets. In conjunction with our other techniques for monitoring user activity, we can effectively provide user-friendly tools for managing the curriculum and improving it continuously. SAML is an open-source framework based on the eXtendable Markup Language (XML) [24]. This library was developed by the Security Services Technical Committee open standards consortium and provided authentication and identity services. SAML enables sending security information between two entities (for instance, client and server). The standard is extendable and customisable for various project types. One is called Shibboleth, an open-source software product for a single sign-on. OPTIMED includes a public part and an authenticated part. The public part is accessible without any authentication. The authenticated part, however, uses the eduID.cz authentication mechanism powered by the Czech academic identity federation. The eduID.cz mechanism provides means for inter-organisational identity management built with Shibboleth middleware [25]. The attributes of authenticated users sent by the Shibboleth are combined with second-layered privacy rules stored in the OPTIMED database. This approach allows users to be assigned different roles depending on the internal system policy and the external identity provided by eduID.cz.

RESULTS

BUSINESS UNDERSTANDING

Since the beginning, we have intensively communicated with all key stakeholders (curriculum designers, senior guarantors, members of the faculty management, as well as students' representatives) in regular meetings, where end-user requirements were continuously defined. We aimed to involve all these stakeholders in the needs identification process, which was necessary to determine the characteristics, properties or qualities the platform must possess. The analysis covered high-level statements from the faculty management's conclusions in the school's long-term plan and general statements emphasised during joint interviews. The different perspectives and opinions were gathered to build a complete picture of what the developed solution should achieve and the project's basic scope. These statements were transformed into a set of user stories (see Table 1) expressing end-user requirements in the case of agile development techniques. User stories [26] were written in business language (not technical) in the following format: As a _____, I ______, so that _______.

The stories provided an effective way to go through acceptance tests at the end of the development phase. The knowledge obtained from user stories was modified to a more detailed system specification list to help developers bridge the gap between system engineering and software design.

Table 1: Examples of user stories

User story
As a curriculum designer, I need to browse and manage (create/edit/delete) my learning outcomes, including all obligatory and optional parameters, at one screen, so I feel comfortable during the curriculum optimisation.
As a student, I want to have a set of advanced filters for browsing the described learning units so that more accurate results can be obtained.
As an end user, I want to log in to a system by entering my credentials from my home institution so I can be authenticated without the need to create/remember a new account.

Use case diagram (see Figure 3) describing the interactions with either the end user or another system was created to specify the expected behaviour but not show the exact method of making it happen. It summarises important relationships between use cases, actors and systems, and significantly helps the development team design a plan from the end user's perspective.

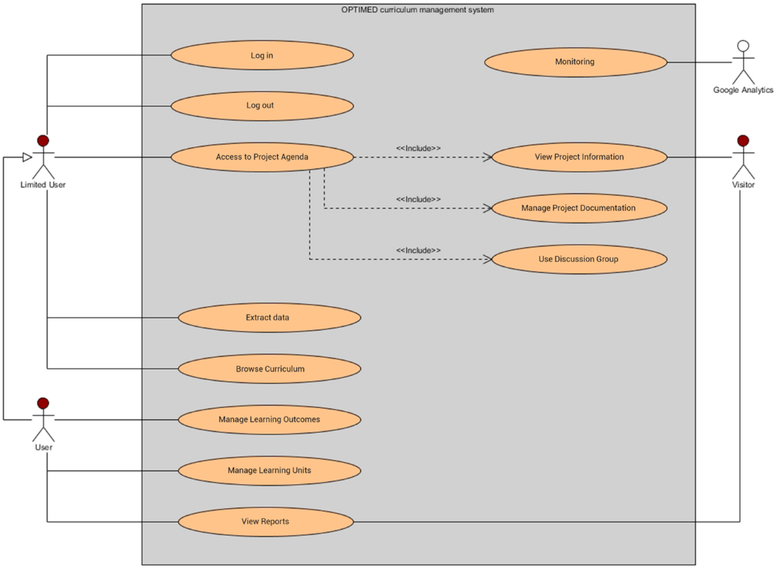


Figure 3: Use case diagram of the OPTIMED portal

DATA UNDERSTANDING

This phase covers all elements of global curriculum harmonisation, including a detailed specification of formal metadata (see Figure 4). Generally, it proves that software architects and designers do understand input data. The organisation of metadata and its linking is provided in the curriculum model, which can be implemented without any restrictions within any relational database technology. The next part of this section is devoted to the description of conceptual data modelling, where all crucial entities (see entity names in capital letters) are shown. The relation called LEARNING_UNIT is one of the essential elements in the OPTIMED system. Every unit describes a small piece of the teaching contents (for example, course lectures representing one coherent topic). It contains information about its creator (PERSON), who has one of the academic functions (SYSTEM_ROLE). The link to the relation MESH_THESAURUS shows that the unit is described by a set of MeSH (Medical Subject Heading) database keywords. The links between the unit and study materials (BIBLIOGRAPHY, ELEARNING_MATERIAL) were shown to be useful for students. The relation LEARNING_OUTCOME describes the knowledge and skills that students should have after completing a learning unit. It is composed of more sentences, each of them having a standardised form: a verb chosen by the rules of Bloom's taxonomy and a completion form (ASSESSMENT). Again, every learning outcome has a creator (PERSON) and a course affiliation (COURSE). The COURSE entity contains parameters of actual courses on the faculty. Courses are split into learning units, as mentioned above. Each COURSE is linked to a study field and a medical discipline (for example, Anatomy, Biochemistry, Neurosciences, etc.). Aggregations tables (UNIT_AGGREGATION, OUTCOME_AGGREGATION) summarise all necessary data from learning outcomes and learning units' relations with other tables. These aggregated tables keep all crucial data together and optimise rendering views or editing modules with easy curriculum export tools. And finally, they provide a strong base for reporting and further curriculum analyses. We use a particular refreshing set of scripts to keep the content of tables up-to-date with every related modification. Finally, a logical data model was designed based on the conceptual data model presenting information gathered from deep requirements analysis. This more complex model sets column types, normalises all entity and attribute names, and shows a detailed representation of curricular data.

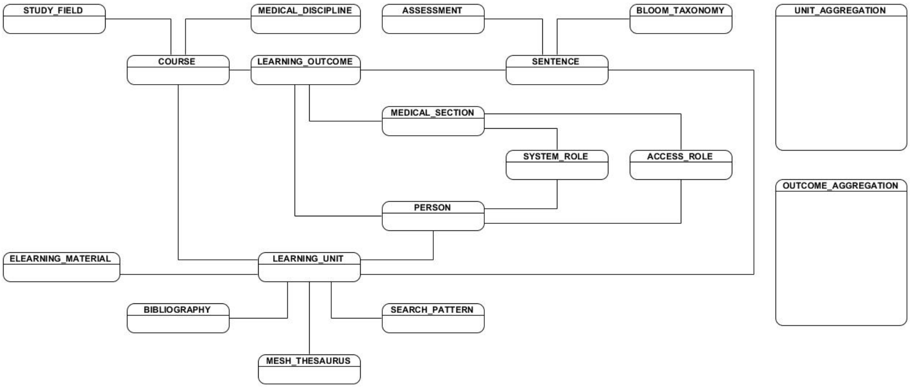


Figure 4: Simplified data model of a formal database arrangement describing the general curriculum

The proposed web-based platform was intended primarily for the academic staff (students, teachers, and faculty management). To make the platform accessible on the Internet, it had to be equipped with a robust authentication/authorisation framework. Federated authentication and authorisation services provided by the eduID.cz federation were therefore employed. We have defined (i) system roles, defining the position of users in the curriculum innovation process, and (ii) access roles, where all access restrictions – as mentioned in the Access Control List (ACL) for entrance to individual modules – are specified. Five essential system roles were idenified (curriculum designers, guarantors, heads of department, coordinators and developers). The curriculum designers and guarantors play an essential role, as they are responsible for the quality of curriculum description, including definitions of input knowledge and skills required in the subsequent practice. ACL represents a powerful tool for expressing the access control policy by clearly describing all permissions attached to any entity involved. Transparency in the definition of access control mechanisms is a crucial issue. Access roles are classified as users without authentication (visitors) and users with authentication (limited-access user, standard-access user, and advanced-access user). User security attributes in ACL cover groups to which the user belongs and corresponding roles assigned to the user. Object security attributes in ACL explain the permissions required to perform operations on the object. It specifies which users or external entities can be granted access to a specific part of the system and what processes are allowed on a given object.

DATA PREPARATION

For the next modelling phase, the input data had to be collected closely with all involved guarantors and teachers. Because of the different levels of ICT literacy among academics at MED MUNI, it was not possible to ensure the definition of a curriculum description through OPTIMED. The technicians prepared a set of structured MS Excel templates, which were then distributed to authors of the learning units together with methodological training and workshops. It was then necessary to unify all these data in terms of format and values, making a set of semi-automated transformations and mapping so that the content could be entered directly into the central OPTIMED database via SQL scripts.

MODELLING

We look at this phase from the perspective of modelling and web application development, where the stage can be compared to the actual development and implementation of defined functionalities. Logically, the development area will be described first, providing the data basis for subsequent analytical processing.

DEVELOPMENT OF THE OPTIMED PLATFORM

Based on the gathered needs during an iterative requirements analysis, a modular structure of the OPTIMED platform was proposed. Each independent module provides a separate functionality according to its practical use.

(i) Learning outcome register and (ii) Learning unit register were built on a data grid component, which makes learning outcomes and units more accessible and manageable for the curriculum experts. The data grid component offers a general functionality for displaying data flexibly in a table of rows and columns. Besides pagination, column resizing and sorting, both registers are also equipped with advanced searching and filtering features. (iii) Curriculum browser is a search engine designed to present the results of a user's search with clear links to the curriculum. Students, teachers and members of the faculty management can easily find what topics are covered in lessons belonging to a particular medical section, discipline, course or year of study. The filtering toolbox with several parametric domains (e.g. type of teaching, hour range, a form of evaluation, etc.) provides a quick specification of a search query and gets more accurate results. (iv) Data export is a powerful utility that allows the user to select, to aggregate, and to export data in a structured CSV (comma-separated values) format with the use of predefined and customised forms. It significantly helps teachers in a periodical control process of learning units.

DESCRIPTIVE REPORTING ON CURRICULUM DATA

During the OPTIMED project, analytical reports were repeatedly prepared and presented to the audience of the OPTIMED stakeholders. The purpose was to show the current status of the defined curriculum data structures and parametric descriptions, including their quantitative and qualitative overview. These outputs help deliver valid and objective summaries as the volume of data could not be assessed in real-time. Concerning the detailed parameters that describe the whole curriculum, it was possible, for example, to create an overview of the distribution of learning units according to time allocation. Figure 5 shows how the four different medical sections of the General Medicine study programme are covered concerning the time commitment of the full-time study without the recommended self-study.

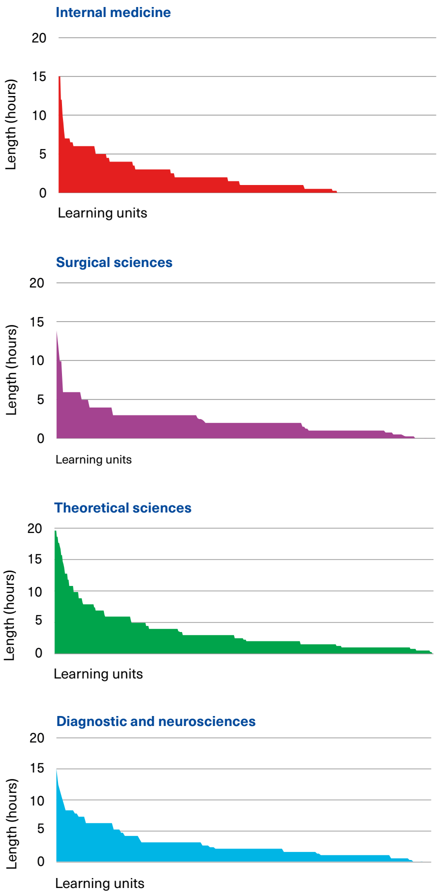


Figure 5: Representation of learning units about length in hours

The following overview (see Figure 6) shows the distribution of learning outcomes in the learning units within the medical sections. The average number of these outcomes per learning unit defined by curriculum designers across the curriculum is 5.56.

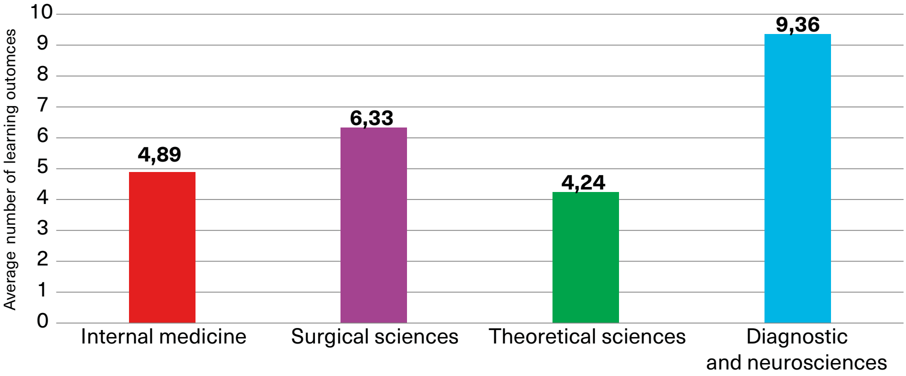


Figure 6: Distribution of learning outcomes in learning units and medical sections

EVALUATION

The modular platform has been implemented at the Faculty of Medicine at Masaryk University, the second largest and second oldest university in the Czech Republic. It offers an appealing way to effectively reform medical education concerning courses' interrelation and logical intersections of a specific study field, where the emphasis is on the product – what sort of graduates should be produced – rather than on the educational process itself. The primary effort is a comprehensive curriculum innovation of the General Medicine study field. The creations are driven towards a smoother continuity between the theoretical and clinical phases of the study and by the need to deliver graduates with 21st-century skills. This project's key point is using ICT tools such as comprehensive registries and browsers to achieve a horizontally innovated structure of compulsory and compulsorily optional courses. The objective is not a radical change in learning or teaching; instead, it is an exploratory mapping of the current state of the General Medicine curriculum with a prospect of innovations toward a more transparent educational environment [16]. The faculty management charged 385 curriculum experts to harmonise the curriculum in different roles and provide feedback to the developers of the ICT mashups from which the web-based platform is composed. They have created a detailed metadata description of the General Medicine study field, which takes over 2,500 pages of text. Table 2 gives a numerical summary describing this particular study field.

Table 2: Summary of the OPTIMED platform content

Medical curriculum domain	Example	Total number
Modules	Internal medicine	4
Medical disciplines	Paediatrics	44
Courses	Paediatrics II - practice	144
Learning units	Acute and late toxicity of treatment	1,342
Learning outcomes	The student describes the treatment principles of these tumours.	6,977

DEPLOYMENT

After its deployment, the OPTIMED platform became the first comprehensive online tool (CurrMS) that enabled a structured medical and healthcare education description. In the academic infrastructure of MED MUNI, OPTIMED was primarily accessible to teachers, guarantors, and members of the faculty management. The agile approach to development provided the ability to respond quickly to new assignments that reflected actual use in updating individual courses, learning units, and outcomes. Beyond the standard CurrMS, OPTIMED also included several interactive visualisations of available data. These were subsequently discussed with course guarantors and at OPTIMED working group meetings and faculty conferences dedicated to education improvement.

DISCUSSION

PRINCIPAL RESULTS

This chapter introduces the original curriculum management system's design, development, implementation, and evaluation framework. The deployed OPTIMED platform, described here, is based on the well-established concept of a standardised definition of learning outcomes. The authors have focused on the reform vision statement as a road map for curriculum revision. The online platform supports curriculum management, and HEIs provide comprehensive services to a broad academic community. Academicians should be able to explore the general curricula of individual medical and healthcare study programmes. One of its key features is its dynamic character and the ability to easily manage any domain closely related to a medical curriculum undergoing upgrades, and to absorb and incorporate all changes into the educational process. The literature overview shows the existence of a variety of systems which are only partially able to meet the needs of today's tertiary education and directly support in-depth curriculum management. OPTIMED offers valuable functionalities and can help academics in their curriculum reengineering efforts. Since 1st April 2014, OPTIMED has been available to all students and teachers at the Faculty of Medicine at Masaryk University (more than 15,500 persons in total). It is used in everyday practice by thousands of students and hundreds of senior teachers, curriculum designers and professional guarantors within the content inspection of the medical curriculum. As a basis for the system, the proposed conceptual data model formally describes a general curriculum structure and can be implemented without any restrictions within any database technology.

LIMITATIONS

The global trend supports international standards integration aimed at improving the quality of life for an increasing number of people [27]. Proven standards exist for financial transactions, telecommunication, environmental protection, food safety, and many other areas of human interest. People who come into contact with medical and healthcare graduates usually expect professional conduct and assume they have met the licensing standards for practising medicine. Today's need for a standardised curriculum, particularly in medical education, is indispensable. Under the MedBiquitous Consortium, the leading organisation in developing and promoting technology standards for health professions, a set of proven standards and recommendations was defined. MedBiquitous has developed eight standards to advance lifelong learning, continuous improvement, and better patient outcomes for health (such as the Competency Framework, the Curriculum Inventory, the Healthcare Professional Profile, etc.). From the perspective of international, systematic, practical, and long-term curriculum innovation approaches, the MedBiquitous recommendation in technical standards must be considered. Several institutional and commercial standard-compliant curriculum management systems mainly support the Curriculum Inventory standard defining core curriculum data for health professions education programmes. Regarding standardised support on the technological level, OPTIMED is currently limited because it is not compliant with any technical standards. Following the outputs of the OPTIMED project, a selected standard (namely, the Curriculum Inventory) was adopted and implemented as a new extension providing a unified communication channel across organisations and various systems for curriculum management in medical and healthcare study fields.

EVALUATION OF THE AIMS OF THE CHAPTER

- To define all relevant requirements on a web application and its module/components.
  - In close collaboration with the most relevant target users (teachers and students), a specification was developed using modelling tools and user stories, which provides a basic set of requirements for a system to collect and to process data describing the medical curriculum.
- To model a proper database structure for curriculum description based on a proven methodological background.
  - The comprehensive data model that was the basis of the OPTIMED portal contained all the essential characteristics in the process of medical curriculum description. Moreover, it provided easy extensibility for implementing internationally valid and proven standards.
- To design and implement a complex reporting system on available data.
  - The two above-mentioned goals were essential to the design and subsequent development of a unique system that included, among other things, interactive data visualisations. These were mainly used as a basis for further harmonisation and updates in well-balanced curriculum development.

LESSONS LEARNED

The key to the project's success was a detailed understanding of the needs and priorities of the faculty, which set a goal of a fully described curriculum for its broadest study programme. Over 350 colleagues representing faculty management, medical discipline guarantors, teachers and the development/analysis team were involved in the project throughout its implementation. After evaluating the entire process, steps were identified that would be implemented differently when repeated, particularly in terms of effectiveness in specifying the assignment and communicating with the various target groups. The most time-consuming stage was the data preparation stage, as the design and development of the system for online data collection on individual subjects was carried out in parallel with the development of the structured curriculum description.

However, it can be clearly stated that the systematic approach in the design of the curriculum description positively influenced the outcome in the form of a unified metadata structure for the different parts of the curriculum, such as course, learning unit and learning outcome.

REFERENCES

[1] Harden RM, Crosby JR, Davis MH. AMEE Guide No. 14: Outcome-based education: Part 1. An introduction to outcome-based education. Med Teach. 1999;21:7–14.

[2] Komenda M, Schwarz D, Vaitsis C, Zary N, Štěrba J, Dušek L. OPTIMED Platform: Curriculum Harmonisation System for Medical and Healthcare Education. Stud Health Technol Inform. 2015;210:511–5.

[3] Komenda M, Schwarz D, Hřebíček J, Holčík J, Dušek L. A Framework for Curriculum Management – The Use of Outcome-based Approach in Practice. In: Zvacek S, Restivo MT, Uhomoibhi J, Helfert M (eds). CSEDU 2014 - Proceedings of the 6th International Conference on Computer Supported Education (Volume 2). Barcelona: SciTePress; 2014. p. 473–478. ISBN 978-989-758-020-8.

[4] Komenda M, Pekárková L. Curriculum planning and construction: How to combine outcome-based approach with process-oriented paradigm. In: Hřebíček J, Ministr J, Pitner T (eds). 10th Summer School of Applied Informatics. Brno: Littera; 2013. p. 97–103. ISBN 978-80-85763-72-0.

[5] Komenda M, Schwarz D, Dušek L. The role of information technologies in medical curriculum harmonisation. PeerJ PrePrints 2015;3:e1093v1. doi: 10.7287/peerj.preprints.1093v1.

[6] Komenda M, Schwarz D, Švancara J, Vaitsis C, Zary N, Dušek L. Practical use of medical terminology in curriculum mapping. Comput Biol Med. 2015;63:74–82.

[7] Komenda M, Víta M, Karolyi M, Kríž V, Pokorná A. Automatic Keyword Extraction from Medical and Healthcare Curriculum. Annals of Computer Science and Information Systems, Volume 8: Proceedings of the 2016 Federated Conference on Computer Science and Information Systems. Warzaw, Los Alamitos: Polskie Towarzystwo Informatyczne, Institute of Electrical and Electronics Engineers; 2016. p. 287–290. ISBN 978-83-60810-90-3.

[8] Víta M, Komenda M, Pokorná A. Exploring medical curricula using social network analysis methods. In: Ganzha M, Maciaszek L, Paprzycki M (eds). 2015 Federated Conference on Computer Science and Information Systems

(FedCSIS). Warsaw, Los Alamitos: Polskie Towarzystwo Informatyczne, IEEE; 2015. p. 297–302. ISBN 978-83-60810-66-8.

[9] Komenda M, Víta M, Vaitsis C, et al. Curriculum mapping with academic analytics in medical and healthcare education. PLoS One. 2015;10(12):e0143748.

[10] Magdaleno AM, Werner CML, de Araujo RM. Reconciling software development models: A quasi-systematic review. J Syst Softw. 2012;85(2):351–69.

[11] Beck K, Andres C. Extreme programming explained: embrace change. Boston: Addison-Wesley Professional; 2004. ISBN 978-0-321-27865-4.

[12] Cunningham W. Manifesto for Agile Software Development [Internet]. 2001 [cited 16 Jan 2022]. Available from: https://agilemanifesto.org/.

[13] Sommerville I. Software Engineering. London: Pearson Education; 2007. ISBN 978-0-321-31379-9.

[14] Mylopoulos J, Chung L, Yu E. From object-oriented to goal-oriented requirements analysis. Commun ACM. 1999;42:31–7.

[15] Rumbaugh J, Jacobson I, Booch G. The Unified Modeling Language Reference Manual. Boston: Addison-Wesley Professional; 2004.
ISBN 978-0-321-71895-2.

[16] Komenda M, Schwarz D, Dušek L. Towards a System of Enhanced Transparency of Medical Curriculum. Eur J Biomed Inform. 2013; 9(3):9–16.

[17] Bakken SS, Aulbach A, Schmid E, et al. PHP manual. Zend Technologies; 1997. Available from:
https://users-cs.au.dk/~bouvin/hyper03/php/PHP_Manual.pdf.

[18] Malkov SN. Customizing a functional programming language for web development. Comput Lang Syst Struct. 2010;36(4):345–31.

[19] Komenda M, Schwarz D, Feberová J, Štípek S, Mihál V, Dušek L. Medical faculties educational network: Multidimensional quality assessment. Comput Methods Programs Biomed. 2012;108(3):900–9.

[20] Schwarz D, Dušek L. Computer applications, Systems and Networks for Medical Education. Brno: Facta Medica; 2014. ISBN 978-80-904731-9-5.

[21] Bowen R, Coar K. Apache Cookbook: Solutions and Examples for Apache Administration. Sebastopol: O'Reilly Media; 2008. ISBN 978-0-596-52994-9.

[22] Momjian B. PostgreSQL: Introduction and Concepts. Boston: Addison--Wesley; 2001. ISBN 978-0-201-70331-3.

[23] Plaza B. Google Analytics for measuring website performance. Tour Manag. 2011;32:477–81.

[24] Armando A, Carbone R, Compagna L, Cuellar J, Tobarra L. Formal Analysis of SAML 2.0 Web Browser Single Sign-on: Breaking the SAML-based Single Sign-on for Google Apps. In: FMSE '08: Proceedings of the 6th ACM Workshop on Formal Methods in Security Engineering. New York: Association for Computing Machinery; 2008. p. 1–10. ISBN 978-1-60558-288-7.

[25] Needleman M. The Shibboleth Authentication/Authorization System. Ser Rev. 2004;30(3):252–3.

[26] Curcio K, Navarro T, Malucelli A, Reinehr S. Requirements engineering: A systematic mapping study in agile software development. J Syst Softw. 2018;139: 32–50.

[27] Wojtczak A, Schwarz MR. Minimum essential requirements and standards in medical education. Med Teach. 2000;22(6):555–9.

SECTION B

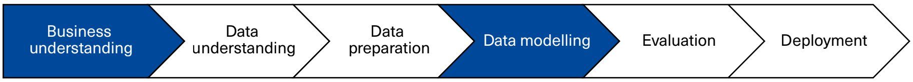
| 02 |

COMPLEX STRUCTURED EVALUATION OF SELECTED PARTS OF THE MEDICAL CURRICULUM

Martin Komenda, Andrea Pokorná, Daniel Schwarz, Petr Štourač

The case study was supported by the following grant project: EVAMED — evaluation of medical education — reg. no.: MUNI/A/1512/2014, funded by the specific research and rector's programme of Masaryk University.

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2015–2015
Keywords	Medical and healthcare education, evaluation, educational data mining, curriculum innovation
Research question	What is the impact of a structured evaluation system on the quality of medical curriculum content?
Type of result	Static analytical report
Level of data processing	Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

The effort to fundamentally simplify and standardise the teaching structure is one of the long-term priorities of the Faculty of Medicine of Masaryk University (MED MUNI). One of the critical steps was the successful implementation of the OPTIMED project. Thanks to this, a unique platform presented a medical curriculum description based on the learning outcomes method. The EVAMED (Evaluation of Medical Education) project has a direct continuity, including ensuring the sustainability of OPTIMED. It mainly focuses on advanced data analysis and visualisation, viewing the central OPTIMED database as a rich source of heterogeneous but valid data. Given that more than 400 educators and guarantors have been involved in the detailed description of the curriculum, it is evident that the granularity of this information shows high variability. The inherent degree of subjectivity of the author-educator in the formulation of learning outcomes and the specification of other descriptive attributes of the study has divided the content into differently formulated components, ranging from general to very detailed – both in the specification of primary and extension knowledge and in the definition of learning outcomes and interconnections within cross-curricular links, both vertical and horizontal. The attempt to systematically evaluate the available textual description of selected parts of the medical curriculum is the fundamental idea of this case study.

EVAMED joined two working groups to identify, through newly defined metrics, inconsistent learning units. These groups involve (a) guarantors and representatives of faculty management and (b) senior students. Using newly designed and implemented evaluation approaches and data mining methods, it was possible to analyse heterogeneous data sources to unify the content of the OPTIMED database (containing more than 1,400 learning units and around 8,000 learning outcomes), leading to an increase in the quality of the description of the whole curriculum. The results in the form of a detailed report not only help to expand the know-how in the field of medical curriculum optimisation and educational data mining but also provide a pilot process for future evaluation of the description of the curriculum not only in the local environment of medical and healthcare faculties in the Czech Republic and Slovakia within the MEFANET educational network but also internationally.

AIMS

— To define a structured system of evaluation templates to help ensure an effective collection of medical curriculum evaluations.

— To design a set of analytical reports that provides a clear view of the evaluation of medical curriculum content in summative and aggregate outcomes.

— To involve upper-year students under the supervision of guarantors in systematically evaluating selected parts of the curriculum.

METHODS

From the methodological point of view, it was crucial to design an algorithmic procedure for the systematic evaluation of teaching (in our case, of learning units available on the OPTIMED portal). The public educational data about the General Medicine study programme stored in the OPTIMED central database were described using a multi-step curriculum innovation process consisting of four phases:

1. Setting up the medical curriculum structure. The study field is divided into individual sections, medical disciplines, courses, and learning units, including the responsible supervisors and guarantors.

2. Defining the descriptive attributes of the LUs, linked MeSH vocabulary keywords and essential terms, and associated outcomes, according to Bloom's taxonomy. The outcomes typically consist of a noun or noun phrase (i.e., the subject matter content) and a verb or verb phrase (i.e. the cognitive process). In this case, each LO defines what students are expected to know, understand, and be able to demonstrate at the end of the learning period, typically as a graduate. Several academic institutions have already applied this concept, especially in medical education.

3. Vertical harmonisation, i.e., content optimisation of learning units and further discussion within individual sections, under the supervision of responsible guarantors.

4. Horizontal evaluation, i.e., follow-up discussions across all sections under the management of supervisors, including an in-depth inspection in collaboration with an established expert committee, which can logically influence the whole structure of the defined curricula.

The data model represents an original methodology for creating a new structured set of learning units and outcomes designed to assess and adjust to real education [1]. The EVAMED project brings an innovative guide for systematically evaluating the curriculum (each phase is briefly described below).

A structured system of evaluation templates for curriculum evaluation was designed by a team of expert guarantors and implemented in practice as three related parametric questionnaires shown below. The questions used are based on published outputs and results [2–6], and the degree of subjectivity in their formulation should therefore be minimised.

Form A: Demographic profile of evaluators

- Unique university identifier of the person
- First name
- Last name
- Age
- Gender
- Year of study
- Average grades for the whole study
- Age in the final year of study

Form B: Evaluation of learning units

- Evaluate the overall clarity of the assessed learning unit.
- The overall scope of the learning unit, as indicated on the identification context card, corresponds to the actual time allocated in the contact teaching.
- Learning objectives in the form of defined learning outcomes are concisely and clearly stated.
- The description of the learning unit corresponds to the stated objectives of the relevant courses according to their syllabuses in the IS MU[1].
- The description of the learning unit corresponds to the teachers' interpretation in actual contact teaching.
- The description of the learning unit is up-to-date and corresponds to the current knowledge of biomedicine.
- The description of the learning unit is clinically relevant, i.e. usable in clinical practice.

[1] https://is.muni.cz/

- The essential concepts of the learning unit have sufficient explanatory value.
- Relevant concepts of the learning unit correspond to critical ideas/topics in contact teaching.
- The learning unit references related courses within the vertical link/level, i.e. previous years' courses.
- The learning unit allows for the integration/connection of knowledge from preclinical courses with clinical experiences.
- The learning unit motivates the search for new knowledge in additional study resources.
- Information resources are appropriate/allow for additional information.
- The learning unit includes recommended study resources in the Information Resources section that are appropriate and applicable to teaching.
- The learning unit contains references to existing electrical support in the Information Resources section that are appropriate and useful for teaching.

Form C: User feedback on the OPTIMED portal

- Evaluate the OPTIMED portal from the user's point of view and, if necessary, provide constructive critical opinions, suggestions, or comments.
- What new functionalities or features would you like to see on the OPTIMED portal?
- Do you know of any similar project/system/publication that focuses on topics like OPTIMED? If so, please provide the name or a link to these resources.
- Can you think of anything else?

Stratification of content is a proportional selection of learning units that will then be given to students for evaluation. The authors assumed a similar level of elaboration of learning units of one or more subjects within the same authoring team of teachers and supervisors. Coverage of all sections, medical disciplines and courses, including the author team (the discipline-course-author triad) will be worked with to ensure that each of the triads above has its learning units in the selected content (the number will be determined in proportion to the overall composition of the curriculum). 40% of the available content was evaluated. It is a realistic option in terms of feasibility (approximately 540 learning units from the OPTIMED central database). Proportional selection of learning units ensured coverage of the entire curriculum of the General Medicine study programme. An equal number of learning units were selected from each course for maximum balance. At the same time, the variety of author collectives within a course was also considered. (i) Constructing a classification of groups of units according to the courses represented, (ii) assigning information about the author collective, (iii) deploying an algorithm that pseudo-randomly selects from the list of instructional units for the actual evaluation.

The assignment of evaluators determines who will evaluate what. Each assessed learning unit was assigned to four students (in an attempt to eliminate the subjectivity of perspective and differences between the knowledge levels of the assessors). Each student was given a unique set of instructional units (ensuring variability). Each student will be assessed on instructional units they have already completed (student's current year of study vs the year that the instructional team is taught).

The pilot evaluation feasibility study provided information on a small sample of students that production evaluations, including optional comments, would be implemented.

The evaluation process was implemented through three structured forms in the Information System of Masaryk University.

Evaluation of the results involved a detailed analysis combining student demographic data with the review itself. When compiling the final report, use probability weighting (setting coefficients) to account for the importance of the evaluation sections of the units within the course/discipline.

RESULTS

BUSINESS UNDERSTANDING

This initial phase focuses on understanding the objectives and defining the problem regarding medical curriculum innovation. The main goal was to understand in detail the area of the structure of the medical curriculum description and to appropriately set up an objective evaluation process in collaboration with the teaching supervisors and students. At the same time, it was necessary to consider the faculty management's needs for feedback regarding the entire course of teaching. As a result of this part, the project guarantors agreed on the methodological solution. A set of assessment instruments in the form of three structured forms was proposed, which will fundamentally help to ensure the effective collection of medical curriculum evaluation.

1. Evaluator demographics (used solely to collect information about the evaluating students and completed by each just once at the start of the evaluation).

2. Evaluation of selected learning units in General Medicine (each student completes a separate form for each learning unit assigned to them; a Likert scale was used – a technique for measuring attitudes in questionnaires, 15 questions in total, including an option to justify non-completion and space for optional commentary, divided into five sections: Overall structure of the learning unit, description, objectives and outcomes of the learning unit, essential concepts, resources and feedback mechanisms, summary evaluation (strengths and weaknesses)).

3. Evaluation of the portal from the user/student perspective (optional to be completed by the student if they want to participate in the further development of the OPTIMED platform actively).

DATA UNDERSTANDING

Crucial to this phase was the successful implementation of the OPTIMED project and the subsequent selection of appropriately parametrically described parts of the curriculum for the follow-up evaluation. These include all elements related to the overall curriculum innovation, including detailed metadata specifications down to the level of learning units and links to learning outcomes. It was also possible to set up the necessary technicalities, such as access to available content, creating forms with links to follow-up data exports, and preparing methodological manuals and workshops for supervisors and students involved in the evaluation.

DATA PREPARATION

This step also represents an important stage in preprocessing the input data and enables the following complete analysis. Data validation is an integral part of this process, where incomplete (evaluations were not fully completed and saved) or invalid (students assessed learning units that they had not personally completed in the past) evaluations had to be removed (see Figure 1). This resulted in three modified datasets: (i) demographic profile of 26 students (first name, last name, unique university identifier of the person, age, gender, year of study, average grades for the whole study and the previous year of study), (ii) evaluation of selected instructional units (1,408 after validation), (iii) export of the list of learning units from the central OPTIMED database (1,366 in total, of which 543 were evaluated and unique).

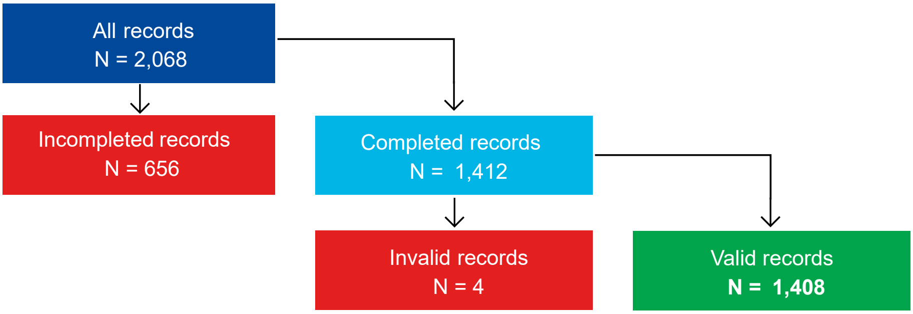


Figure 1: Data preprocessing schema

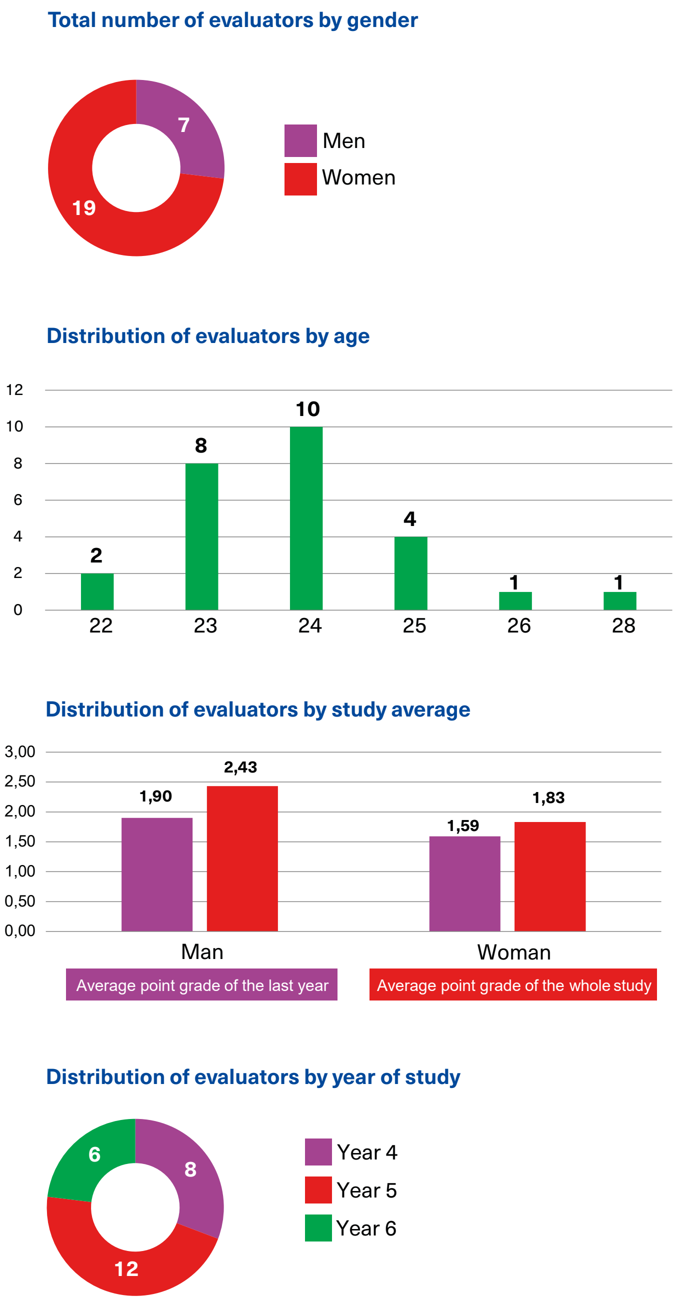
MODELLING

This phase involved mapping the final datasets, in-depth analysis and visualisation of the results, and a description of the critical characteristics of the evaluation process. The final reports contain views of both aggregated and row data (in this case, individual evaluations) with links to students. The proportion of learning units in the four medical sections (Theoretical Sciences, Diagnostic Disciplines and Neurosciences, Surgical Sciences, Internal Medicine) that were assessed (this was 40% of the total number of learning units in the General Medicine study programme) was kept. Below is a summary table showing the main descriptive characteristics of the evaluation.

Table 1: Summary of evaluation of selected medical disciplines

Total evaluation time per group of students (hours)	1,750
Total number of learning units on the OPTIMED portal	1,348
Percentage of assessed learning units (%)	40.2
Number of unique assessed learning units	542
Number of independent evaluations of the unit	4
Total number of units assessed	2,068
Number of students	26
The average number of learning units per student	80
Total time burden per student (hours)	67.3

The detailed profile of the target group was based on the results of the first questionnaire survey. The output is a list of evaluators divided according to essential characteristics in the form of descriptive summary such as gender, age, year of study or grade point average.


Total number of evaluators by gender

Distribution of evaluators by age

Distribution of evaluators by study average

Distribution of evaluators by year of study

Figure 2: Form A – Demographic profiles of 26 students in total

The average number of all evaluations (one evaluation record means one learning unit) per one student evaluator was 79.5 learning units, the total number of evaluated units was 542 (which accounts for 40.2% of the curriculum), and the total number of evaluations (each learning unit was reviewed by four independent student opinions, which was done in order to guarantee higher objectivity) was 2,068.

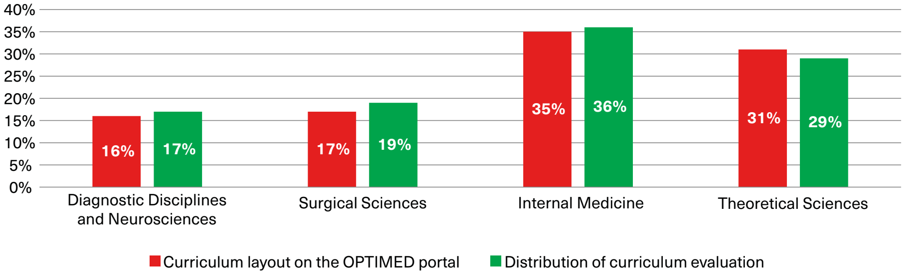


Figure 3: Overview of evaluation of curriculum divided into four medical sections

The aggregate results from all the evaluations provided several different perspectives, such as the average ratings for each question or the overall average rating for each medical section (Table 2).

Table 2: Average curriculum rating by medical section (the school scale)

Medical discipline	Theoretical Sciences	Diagnostic Disciplines and Neurosciences	Surgical Sciences	Internal Medicine
Average rating	1.79	1.97	1.87	1.79

One of the essential prerequisites for a correct evaluation and valid results was that the student, as the evaluator, had already passed the given learning unit. They could therefore provide relevant feedback. A look (Figure 4) at the aggregate average rating by year of the study clearly shows that students in the lower years (study years 1, 2, 3), on average, rated less critically than students in the higher years (4, 5, 6). Despite this, the average rating of any year was never worse than 2 (on the school scale), which is an excellent result.

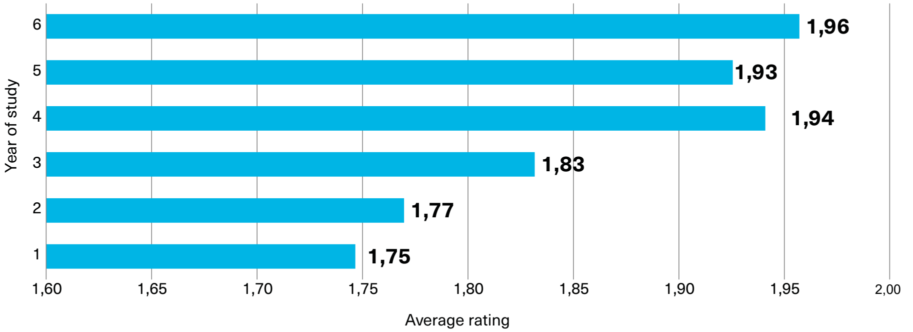


Figure 4: Average learning unit rating depending on the evaluator‘s year of study (the school scale)

When evaluating the questions from each section of form B, the learning units' descriptions, objectives, and outcomes were the best (averaging 1.47 on the school scale). In contrast, the resources and feedback mechanisms sections were rated the worst (averaging 2.58 on the school scale).

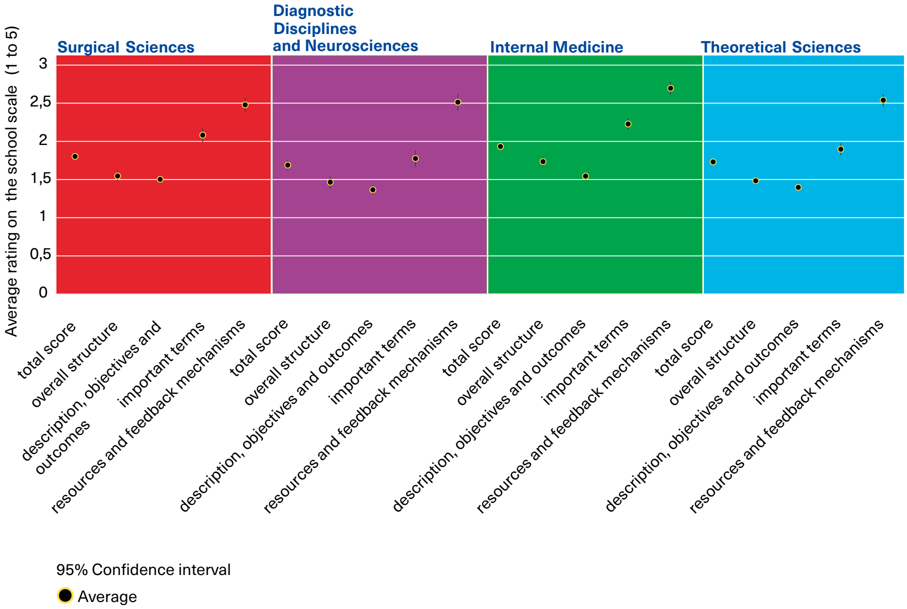


Figure 5: Evaluation ratings by medical sections

A final example (Figure 5) of outcomes is a comprehensive view of the curriculum, divided into medical sections and thematically related issues grouped into domains. It can be seen here that the trend of evaluation is similar from section to section. Despite this, it is easy to identify the weakest parts in each section.

EVALUATION

At this stage, a thorough back-check of all results was carried out. Independently of the methodological and analytical teams, the reports were validated against the original information from the systems from which the data were exported. The final outputs were also validated by the guarantors involved in the EVAMED project, where key findings have also been identified that need to be systematically addressed and solved as part of global curriculum innovation.

DEPLOYMENT

The primary output of this phase was the final report of the one-year EVAMED project, which included a description of the objectives, the project research team together with students, the timeline, the curriculum evaluation process, an algorithmic solution for selecting the content to be assessed, a set of structured evaluation forms, summary overviews of the aggregated data, detailed outlines of the individual questions ordered by medical disciplines, textual comments, and project conclusions.

DISCUSSION

The EVAMED project provided a new way to systematically evaluate data stored in a curriculum management system through a parametric set of questionnaires. The theoretically proposed concept of the evaluation was successfully implemented in practice. Its output is an objectified student's view of the current description of the General Medicine study programme teaching. Critically and to eliminate subjective opinion, 40% of the content of the OPTIMED portal was evaluated by a group of 27 students from higher years of study. The curriculum distribution on the OPTIMED portal corresponds entirely with the distribution of the content selected for self-evaluation. The global view of the evaluation is positive (the average rating across all attributes is 1.87, which corresponds to "very good" on the standard grading scale). Students fundamentally perceive the need for more interconnectedness of the courses and the lack of online learning materials to supplement the necessary information. The aggregate average ratings of the given questions are shown below.

Table 3: Basic characteristics of evaluation of learning units

Section	Question	Average rating[2]	Number of ratings	Number of comments
The overall structure of the learning unit	Question 1: Evaluate the overall clarity of the assessed learning unit.	1.64	2,068	393
The overall structure of the learning unit	Question 2: The total length of the learning unit indicated on the Identification context card corresponds to the actual time allocation in contact teaching.	1.55	1,952	139
Description, objectives, and outcomes of the learning unit	Question 3: The learning objectives in the form of defined learning outcomes are concisely and comprehensibly stated.	1.48	2,058	237
Description, objectives, and outcomes of the learning unit	Question 4: The description of the learning unit corresponds to the stated objectives of the relevant courses according to their syllabuses in IS MU.	1.69	2,029	528
Description, objectives, and outcomes of the learning unit	Question 5: The description of the learning unit corresponds to the teachers' interpretation of actual contact teaching.	1.53	1,995	164
Description, objectives, and outcomes of the learning unit	Question 6: The description of the learning unit is up-to-date and corresponds to the current knowledge of biomedicine.	1.23	1,716	42
Description, objectives, and outcomes of the learning unit	Question 7: The description of the learning unit is clinically relevant - applicable to clinical practice.	1.56	2,036	171
Important terms	Question 8: The important terms of the learning unit have sufficient meaning.	1.79	2,043	346
Important terms	Question 9: The important terms of the learning unit correspond to the fundamental concepts/topics in contact teaching.	1.75	2,043	90
Important terms	Question 10: The learning unit includes links to related courses within the vertical link/level - previous years' courses.	2.92	1,894	244

[2] The school rate: 1 – positive, 5 – negative.

Section	Question	Average rating[2]	Number of ratings	Number of comments
Important terms	Question 11: The learning unit allows integration/connection of knowledge from preclinical courses with clinical experience.	1.88	2,005	102
Resources and feedback mechanisms	Question 12: The learning unit motivates you to search for new knowledge in additional study resources.	1.96	2,052	131
Resources and feedback mechanisms	Question 13: Information sources are suitable/allow for additional information.	1.74	2,016	161
Resources and feedback mechanisms	Question 14: In the Information Resources section, the learning unit contains recommended study resources suitable and applicable for teaching.	1.60	2,007	269
Resources and feedback mechanisms	Question 15: The Information Resources section of the tutorial contains links to existing electrical support that are appropriate and useful for teaching.	3.60	1,979	318

The final analytical report for the management of the Faculty of Medicine of Masaryk University is one of the case studies where objective and valid information was prepared as material for decision support and innovation implementation in selected courses of the given curriculum.

EVALUATION OF THE AIMS OF THE CHAPTER

- To define a structured system of evaluation templates to help ensure the effective collection of medical curriculum evaluations.
  - In a collaboration between senior expert guarantors and students at the Faculty of Medicine, it was possible to design structured feedback through a set of templates, including a methodology for dividing the assessed content. This output subsequently enabled feedback collection from student evaluators across different years of study.
- To design a set of analytical reports that provide a clear view of the evaluation of medical curriculum content in summative and aggregate outcomes.
  - The design of the final analytical reports was inspired by long-term experience in implementing academic projects that focused on data

[2] The school rate: 1 – positive, 5 – negative.

processing and visualisation. The output consisted of basic descriptive statistics and available data and a detailed analysis where the project team's requirements were taken into account.

- To involve upper-year students under the supervision of guarantors in systematically evaluating selected parts of the curriculum.
    - The active involvement of student evaluators in the EVAMED project was the result of long-term collaboration on previous projects that focused on curriculum building and optimisation. MED MUNI generally respects the opinion of the student community, and in this case study, it proved to be quite crucial as well. In particular, a set of educational workshops was a vital aspect of the valid evaluation outputs. The methodology and established curriculum evaluation system will be addressed in detail with the project sponsors.

LESSONS LEARNED

This project was unique in the level of involvement of students from the Faculty of Medicine of Masaryk University, who were gradually involved in all stages of its solution from the very beginning. Therefore, they were thoroughly familiar with the structured feedback and knew the parts of the curriculum being evaluated. The data collection itself was relatively straightforward, although time-consuming, from the students' (evaluators') point of view. The modelling phase produced interesting and informative outputs for subsequent review and presentation across the target groups involved (course guarantors, faculty management).

The main benefit of the structured evaluation of the quality of the medical curriculum was its comprehensiveness and objectivity regarding the overall setting of who would evaluate what, how and how many times.

REFERENCES

[1] Komenda M, Víta M, Vaitsis C, et al. Curriculum mapping with academic analytics in medical and healthcare education. PLoS One. 2015;10(12):e0143748.

[2] General Medical Council. Tomorrow's doctors [Internet]. 2003 [cited 16 Jan 2023]. Available from: https://www.educacionmedica.net/pdf/documentos/modelos/tomorrowdoc.pdf.

[3] Brown R. Quality assurance in higher education: The UK experience since 1992. Routledge; 2004.

[4] Simpson JG, Furnace J, Crosby J, et al. The Scottish doctor--learning outcomes for the medical undergraduate in Scotland: a foundation for competent and reflective practitioners. Med Teach. 2002;24(2):136–43.

[5] Metz JCM, Verbeek-Weel AMM, Huisjes HJ. Blueprint 2001: training of doctors in The Netherlands [Internet]. 2001 [cited 10 Jan 2023]. Available from: http://www.medidak.de/de/didaktik/tbl/Dutch_blueprint.pdf.

[6] Bloch R, Bürgi H. The Swiss Catalogue of Learning Objectives. Med Teach. 2002;24(2):144–50.

SECTION B

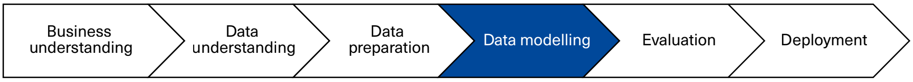
03

USER BEHAVIOUR ANALYSIS AND INTERACTIVE VISUALISATION IN A CURRICULUM MANAGEMENT SYSTEM

Matěj Karolyi, Martin Komenda

Parts of the authored thesis Karolyi M. User Behaviour Analysis in Curriculum Management System [M.Sc. thesis]. Brno: Masaryk University, Faculty of Informatics; 2018. Available from: https://is.muni.cz/th/m9sd8/. Thesis supervisor: Martin Komenda.

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2012–2024
Keywords	User behaviour, data analysis, curriculum management system, OPTIMED
Research question	What are the benefits of online visual analytics tools reporting user behaviour data in the given curriculum management system?
Type of result	Interactive visualisation
Level of data processing	Descriptive statistics

DATA TO DOWNLOAD

INTRODUCTION

It is a crucial priority to keep the quality of medical and healthcare education at the same quality level and ensure its efficiency for more students than ever before. Such a situation requires the development of supporting digital tools and learning materials in the form of educational portals and complex curriculum management systems. These systems are often complemented by information on user behaviour: which actions users perform, which parts of portals are used at different times of the day, and whether they reach their intended goals. All these details are primarily private and are used by product owners only to optimise their business objectives. However, we decided to publish this information publicly within the content of the OPTIMED curriculum management system [1].

This case study is focused on the analysis of user behaviour in OPTIMED. The OPTIMED curriculum management system is a web-based platform for harmonising medical and healthcare curricula effectively. With the curriculum described by this platform, curriculum designers and guarantors of courses can offer a well-arranged set of systems (mandatory and optional) and identify possible duplications or overlaps across medical and healthcare disciplines in education [1,2]. Approaches to the data collection and visualisation are shown, as well as the final results of the analysis above – the collected and processed data, presented user tracking, and analysis concepts applicable in various spheres of human activities in the online environment.

The motivation to collect data about users' actions is significant. Information about what users do, where they look or click, and what makes them leave a website is very valuable. If data about all these factors are collected and stored, it is possible to analyse them. Keeping all data is unnecessary; only a subset or aggregated results are suitable. The outputs of these types of research usually serve as an excellent foundation to help to form plans and future strategies or make changes within portals or systems.

As regards the user behaviour analysis and visualisation of its outputs, several steps are necessary. The first action is data extraction, where a data source is accessed to obtain raw data. The obtained dataset is subsequently processed and cleaned, resulting in a well-structured dataset. In the next step, invalid and irrelevant entries are removed from the dataset. This usually depends on the purpose of the particular analysis. Subsequently, an evaluation is performed. Again, it mainly depends on the type of analysis – used methods, involved experts, and several iterations. The final phase consists of modelling outputs, visualising them using the appropriate techniques, and respecting previously defined rules. In this case, the user behaviour analysis is part of the more extensive reporting module of OPTIMED. The OPTIMED Reporting module shows statistics about the teaching volume, the educational content, and some summary overviews for stakeholders.

AIMS

- To identify an audience for the OPTIMED curriculum management system. From whom is the audience composed?
- To identify when the OPTIMED platform is most frequently used, in terms of the time of the day concerning the academic year.
- To discover the most frequently searched words and phrases.

METHODS

We have proceeded with the data acquisition process using a standardised and steady methodology called CRISP-DM [3,4]. It significantly simplifies and speeds up the whole process, makes it safer and helps to avoid known pitfalls. The reporting module[1] is formed as a Symfony project [5]. Usage of this PHP language framework accelerates the development of individual requirements. Another advantage is using embedded functionality and keeping a good project structure. And finally, the framework provides good security (especially in database queries). The Symfony project at version 2.x is the foundation of the OPTIMED reporting module. At the start of implementation, it was the latest long-term supported (LTS) version. In case of upgrading to the newest Symfony version, some parts had to be refactored. Documentation pages contain more information about migrations to higher versions [6].

RESULTS

BUSINESS UNDERSTANDING

Understanding a specific domain is the starting point of any analysis. In our case, the domain is behaviour analysis and all its aspects. We can speak about best practices of website composition, using standard building blocks of websites as they are meant to be used, the definition of the focus user group and so on. During the design and development of the OPTIMED platform, multiple functionalities were defined, which are now provided by the curriculum management

[1] https://opti.med.muni.cz/new/cs/reporting/web/

system. User behaviour analysis within the forum is vital to provide information on whether the intended objectives meet reality.
We wanted to examine and measure the attributes listed below. Some of them helped us understand the audience composition; some told us something about the times of most frequent use; but most of all, we could see patterns in user searches.

The mentioned attributes were:

- Who are the users of the platform?
- How much do users search within the platform depending on their academic status (students, academics)?
- When is the platform most used (during the academic year, year-on-year, during the day)?
- Which phrases are most frequent?
- Which modes of searching are most frequently used?
- Which learning units describing the curriculum are visited by users and which are not?

DATA UNDERSTANDING

The data model for user analysis behaviour is implemented inside of the data model of the whole OPTIMED curriculum management system. The entire model enables curriculum development and management. Some entities are developed exclusively for user behaviour analysis. Information from them is completed by other entities whose primary purpose is different. The entity-relationship diagram (see Figure 1) shows all entities which hold data for current and future user behaviour analysis in the OPTIMED curriculum management system. Entities coloured in grey are entities not designed and implemented just for user behaviour analysis.

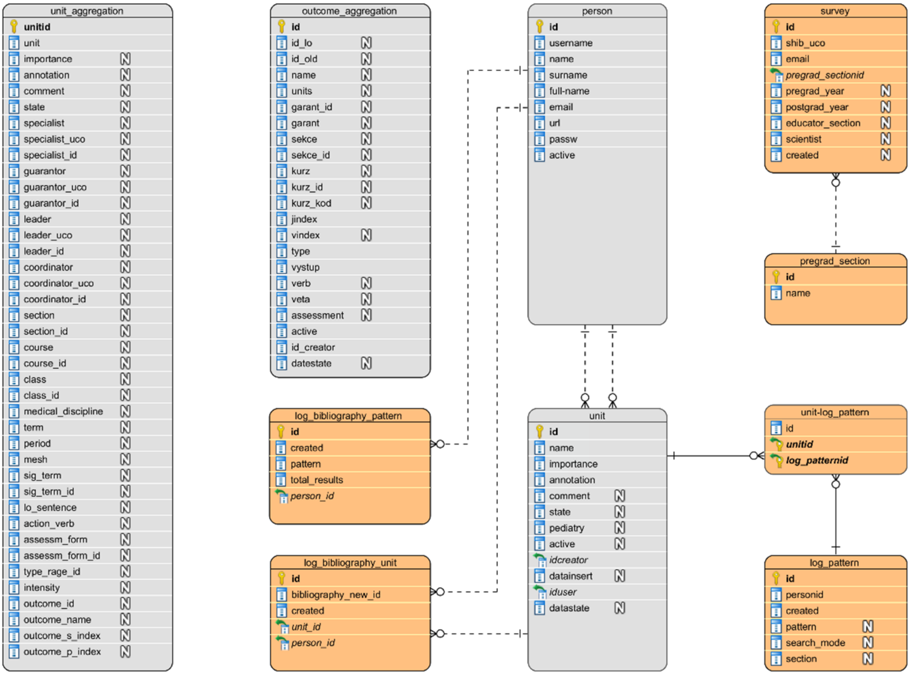


Figure 1: The entity-relationship diagram

DATA PREPARATION

The data layer of the user behaviour module consists of six primary database tables and four supplementary database tables:

- survey – every user who comes to the platform for the first time must fill in the survey. The information from the initial study is used in further analysis for sorting users into groups.
- pregrad_section – list of sections to which the user may be assigned. Every team can be a user‘s specialisation or area of interest.
- log_pattern – log of searches in the curriculum browser. A query, search mode and time is stored during every performed search.
- log_unit – log of openings of learning units. Every time users open a learning unit detail, an identification of the learning unit and previous search query is stored.
- log_bibliography_pattern – log of searches in the bibliography.

- log_bibliography_unit – log of openings of learning units through the bibliography.
- unit – learning unit entity is used for essential attributes (e.g. name of learning unit).
- person – person entity is used for its association with a survey.
- unit_aggregation – aggregated data of learning units (linked courses, linked medical disciplines, etc.).
- outcome_aggregation – aggregated data of learning outcomes/competencies (linked courses, linked learning units, etc.).

The user behaviour module of the OPTIMED reporting module and the whole OPTIMED platform runs above PostgreSQL [7] in version 9.5. Version 10 was released, and migration to this version is being considered. PostgreSQL is very suitable for this purpose. Its configuration is easy, and in most cases, the default settings are appropriate. Besides PostgreSQL, we run a monitoring system called MUNIN [8]. Thanks to obtaining statistics about operational procedures and databases, we can optimise database queries or reveal potential problems in the future.

MODELLING

The final deliverable is a sub-module of the OPTIMED reporting for analysing user behaviour. It consists of two main parts: (i) the user tracking mechanism, which is fully integrated into the core of the OPTIMED curriculum management system and (ii) the reporting and visualisation part, which shows the outcomes of all analyses above-collected data. Complete tracking mechanisms and final reporting are deployed on all existing production instances. The current state of analysis (at the time of user access to the page) is presented to the user in the form of a wide variety of interactive graphs with additional information that would help them focus on the exciting parts of the analysis.

The whole module of OPTIMED Reporting consists of more sub-modules, and the user behaviour analysis is one of them, divided into smaller unit. Some of them are straightforward, while others are more complex. Much attention was paid to visualisations so that they would show the data clearly, would not misinterpret the contained information and would not distract readers by excessive, unnecessary decorations. There are two main approaches to looking at the obtained data: (i) audience perspective – who are users when they use the platform; (ii) searching perspective – what users search and what results they get.

We have created the following graphical overviews for both approaches that are freely accessible from web browsers[2]:

- Infographics of complex audience analysis: Basic facts describing OPTIMED users and their behaviour in storytelling-oriented infographics.
- Audience analysis: An overview of who uses the OPTIMED curriculum browser and the complete audience composition.
- Search analysis: An overview of the most used search modes and the most frequent phrases.
- Time distribution of search queries in the browser: Analyses of the most demanding part of the day and overview of what time of day the curriculum browser is most used.
- Timeline of queries: A flat stacked graph shows browser utilisation from its inception to today.
- Annual comparison of browser queries: Stacked or grouped bar chart presents the history of search queries.

Two selected topics that document the audience and search analysis are described in detail below for demonstration purposes.

AUDIENCE ANALYSIS

The main outputs of the audience analysis are shown in part called a complex analysis of the audience. This web page should be the entry point for every reader looking at the user behaviour analysis for the first time. It gives readers complex information about the system and uses data storytelling techniques [9,10] At the bottom of this web page are links to other types of analyses. Therefore, the user can continue to the area that interests them. One of the exciting graphics is dedicated to the number of user searches in the curriculum content depending on the academic year (see Figure 2). The OPTIMED platform is mainly used at the start of the academic year. Over the rest of the academic year (yellow colour), as well as during the examination period (red colour). On the other hand, the traffic is relatively low during the holiday (green colour).

[2] https://opti.med.muni.cz/new/cs/reporting/web/o-uzivatelich/komplexni-analyza-publika/

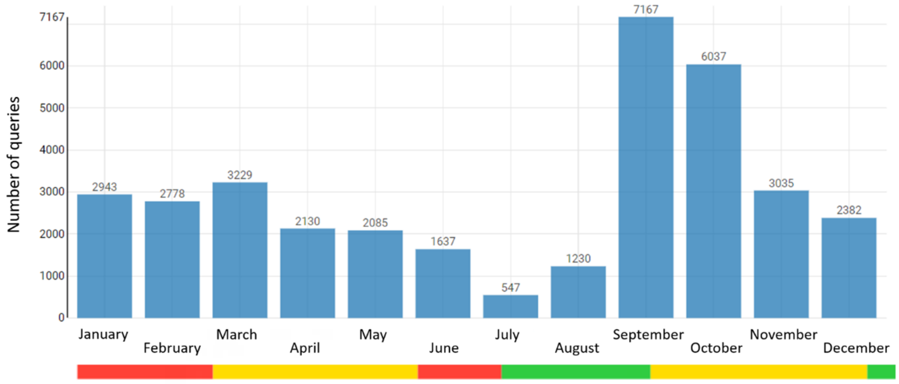
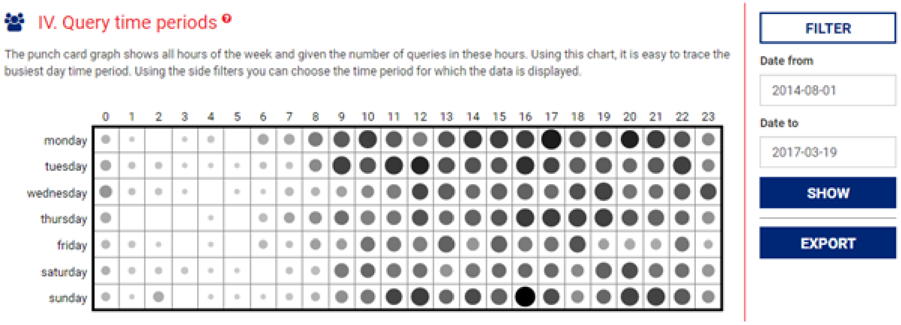


Figure 2: Searches within the OPTIMED platform depending on the academic year

Thanks to other audience analyses, users can see how many students and employees (teachers, curriculum designers, etc.) use the OPTIMED browser or how many search queries are performed during each hour of the week (see Figure 3). Using filters on the right side, it is possible to set the specific time extent.



IV. Query time periods

The punch card graph shows all hours of the week and given the number of queries in these hours. Using this chart, it is easy to trace the busiest day time period. Using the side filters you can choose the time period for which the data is displayed.

Date from: 2014-08-01

Date to: 2017-03-19

Figure 3: The punch card graph of user accesses during the week

SEARCH ANALYSIS

The OPTIMED curriculum browser is intended to help students find learning units after entering the search phrase. The most frequently entered words are interesting for our purposes of search engine improvements. The search phrase could be a medical keyword, a teacher's name or code of an enrolled course. The entered word is searched in the whole content of each learning unit, and all other elements (learning outcomes, persons, study materials, etc.) that are connected with it. The final results are then returned to the user, sorted by their relevance to the entered query. The number of incomplete learning units of particular questions was an interesting phenomenon we could observe. The ten most frequently used search phrases (translated to English) of one instance of the OPTIMED platform are listed in the table below (see Table 1). These results showed that more than half of the top 10 search phrases were codes of courses in which students were enrolled, and therefore, they looked for information connected with a specific subject.

Table 1: The ten most frequent search phrases

Search phrase	Percentage of occurrence
"no phrase"	12.30%
anatomy	11.09%
biophysics	8.56%
VSPF0622p	4.88%
VLBF011p	4.77%
VLON091	4.18%
pathological physiology	4.15%
VLNE9X1p	3.64%
VLFA0722p	3.48%
VSBF011c	3.12%

EVALUATION

Each analysis is created as a separate web page. After its creation, the outputs are added to the portfolio of existing and accessible visualisations. In general, the Symfony framework helps write code in a standardised form, suitable for further maintenance and extension. When the whole creation process is completed, the result must be opposed. The selection of opponents depends on the domain and complexity of created visualisation (curriculum designers, medical informatics specialists, guarantor of a particular course, etc.).

All created reports underwent several levels of complex review: (i) technical in-house user testing (the functionality and visual style of the web presentation is correct and optimised for most of the modern internet browsers), (ii) content-based internal validation of the available data (raw input data corresponds to the final reports), (iii) external review by the guarantor of study programmes (evaluation of the added value of data visualisations in the context of decision support).

DEPLOYMENT

After the complete design, development, and implementation process of these interactive visualisations, the sub-module was deployed as a new feature of the OPTIMED portal. The reporting tools module was extended with a section dedicated to user behaviour analysis. The extension was manageable thanks to the appropriate choice of technologies during the development. The maintenance of this sub-module is part of the overall comprehensive solution for curriculum management and optimisation, which helps present the available data comprehensibly and efficiently across the curriculum.

DISCUSSION

This case study is focused on the behaviour analysis of the audience of the OPTIMED platform, which is the curriculum management system developed at Masaryk University. The goal is to show the sub-module, which helps to describe the platform's audience, understand their activities within the forum and provide a set of analyses which help to optimise the platform in the future or use these concepts elsewhere. The application which is described has already been running in three production instances.

EVALUATION OF THE AIMS OF THE CHAPTER

- To identify an audience for the OPTIMED curriculum management system. From whom is the audience composed?
  - This case study allowed us to determine the primary distribution of users (students, teachers) and their behaviour when working with the OPTIMED browser, which is used for a comprehensive search of the curriculum of the General Medicine study programme. In addition, it was also possible to distinguish the use of the different search modes (full text, only course, only person, only study material etc.).

- To identify when the OPTIMED platform is most frequently used depending on the time of the day concerning the academic year.
  - The frequency of use of the OPTIMED browser depends on the academic year period. One of the outputs clearly shows the number of queries entered into the browser in each month of the year, including overlaps with different parts of the academic year (semester, examination period, holidays). Visualising the hour distribution of searches shows the days and hours of the week and the number of entered queries.

- To discover the most frequently searched words and phrases.
  - A detailed keyword analysis defining the most common keywords (e.g. course code, lecture topic, important term) was an integral part of this task. It was thus possible to prepare an analysis of the teaching coverage (depending on whether the search term occurred in the curriculum description) concerning the most frequently asked queries in the OPTIMED browser.

LESSONS LEARNED

The outputs of this project are an extension of the OPTIMED portal and the data it contained. The main added value is a clear and understandable view of user behaviour throughout the portal. This provided the development team with a very valuable and necessary tool for a comprehensive overview, including a detailed analysis of who searches for what and when. These outputs provided a valid basis for subsequent changes, innovations and optimisations to the portal based on real data.

The main benefit of the online visualisations over the data describing OPTIMED user behaviour was the reusability and timeliness of the reports, which were primarily used for internal needs to modify existing and design and develop new functionalities.

REFERENCES

[1] Komenda M, Schwarz D, Vaitsis C, Zary N, Štěrba J, Dušek L. OPTIMED Platform: Curriculum Harmonisation System for Medical and Healthcare Education. Stud Health Technol Inform. 2015;210:511–5.

[2] Komenda M. Towards a Framework for Medical Curriculum Mapping [Ph.D Thesis]. Brno: Masaryk University, Faculty of Informatics; 2016. Available from: https://is.muni.cz/th/hcl4g/.

[3] Azevedo A, Santos, M F. KDD, semma and CRISP-DM: A parallel overview. In: Abraham AP (ed). Proceedings of the IADIS European Conference Data Mining, Amsterdam, The Netherlands, 24–26 July 2008. IADIS; 2008. p. 182–185.

[4] Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R. CRISP-DM 1.0 Step-by-step data mining guide [Internet]. CRISP-DM consortium; 2000 [cited 6 Jun 2023].
Available from: http://www.statoo.com/CRISP-DM.pdf.

[5] About Symfony Project [Internet]. Symfony SAS; 2018 [cited 6 Jun 2023].
Available from: https://symfony.com/what-is-symfony.

[6] Upgrading a Major Version (e.g. 3.4.0 to 4.0.0) (Symfony Docs) [Internet]. Symfony SAS; 2018 [cited 6 Jun 2023].
Available from: http://symfony.com/doc/current/setup/upgrade_major.html.

[7] PostgreSQL: About [Internet]. 2018 [cited 8 March 2023].
Available from: https://www.postgresql.org/about/.

[8] Munin. [Internet]. 2018 [cited 8 March 2023].
Available from: http://munin-monitoring.org/.

[9] Boy J, Detienne F, Fekete JD. Storytelling in information visualizations: Does it engage users to explore data? In: Begole B, Kim J, Inkpen K, Woo W (eds). Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery; 2015. p. 1449–1458.

[10] Kosara R, Mackinlay J. Storytelling: The next step for visualization. Computer 2013;46(5): 44–50.

SECTION B

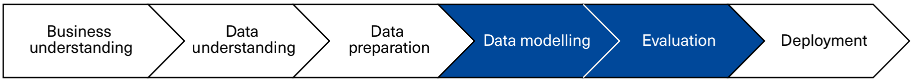
| 04 |

DETECTION OF INTERSECTIONS BETWEEN CURRICULUM MAPPING AND VIRTUAL PATIENT INFORMATION SYSTEMS

Martin Komenda, Petr Štourač, Daniel Schwarz

The case study was supported by the following Erasmus+ grant projects: (i) MERGER project – reg. no. MUNI/A/1339/2016 funded from the Grant Agency of the Masaryk University; (ii) Masaryk University Strategic Investments in Education SIMU+ (CZ.02.2.67/0.0/0.0/16_016/0002416) funded from the European Regional Development Fund; (iii) Masaryk University 4.0 (CZ.02.2.67/0.0/0.0/16_015/0002418) funded from the European Social Fund.

CRISP-DM CRUCIAL PHASES



Business understanding

Data understanding

Data preparation

Data modelling

Evaluation

Deployment

GENERAL INFORMATION

Year	2017–2017
Keywords	Virtual patients, curriculum mapping, natural language processing, machine learning
Research question	How to design and use in practice an algorithm to automatically determine the overlaps between virtual scenarios and the curriculum description, including the implementation of feedback.
Type of result	Interactive visualisation
Level of data processing	Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

This case study focused on information science, applied informatics and biomedical engineering to develop methods for automatically detecting links between information systems to support clinically oriented stages of study in medical and health disciplines. The MERGER project, which addressed this issue, took advantage of a large amount of textual data available in the local systems of the Faculty of Medicine, Masaryk University (MED MUNI) for creating and mapping the medical curriculum and in systems for creating and playing virtual patients. This pilot project followed the author's previously published work [1–6], where fundamental aspects of a curriculum mapping domain and virtual scenarios framework were highlighted. The project team analysed a large amount of text data collected in systems for medical curriculum management and mapping (e.g. OPTIMED[1]) and systems for virtual patient creation (e.g. AKUTNE.CZ[2])—the developed detection algorithms aimed to employ natural language processing techniques and machine learning methods.

The underlying idea was (i) to use appropriate machine learning and natural language processing or analytical methods to reduce the capacity associated with the time-consuming manual work required for annotating content entities (e.g. learning units or learning outcomes in a local medical curriculum management system) or linking them to entities in systems supporting clinical reasoning (e.g. linear or branching interactive algorithms and virtual patients), (ii) to establish an evaluation/feedback recommending system for students. Moreover, the results of manual annotation or relating between different information systems may be biased by the subjective approach of annotators. Therefore, it was necessary to replace manual work with automatic algorithms. The evaluation phase actively involved students of MED MUNI at the undergraduate level of study. They participated in consolidating interesting datasets from different learning information systems, and they provided valuable feedback in determining the accuracy of this detection.

AIMS

- To understand the structure and all relevant descriptive characteristics of medical and healthcare curriculum building blocks and virtual patient records.

[1] https://portfolio.med.muni.cz/

[2] https://www.akutne.cz/

— To propose proper methods to identify text-based similarity between medical curriculum and virtual patients.

— To provide an accurate evaluation of achieved analytical results.

METHODS

Most of the experience related to methodological background came from the educational network of all medical faculties in the Czech Republic and Slovakia called MEFANET[3] (Medical Faculties Network). Through this unique system for electronic publishing of educational content, academic institutions focused on teaching medical and healthcare study programmes shared and offered each other the created content and centralised metadata about this content in one place, the so-called central gateway[4] of the MEFANET network. Specialised extensions and dedicated projects were gradually added to the electronic publishing system. In particular, the MERGER project focused on an existing component for the so-called serious games, which collected data on learning objects that work in some way with simulation medicine – usually virtual patients, and more broadly virtual scenarios. Generally, virtual scenarios are more comprehensive in scope compared to virtual patients alone.

MEDICAL AND HEALTHCARE CURRICULUM

The information system for mapping medical and healthcare curricula started in 2013, then as part of the OPTIMED project. Today, as part of the SIMUportfolio platform, the information system provides curriculum managers with complex overviews and basic visual analytics needed for introducing innovations into the selected study programmes and medical disciplines. Crucially, the curriculum description was fully standardised according to the internationally accepted recommendations of the MedBiquitous association[5]. For the MERGER project, the OPTIMED platform provided a large text file aggregating data on learning units, learning outcomes and other annotation entities (keywords from the MeSH glossary, important terms, etc.). For this case study, sets of data in the English language were extracted (1,360 learning units described by approximately 2,600 standard pages of text).

[3] https://www.mefanet.cz/

[4] http://portal.mefanet.cz

[5] https://www.medbiq.org/standards

VIRTUAL PATIENTS

Virtual patients are clinical scenarios that play out on the computer screen. The student interacts with the patient (the computer) by selecting the most appropriate choices for further patient management or scenario progression, thereby requesting information on physical examination findings and laboratory tests. The computer supplies patient responses or other requested information. Students are typically required to commit to a diagnosis and management plan at some point [6]. For this case study, the AKUTNE.CZ platform provided virtual patients were used to verify the accuracy of developed algorithms quantitatively (77 of the virtual patients were described by approximately 550 standard pages of text in total).

NATURAL LANGUAGE PROCESSING

Natural language processing and subsequent exploratory similarity analysis were used to reduce the time-consuming manual agenda of processing, transforming and mapping curriculum content entities (descriptions of learning units and outcomes from the OPTIMED system) and their linking to virtual patient entities (interactive algorithms from the AKUTNE.CZ portal). The main reason is that the results of manual annotation or the creation of links between different objects from different information systems can be significantly distorted by evaluators' subjective approaches and opinions. Therefore, it was desirable to replace manual work with automatic algorithms. Using the selected distance and similarity metrics, it was possible to identify how the different records (learning units and virtual patients) are linked; for this purpose, keyword frequency vectors based on the standard corpus of the OPTIMED system and AKUTNE.CZ were used. The project was based on several successfully implemented classification and natural language processing tasks over text datasets.

TECHNICAL BACKGROUND

The development toolkit was based on the R programming language and software tools. Data pre-processing, analyses and visualisations were performed in R using special libraries and packages (readxl, dplyr, tm, proxy, visNetwork, data.tree, DiagrammeR, Shiny, shinydashboard). Afterwards, an interactive web-based visualisation toolbox was developed through the R Shiny package, which was available on CRAN, and made it easy to build a standalone online application that benefits from the computational power of R and the interactivity of the modern web [5, 7, 8].

RESULTS

BUSINESS UNDERSTANDING

The research team adequately included all the necessary multidisciplinary roles (senior guarantor of study programmes, technical guarantors of both platforms providing input data, computer scientist, analyst, developer and student). As a result, the understanding phase was quickly resolved. English versions of the curriculum and virtual scenarios were chosen to use natural language processing methods.

DATA UNDERSTANDING

The technical guarantors of the platforms (selected curriculum pieces (Table 1) and virtual patients (Table 2)) knew in detail the structure of descriptive attributes, individual relations, and database models of both systems. Initial data collection and retrieval required information from databases in a designed form were seamless. Fields of medical disciplines were chosen to standardise the results to the existing study programmes at MED MUNI, which easily allows the categorisation of learning units to a specific field (i.e. anatomy, laboratory diagnostics or neurosurgery).

Table 1: A complete description of one learning unit exported from the OPTIMED system

Learning unit	Venous surgery
Importance	In general, the varicose veins of the lower extremities are one of the most common surgical diseases. Severe venous insufficiency affects about 20% of the adult population. A hazardous disease of the deep venous system is deep vein thrombosis. Surgery often occurs as a complication during other procedures but may have serious consequences. The creation of vascular access for dialysis represents specific problems of venous surgery. The teaching unit aims to introduce the students to the issues of varicose veins of the lower limbs and chronic venous insufficiency in terms of vascular surgery. It will present the case of deep vein thrombosis, particularly surgical procedures, and a basic overview of possibilities to provide vascular access for patients in chronic hemodialysis programs. The students will also learn about congenital fistulas and surgical problems of peritoneal dialysis catheters in connection with artificial hemodialysis arteriovenous fistulas.
Annotation	Varicose veins of the lower limbs, chronic venous insufficiency etiopathogenesis of primary and secondary varicose veins clinical picture CEAP classification diagnosis treatment conservative surgical (sclerotherapy, standard operations, endovenous laser therapy – EVLT, radiofrequency ablation – RFA) Deep vein thrombosis ...

Indexes	Artificial arteriovenous fistula for dialysis Congenital fistulas Examination, diagnosis, treatment Medical history Physical examination
Learning outcomes[6]	Examination of a patient with venous system disease. Indication, fundamentals of technical implementation and use of AV shunts. Principles of diagnosis and treatment principles. Conservative and surgical treatment of venous diseases. Typical symptoms of venous disease and obtaining a targeted medical history.

Table 2: A complete description of one virtual patient exported from the AKUTNE.CZ platform

Virtual patient	Surviving sepsis
Annotation	Sepsis is the systemic inflammatory response syndrome of an organism caused by infection. It can easily extend to a stage of severe sepsis and septic shock with the signs of multiple organ dysfunction, even failure requiring extensive organ function support. An increasing number of septic patients reverse the fact that the number of predisposed individuals in various ways increases patients suffering from immunosuppression and undergoing invasive diagnostic and therapeutic procedures. Early diagnosis and immediate and accurate therapy are crucial.
URL	https://www.akutne.cz/algorithm/en/66-surviving-sepsis/
Keywords	Antibiotics Volumotherapy Monitoring
Introduction	A 28-year-old man was brought to the emergency department. He is conscious, worn out, BP 90/50, regular heartbeat, sinus tachycardia 120 /min, tachypnoea 25 / min, and SpO_2 89 %. From the history: the patient has not been seriously ill, takes no medication, and denies any allergies. He has been complaining about a sore throat for a few days. The patient's overall condition has worsened rapidly throughout the day. He is gasping, has chest pain on the right side, expectorates purulent sputum and has a temperature of up to 40 °C. You suspect him of being septic. Your first step is going to be...
All text descriptions of scenario nodes and all proposed answers.	

DATA PREPARATION

This phase primarily covered table, record, descriptive attributes selection, data aggregation, transformation, mapping, and pre-processing. A set of steps had to be performed for machine processing of the input data: (i) Removing HTML tags, punctuation, words from the stop-word list, and building a text corpus. (ii) Text-based objects in the form of virtual patients and learning units were transformed into numeric vectors, where each element of a given vector indicated the frequency of occurrence of a word of a given index. The index word was common to both source datasets due to the same text corpus.

[6] After the text processing procedure.

MODELLING

The goal of the modelling/ exploring similarity phase was to determine the similarity between a particular virtual patient (its textual representation) and the learning units (its textual representation) based on the frequency of occurrence of keywords. The modelling phase was divided into three linked steps: (i) vector-space representation of available text corpus, (ii) similarity computation, (iii) graph creation. In general, there are several approaches to determining the overall similarity between texts. Experimentally, one of the proven analytical techniques was used during the modelling phase. Namely, exploratory distance/ similarity analysis (example of = unsupervised machine learning) determined how similar the virtual patients and learning units are based on selected metrics (extended Jaccard coefficient, Pearson correlation coefficient, cosine distance). The first two approaches returned absolute 100% similarity to a greater extent. Moreover, the match pairs (virtual patient, learning unit) were identical for both metrics. For this reason, cosine distance was chosen as the proper one for the pilot case study purposes. Using the thresholding method, it was then possible to quantify these relations. In this case, no student evaluation data was needed. Each virtual patient's five most similar learning units were identified as a static overview table and graph visualisation. These interactive graphs displayed the relations/links between virtual patients and learning unit objects. Figure 1 shows the five most similar learning units for a virtual patient called "Surviving sepsis".

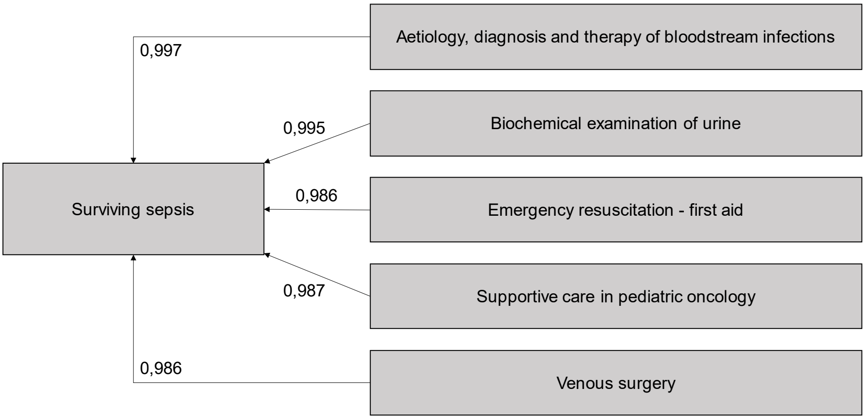

Aetiology, diagnosis and therapy of bloodstream infections
0,997
Biochemical examination of urine
0,995
Surviving sepsis
0,986
Emergency resuscitation - first aid
0,987
Supportive care in pediatric oncology
0,986
Venous surgery

Figure 1: Similarity graph shows the five most relevant learning units to a virtual patient called "Surviving sepsis"

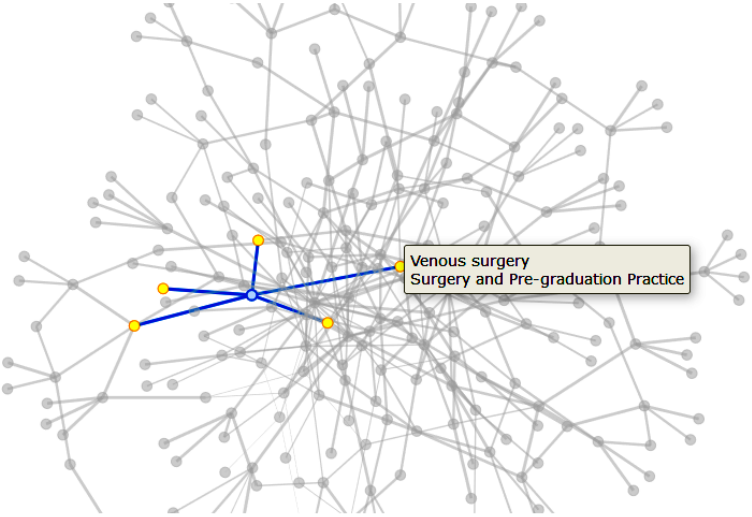


Figure 2: This interactive network graph shows 77 virtual patients from AKUTNE.CZ and all related learning units from the OPTIMED portal (a zoom-in/out function is available). The virtual patient is called "Surviving sepsis".is highlighted by a blue node, while yellow nodes highlight the most relevant learning units. A tooltip shows the title and course of the appropriate learning unit

EVALUATION

The evaluation board consisted of 12 students, mainly from the most extensive General Medicine programme and from different years across MED MUNI. Each virtual scenario (virtual patient) had two independent evaluations. Each student evaluated ten algorithms. Google Forms platform was used as the technical solution, where a structured questionnaire was created to collect individual ratings using the Likert scale. Students indicated the relation (Table 3) of appropriate learning units to a given virtual patient (the strength of this relation was defined on a scale of 1 (no similarity) to 5 (total similarity). Based on this evaluation, the results obtained from the machine similarity identification using the selected metric were then checked.

Table 3: Example of one evaluation record.

Discipline	Diagnostic Imaging Methods
Learning unit	Protection against radiation, the principle of skiagraphy and skiascopy
Virtual patient	Craniocerebral trauma
Evaluation score	3

DEPLOYMENT

After several iterations of development, testing and involvement of students as evaluators, the outputs were deployed and published in several formats to maximise readability for the selected target groups: (i) a comprehensive table in MS Excel format, where all links and similarities are visible, including student ratings, (ii) a complete similarity matrix of all objects from both systems (OPTIMED, AKUTNE. CZ), (iii) an interactive visualisation showing the network of virtual patients with the most similar learning units, (iv) a set of conference papers and publications describing the achieved results of the MERGER project.

DISCUSSION

In accordance with the proven CRISP-DM methodology, the different phases were addressed step by step; at the same time, several problems were highlighted, which were invisible and unrecognisable at first sight. (i) The reasons for this were both the complicated path to export the required data and the overall size of the aggregated texts. In order to use similarity algorithms effectively, it is advisable to have the corpus as large as possible. (ii) The modelling phase was somewhat experimental, as it was not known in advance which metric or similarity coefficient would perform best on the available data. It turned out that using multiple options and then comparing the results did not pose a problem, and that the resulting functional metric of cosine distance led to valid results when recommending similar content. (iii) Using the R package Shiny, it was possible to develop and implement an interactive web visualisation with CSS themes, HTML widgets and JavaScript actions. Due to long-term compatibility with the OPTIMED portal, which has been replaced by the SIMUportfolio platform, this visualisation is only available locally.

The results were also available in the static form of a similarity matrix; in this case, the aim was to use them for the necessary evaluation phase, with the active involvement of MED MUNI students. This involvement proved to be a very positive experience that produced valuable and valid results. Overall, the MERGER project brought an innovative perspective on the use of proven advanced methods for natural language processing combined with selected machine learning approaches, which were appropriately complemented by the human factor in the form of a comprehensive structured evaluation.

EVALUATION OF THE AIMS OF THE CHAPTER

- To understand the structure and all relevant descriptive characteristics of medical and healthcare curriculum building blocks and virtual patient records.
  - For this case study, data were collected from the OPTIMED curriculum management and mapping system (1361 learning units) and the AKUTNE.cz interactive education platform (77 virtual patients). Both data exports fully reflected the required structure and content for the subsequent similarity analysis.
- To propose proper methods to identify text-based similarity between medical curriculum and virtual patients.
  - Several proven approaches to identifying the similarity of objects (learning units and virtual patients) using keywords have been experimentally verified. The selected metric was used as the basis of an algorithm for automatic similarity detection.
- To provide an accurate evaluation of achieved analytical results.
  - To obtain relevant feedback, 12 students from MED MUNI were involved in the evaluation process. They evaluated the similarity between the virtual patients and the five recommended learning units through a structured form.

LESSONS LEARNED

This project brought an interesting experience with machine learning methods in identifying content overlaps between the declared teaching (structured curriculum description on the OPTIMED portal) and specific study materials (virtual clinical case studies AKUTNE.cz). Mainly due to the implementation of these two projects and the active involvement and support of guarantors of these activities, the initial phases of the CRISP-DM model were relatively trivial. The most complex phase was undoubtedly the modelling phase, where a thorough survey of existing methods suitable for keyword-based similarity determination over available texts was required. Subsequently, the identified methods were piloted and the results were evaluated in order to select the approach that returns the best results.

Experimentally, based on the available literature and in collaboration with senior mentors, the selected machine-learning methods were implemented in practice to identify similar learning fragments that correspond on the basis of textual similarity to the selected learning materials.

ACKNOWLEDGEMENT

Special thanks to our colleagues Roman Vyškovský and Jakub Ščavnický, who contributed significantly to the modelling phase, especially to the selection, adoption and application of similarity metrics over the available data.

REFERENCES

[1] Schwarz D, Štourač P, Komenda M, et al. Interactive algorithms for teaching and learning acute medicine in the network of medical faculties MEFANET. J Med Internet Res. 2013;15(7):e135.

[2] Komenda M, Víta M, Vaitsis C, et al. Curriculum Mapping with Academic Analytics in Medical and Healthcare Education. PLoS One. 2015;10(12):e0143748.

[3] Karolyi M, Komenda M, Janoušová R, et al. Finding overlapping terms in medical and health care curriculum using text mining methods: rehabilitation representation – a proof of concept. MEFANET J. 2017;4(2):71–7.

[4] Komenda M, Karolyi M, Vyškovský R, et al. Towards a keyword extraction in medical and healthcare education. In: 2017 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE; 2017. p. 173–6.

[5] Komenda M, Ščavnický J, Růžičková P, et al. Similarity Detection Between Virtual Patients and Medical Curriculum Using R. Stud Health Technol Inform. 2018;255:222–6.

[6] Cook D A, Triola M M. Virtual patients: a critical literature review and proposed next steps. Med Educ. 2009;43(4):303–11.

[7] Karolyi M, Scavnický J, Komenda M. First Step Towards Enhancement of Searching Within Medical Curriculum in Czech Language using Morphological Analysis. In: CSEDU 2018 – Proceedings of the 10th International Conference on Computer Supported Education. 2018. p. 288–293.

[8] Ščavnický J, Karolyi M, Růžičková P, et al. Pitfalls in users' evaluation of algorithms for text-based similarity detection in medical education. In: 2018 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE; 2018. p. 109–116.

SECTION B

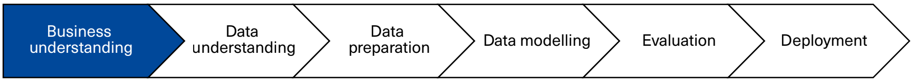
05

MEDICAL CURRICULUM INNOVATIONS USING TECHNOLOGICAL STANDARDS

Martin Komenda, Dimitris Spachos, Christos Vaitsis, Luke Woodham, Panos Bamidis

The case study was supported by the following grant projects: MEDCIN – Medical Curriculum Innovations – reg. no. 2015-1-CZ01-KA203-013935; this project has been funded by the European Commission ERASMUS+ programme. OPTIMED – optimised medical education: horizontal and vertical connections, innovations and efficiency in practice – Reg. No. CZ.1.07/2.2.00/28.0042, which was funded by the European Social Fund and the national (state) budget of the Czech Republic.

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2015–2017
Keywords	Outcome-based curriculum, standardisation, standard-compliant systems, curriculum comparison
Research question	How does standardise medical and healthcare curricula while you need to explore details in the form of gaps, overlapping areas, and similarities?
Type of result	Interactive visualisation
Level of data processing	Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

A standardised curriculum, particularly in medical and healthcare education, is indispensable. A comprehensive platform covering all necessary instruments for easy in-depth curriculum management is crucial. For any medium to achieve these goals, data standards must enable the systems to communicate across organisations and implementations. The MEDCIN (Medical Curriculum Innovations) project intended to focus on these standards developments through MedBiquitous Consortium[1], the leading organisation in developing and promoting technology standards for the health professions. MEDCIN aimed to systematically support the medical and healthcare educational process by the unification of theoretical background (existing standard-based methodologies), reform endeavour (series of agreements at the European higher education area commonly known as the Bologna Process [1]) and powerful technological performance, and ensure more comparable, compatible and coherent systems of higher education in Europe.

MEDCIN proposed an innovative methodological background, including a web-based visualisation tool for a comprehensive evaluation and a map of standard-compliant medical and healthcare curricula with modern information and communication technologies. Using MedBiquitous standards[2] ensures broader applicability beyond the partnership for MEDCIN's approach and model, disseminating more widely as an example of best practices across Europe. The various standard-compliant curriculum management systems then provide a comprehensive and structured curriculum description with links to guaranteed study materials (printed books and electronic multimedia tools, including virtual patients), interactive reporting of teaching contents and a particular module dedicated to supporting modern assessment methods like OSCE (Objective Structured Clinical Examination).

MedBiquitous develops a set of proven standards to advance lifelong learning, continuous improvement, and better patient outcomes. It covers a wide range of XML (eXtensible Markup Language) and web service standards, including (i) the Competency Framework for integrating competency frameworks into educational technologies such as curriculum management systems, (ii) the Curriculum Inventory for aggregation of curriculum data for research and benchmarking purposes, and (iii) the Medical Education Metrics for collecting validation data for standardised survey items. All these standards go through a rigorous ANSI-accredited development process. From an international, systematic, practical, and long-term perspective, the MedBiquitous outputs must

[1] https://www.medbiq.org/

[2] https://www.medbiq.org/standards

be considered for implementation in solutions like curriculum management systems.

AIMS

- To adopt and implement a proven methodological background addressing the current challenge faced by the health professions to evaluate and map medical and healthcare curricula.
- To systematically build an origin model using proper data mining techniques to compare standardised medical curricula.
- To find a straightforward way to present data analytics interactive reports covering keyword-based exploration and similarity comparison on available data.

METHODS

During the MEDCIN project, a best-practice methodology for achieving comparability of higher education quality across Europe was proposed, including an exemplar case study illustrating how these standards could be effectively integrated into the practice. This pilot implementation demonstrated how the MedBiquitous standards could effectively integrate into the course. The guaranteed and high-quality curriculum in medical and healthcare education is still essential because medicine does not allow gaps in the knowledge obtained during studies, and any error in medical practice may lead to fatal consequences. From the perspective of human cognition abilities, it is impossible to carefully read and verify the whole curriculum (the content of all learning units with all their linkages and co-dependencies to learning outcomes and recommended study materials). The use of proper data mining techniques and analytical methods can in-depth explore all sections of a curriculum.

This is why MEDCIN proposed an innovative methodological background, including a web-based visualisation tool for comprehensive evaluation and medical curricula mapping with modern information and communication technologies. The aim was to identify and validate novel, potentially valuable patterns which significantly help curriculum managers/evaluators make the right decisions and build a well-balanced medical curriculum. MEDCIN didn't only eliminate poor transparency in curricula and helped improve teaching. Many medical and healthcare institutions have already described their curricula. At the time of project proposal, comparing curriculum content was a challenging area where no proven approach had been published before. MEDCIN introduced a new computational system, including data mining, which compared curricula based on a standardised format. It allowed users to draw a comparison between two independent profiles of graduates. MedBiquitous standards ensured broader applicability beyond the partnership for MEDCIN's approach and model, disseminating more widely as an exemplar of best practices across Europe. The strength of the collaboration and the experience obtained allowed MedBiquitous Europe to provide a more robust and coherent view of the European needs in developing data standards. The influence of this partnership on the discussion ensured that any future standards would have broader applicability and widespread use, allowing them to further progress towards widespread European directives such as the Bologna process.

Specifically, MEDCIN covered the underpinning methodology enhancing the process of medical and healthcare curriculum standardisation with the following outputs:

1. Preparation for the project through data and related information on the use of existing Medbiquitous standards implemented in curriculum management systems.

2. Overview of usable healthcare informatics standards in higher institution education.

3. Implementation of approved standards into a selected existing platform for curriculum management.

4. Assessment study of proposed guidelines.

5. Preparation of best-practice methodology ("how to standardise medical and healthcare curriculum by series of agreements the at European higher education area").

6. Perform a practical-oriented workshop, where all the necessary topics of medical curriculum standardisation are introduced.

7. Dissemination of the complex guidelines, including best-practice methodology, to the broader community, particularly across the MEFANET network[3] and the MedBiquitous consortium.

[3] https://www.mefanet.cz/index-en.php

TECHNOLOGIES

Concerning technological background, the MEDCIN platform ran on commonly used and widespread web servers, including an Apache or Microsoft Internet Information Server. An Apache server running on Ubuntu Long Term Support (LTS) and a Linux distribution of an operating system (concerning its stability and proven performance) were used. The server-oriented scripting language PHP was used to build the application layer of the platform. It represented a hybrid semantic programming language, according to Malkov [2]. The whole web application was made on a Symfony framework because it generally supports object-oriented design and event-driven development and provides a set of tools for debugging. This framework was released under an open-source license and could be extended with many plug-ins and add-ons. PostgreSQL technology – an open-source Object-Relational Database Management System – was used for the data layer. This model was based on objects, classes, and inheritance, directly supported by database schemas and the query language. It also supported extending the data model with custom data types and methods like relational systems. For statistical computations, R language was used. The R scripts lived in the dedicated R server, and a Symfony application could call R methods using the OpenCPU system. For rendering the graphs in the web environment, the JavaScript library d3.js was used. An example of deployment of the production instance of the MEDCIN platform is shown in Figure 1 [3].

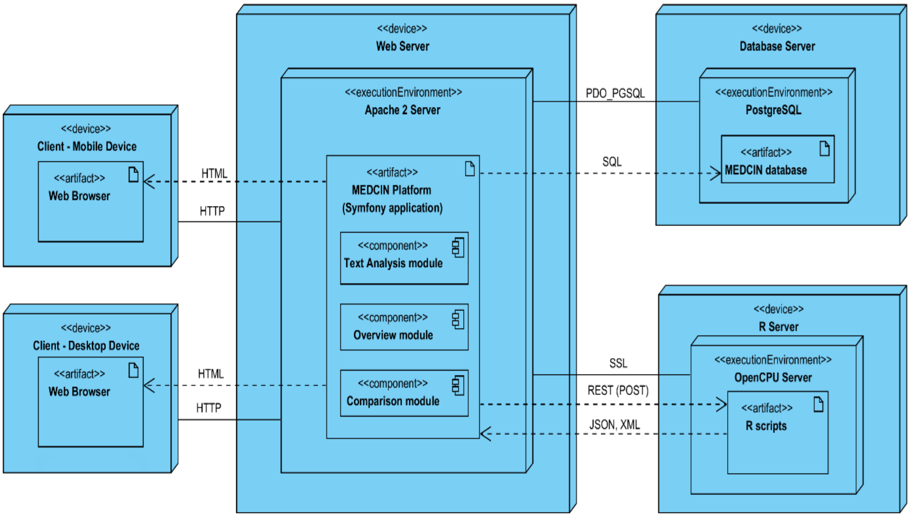


Figure 1: The MEDCIN platform deployment diagram

RESULTS

BUSINESS UNDERSTANDING

Generally, a curriculum is a statement of an educational programme's intended aims, objectives, content, experiences, outcomes, and processes. A curriculum includes at least the following information: (i) a detailed description of the training structure (entry requirements, length, and programme organisation, including its assessment system), (ii) a description of expected learning methods, teaching, assessments, feedback, supervision, and evaluation, (iii) it typically covers both generic professional and speciality-specific areas. How a curriculum for medical and healthcare education is constructed depends on the designers' views about learning theories, and experiences, how medicine is practised, social responsibility and accountability issues, the role of the knowledge base, professional values, and health service development. A syllabus is simply a list of the main topics, the content to be covered by a course of study, and is only part of the curriculum. The syllabus content of the curriculum should be stated in terms of what knowledge, skills, attitudes, and expertise the learner will achieve. In its final format, a curriculum is an ideological, social, and aspirational document that must reflect the local institute's circumstances and needs.

The curriculum design focused on two main components: (i) the structure of the curriculum, (ii) the content of the curriculum. The design process is a complex task requiring much experience and effort. There is a lot of conversation and research based on the curriculum designers' values, vision, assumptions, and institutions or social, economic, political, and cultural influences. Below is a summary of some of the essential characteristics that a curriculum should follow:

- A well-developed curriculum must tell the learner precisely what to expect, including the methods of student support.
- It has to advise the teacher on delivering the content and supporting the learners in their personal and professional development tasks.
- It should help the institution to set appropriate assessments of student learning and implement relevant evaluations of the educational process.
- It must inform society how the school is executing its social responsibilities.
- The curriculum should present a reasoned picture of the subject to be studied and define the teaching and learning processes and the intended learning outcomes of that study.

- A prior statement of vision, mission or values must exist, and all decisions about the curriculum must be taken under this statement. The information must be made for the local context. Contextual opinions are expressed in concrete terms. General comments are of limited value and must be avoided.

Medical and healthcare education standards are heavily used in curriculum design and implementation. Since this is an essential part of medical and healthcare education, definitions and basic principles related to the curriculum design using standards are crucial to understanding in detail. Beyond medical and healthcare vocabularies and terminologies (for example, UMLS – Unified Medical Language System, MeSH – Medical Subject Headings, ICD – International Statistical Classification of Diseases and Related Health Problems, or Systematized Nomenclature of Medicine Clinical Terms), specific standards are created for medical education. According to the MEDCIN project's set outcomes, the selection and adoption of appropriate standards and their implementation into the selected curriculum management system, along with the development of tools that extracts and represents the standardised curriculum data overview, play a significant role. A detailed state-of-the-art review [4] of available medical and healthcare education technical standards and an overview of available standard-compliant systems identified those most relevant to the field. Based upon this initial pool of bars, a selection criterion was applied to explore these required standards and best suited to developing a standardised model for describing medical and healthcare curricula. MedBiquitous standards [5] (high level of adoption in the United States), namely the Curriculum inventory (it standardises instructional, assessment methods, and resource types), the Competency framework (it uses competencies and learning outcomes to structure education and performance management systems, allowing for searchable curriculum) and the Competency object (together with the Competency Framework, it allows for representation of unique competency objects, including learning outcomes, objectives, and goals), have been identified as the most relevant and proven standardised concept for further implementation. Table 1 summarises the criteria (C1–C5) considered during the state-of-the-art review.

- C1: The standard can be integrated into a Curriculum Management System.
- C2: The standard can standardise an entire medical curriculum.
- C3: The standard can communicate the entire curriculum for comparison and benchmarking purposes.
- C4: The standard can be used to standardise only a specific part of a curriculum (competencies, objectives, outcomes, learning activities and purposes) and is associated and works with other standards that satisfy the criteria C1, C2 and C3.

- C5: The standard reports various educational and administrative procedures and processes other than those related to project goals or is not associated with different measures that satisfy the criteria C1, C2 and C3.

From the listed criteria, C1 to C4 were crucial for achieving the MEDCIN project goals, and therefore, standards that satisfied one to all of them were included. In contrast, standards that met the C5 criterion were excluded as non-appropriate for the project's goals.

Table 1: Criteria summarisation of MedBiquitos standards

Standard	C1	C2	C3	C4	C5	Decision
Activity Report (AR)	-	-	-	-	√	Exclude
Curriculum Inventory (CI)	**√**	**√**	**√**	**-**	**-**	**Include**
Competency Framework (CF)	**√**	**√**	**-**	**√**	**-**	**Include**
Healthcare LOM (HLOM)	-	-	-	√	√	Exclude
Healthcare Professional Profile (HPP)	-	-	-	√	√	Exclude
Medical Education Metrics (MEM)	-	-	-	-	√	Exclude
MedBiquitous Virtual Patient (MVP)	-	-	-	-	√	Exclude
Performance Framework (PF)	-	-	-	-	√	Exclude
Competency Object (CO)	**-**	**√**	**-**	**√**	**-**	**Include**

DATA UNDERSTANDING

In general, the innovations, if performed in teaching domains formalised with the use of a detailed parametric description and entities adopted from the outcome-based concept (learning units, learning outcomes), enhance the transparency and continuity of the environment in which the authors of teaching materials, as well as students, work daily [6]. For example, the database schema describing the curriculum of the Faculty of Medicine at Masaryk University consists of several entities. The most crucial objects from the MEDCIN project perspective are the following:

- **Programme** – information about the whole curriculum binding to the field
  - MEDCIN standard-compliant curriculum building block: Curriculum
- **Medical discipline** – categories for medical and healthcare specialities
  - General Medicine study programme has 44 medical disciplines
  - MEDCIN standard-compliant curriculum building block: Sequence block

- **Course** – information about particular courses binding to the semester and medical discipline
  - General Medicine study programme has 138 active courses
  - MEDCIN standard-compliant curriculum building block: Sequence block
- **Unit** – in other words, learning units are parts of courses (the course comprises learning units).
  - General Medicine study programme has 1,400+ learning units
  - MEDCIN standard-compliant curriculum building block: Event
- **Learning outcome** – competencies of each student after finishing learning units, courses or the whole programme
  - General Medicine study programme has 6,900+ outcomes
  - MEDCIN standard-compliant curriculum building block: Competency object

In our case, the curriculum is described by its courses, learning units, and learning outcomes and their attributes (titles, descriptions, keywords, annotations, and study materials). One example of the Connective tissue learning unit is shown below (Table 2).

Table 2: Example of learning unit description

Learning unit	Connective tissue
Study programme	General Medicine
Medical section	Theoretic sciences
Discipline	Histology and embryology
Courses	VSHE0221c - Histology and Embryology I - practice VSHE0221p - Histology and Embryology I - lecture VSHE0322c - Histology and embryology II - practice VSHE0322p - Histology and embryology II - lecture
Learning outcome 1	Student characterises standard features of the connective tissue, extracellular matrix composition, fibre types, their characteristics, occurrence and arrangement of the connective tissue, origin, and regeneration ability of the connective tissue.
Learning outcome 2	Student characterises the microscopic structure of the adipose tissue, adipose tissue types and their function, embryonic origin, and adipose tissue regeneration.
Learning outcome 3	Student characterises the microscopic structure of the bone tissue, bone cell types and intercellular substance composition, types of the bone tissue, process of formation (endochondral and intramembranous ossification), bone remodelling and regeneration, and types and structure of bone joints.

DATA PREPARATION

As was mentioned before, the MedBiquitous standards are described in XML format. A database entity relationship diagram was designed for a proper and precise curriculum data arrangement (see Figure 2). Therefore, the XML standardised concept has been successfully implemented on the data layer.

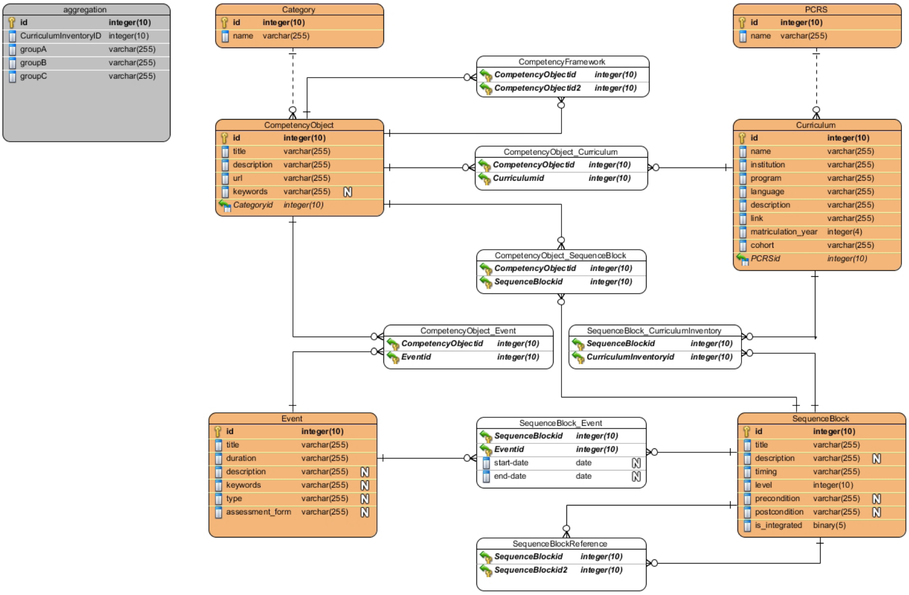


Figure 2: MEDCIN Platform database scheme

This approach enabled to simulate using two selected standards in practice, namely the Curriculum Inventory and the Competency Framework, by finding the attribute matching between the XML scheme of these standards, newly designed entities, and existing entities representing various pieces of curricula. A detailed description of crucial entities is provided below.

CURRICULUM

— Information about the curriculum concerning the Physician Competency Reference Set (PCRS)[4].

— This entity had multiple links to sequence blocks and competency objects (programme-level competencies).

[4] https://www.aamc.org/initiatives/cir/about/348808/aboutpcrs.html

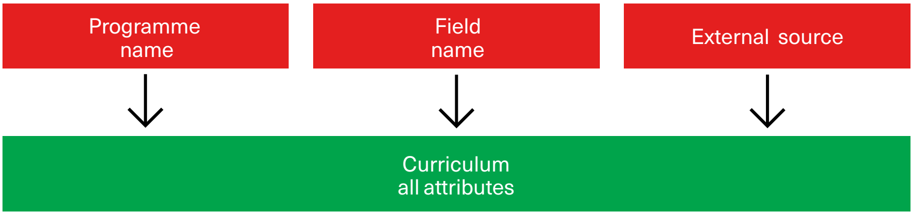


Figure 3: Standardisation towards “curriculum” entity

SEQUENCE BLOCK

— Information about sequence blocks connected to the curriculum.

— This entity could reference other sequence blocks, events, curricula, and competency objects (sequence block-level competencies).

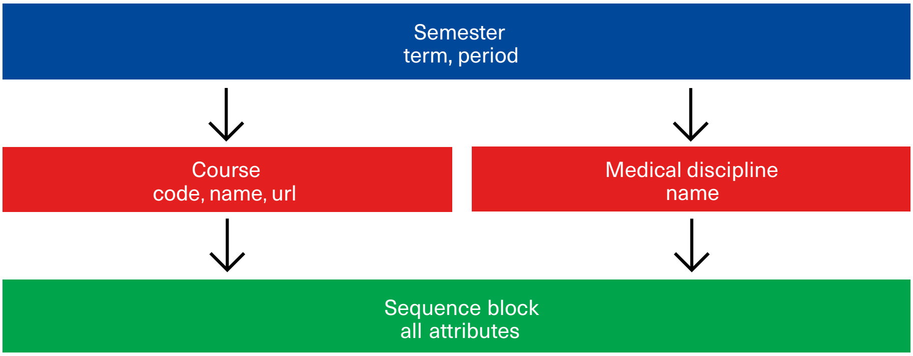


Figure 4: Standardisation towards “sequence block” entity

EVENT

— An entity represented events inside the sequence blocks.

— This entity could be directly linked to multiple sequence blocks and competency objects (event-level competencies).

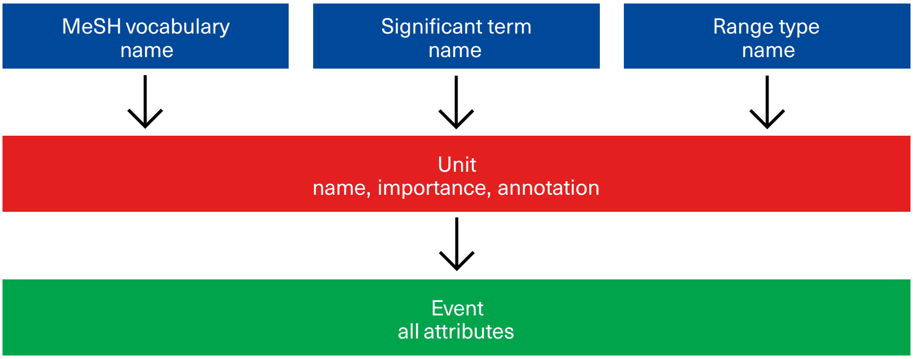


Figure 5: Standardisation towards "event" entity

COMPETENCY OBJECT

- Competencies were divided into three categories derived by their binding to other entities – program, sequence block, or event level.
- By binding to itself, we simulated the Competency Framework standard.

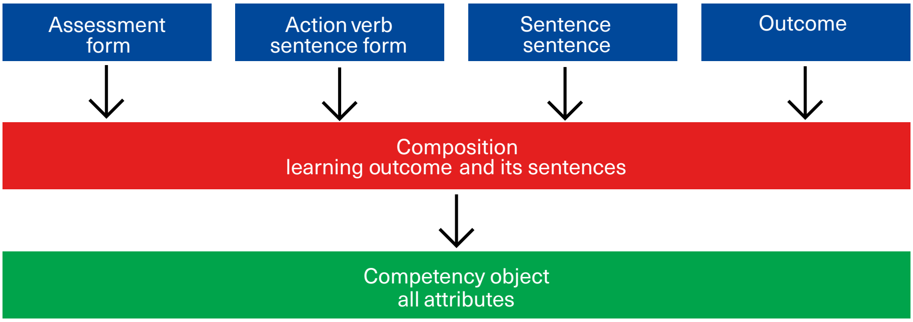


Figure 6: Standardisation towards "competency object" entity

MODELLING

Based on the outcome-based paradigm [7] and Medbiquitous standard-compliant database structure, MEDCIN's web-based platform for prototype curriculum analysis and mapping was developed. The primary goal had been to provide a more transparent overview of curriculum building blocks, including a new perspective on showing and explaining a complicated structure of mentioned standard in practice. For this purpose, data mining and statistical methods were applied in compliance with standardised and approved approaches. Interactive reporting on available curriculum data assisted in identifying potentially problematic areas and constructing a comprehensive overview of the imported curriculum because, from the perspective of human cognition abilities, it was not possible to carefully read, verify and understand all building blocks with all their linkages and co-dependencies. Moreover, this activity was expected to bring a new communication channel between stakeholders such as curriculum designers, guarantors, department heads and faculty management.

MEDCIN addressed the need for an innovative concept which helped to identify automatically the most frequent topics in the form of the most occurring keywords taught over the study of medicine and healthcare. The module for curriculum text analysis and comparison on keyword occurrence aimed to compare various curricula to local institutional needs easily. Keyword extraction was an essential technique for document retrieval. The curriculum's most frequent and potentially relevant topics could be readily displayed by extracting appropriate keywords. The comparison module facilitated decision-makers and audit bodies in evaluating a medical programme as they could see how the curriculum met the requirements set by a higher education board and quickly discover possible gaps and overlaps. MEDCIN covered three independent medical curricula, namely those from Masaryk University (General Medicine study programme), Karolinska Institutet (Clinical Medicine – Surgery), and Aristotle University of Thessaloniki (School of Medicine).

For example, the curricula comparison module (see Figure 7) could quickly compare two curricula or courses within one curriculum. It provided three main functionalities – summary report, search by phrase, and course comparison. The whole module could be used for every imported curriculum to the MEDCIN platform.

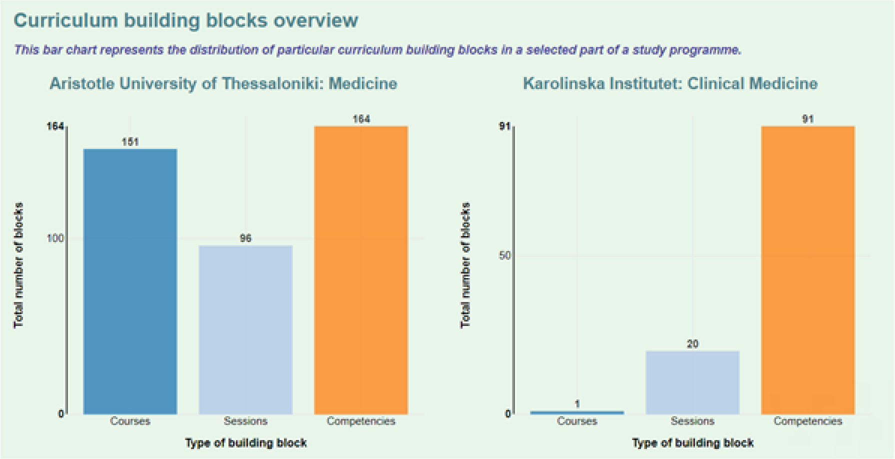


Figure 7: Basic statistics about compared curricula

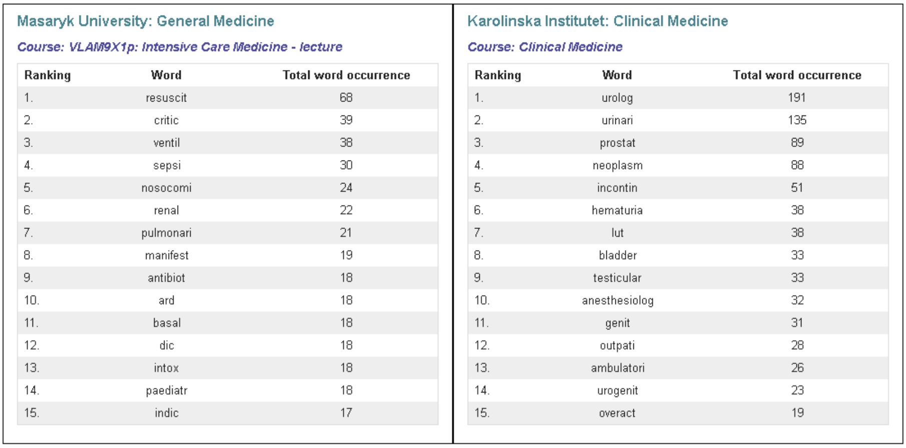
Course comparison was carried out on the course level and was performed using frequency or relevance analysis. Frequency analysis provided a text-analytical report based on the keyword occurrence. Relevance analysis delivered text-analytical information found on the weighted keyword occurrence. Afterwards, the text-analytical report consisted of several parts for each curriculum:

— word cloud showing the most frequent keywords occurring in selected courses' sequence blocks,

— frequency table (see Figure 8) and histogram showing the 15 most frequent words occurring in selected courses' sequence blocks,

— dendrogram (see Figure 9) showing the similarity between selected course events.

All computing and visualisations in MEDCIN curricula comparison were performed by R using the OpenCPU server.



Masaryk University: General Medicine

Course: VLAM9X1p: Intensive Care Medicine - lecture

Ranking	Word	Total word occurrence
1.	resuscit	68
2.	critic	39
3.	ventil	38
4.	sepsi	30
5.	nosocomi	24
6.	renal	22
7.	pulmonari	21
8.	manifest	19
9.	antibiot	18
10.	ard	18
11.	basal	18
12.	dic	18
13.	intox	18
14.	paediatr	18
15.	indic	17

Karolinska Institutet: Clinical Medicine

Course: Clinical Medicine

Ranking	Word	Total word occurrence
1.	urolog	191
2.	urinari	135
3.	prostat	89
4.	neoplasm	88
5.	incontin	51
6.	hematuria	38
7.	lut	38
8.	bladder	33
9.	testicular	33
10.	anesthesiolog	32
11.	genit	31
12.	outpati	28
13.	ambulatori	26
14.	urogenit	23
15.	overact	19

Figure 8: Frequency tables of two courses comparison

The dendrogram below represents similarity where distinct colours in legend help to identify clusters of similar events in both selected courses.

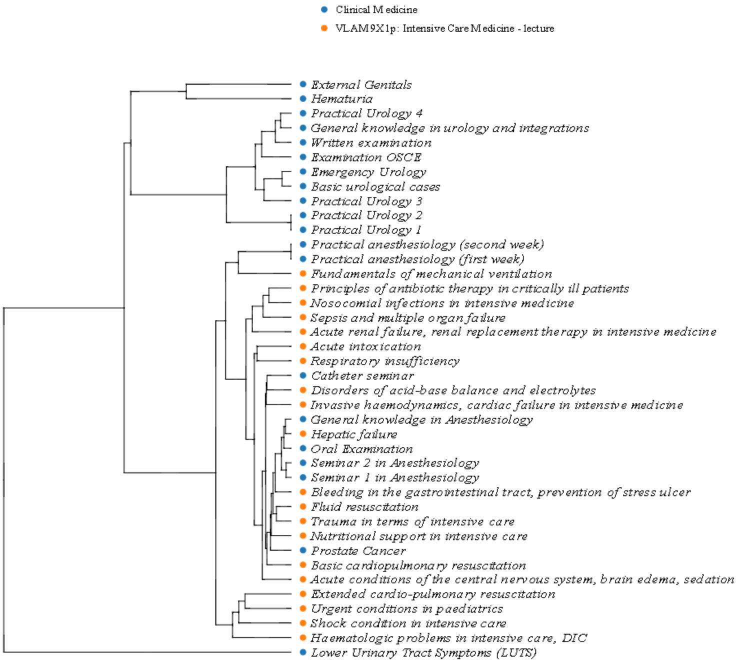


Figure 9: Dendrogram of two courses comparison

EVALUATION

Software testing and user feedback on different levels are critical components of software product development. It improves consistency and performance and makes the software more accurate and reliable. This phase covers in-depth testing and evaluation from various target group perspectives of the developed platform during the pilot phase. The MEDCIN project team performed a continuous evaluation procedure in a 3-axis system: (i) heuristic evaluation, (ii) user testing, (iii) face-to-face workshop presentations to a vast community of educators and students across MEFANET, MedBiquitous and all involved partners.

HEURISTIC USER INTERFACE DESIGN EVALUATION

During the module testing and the pilot evaluation period, we performed a thorough usability design evaluation based on Jakob Nielsen's [8] Ten Usability Heuristics for User Interface Design, where general principles for interaction design were applied; this is a standardised method to measure the usability of a software system from a user's perspective. They are called "heuristics" because they are broad rules of thumb, not specific usability guidelines. A group of software usability experts examined the MEDCIN platform through the ten aspects. The heuristic evaluation analysis for the MEDCIN platform was as follows:

1. Visibility of system status
   - The system succeeds in always keeping users informed about what is going on. This happens within a reasonable time. Every user action gives appropriate feedback to the user.
   - Overall rating: 5/5

2. Match between the system and the natural world
   - The system, in most cases, speaks the users' language, with words, phrases and concepts familiar to the user rather than system-oriented terms. Real-world conventions make information appear in a natural and logical order. Tooltip texts, though, where exists, could be more explainable.
   - Overall rating: 4/5

3. User control and freedom
   - The MEDCIN platform usually allows users who choose system functions by mistake and need a marked "emergency exit" to leave the unwanted state without going through an extended dialogue. The system supports undo and redo actions but lacks the same functionality through the browser's buttons.
   - Overall rating: 4/5

4. Consistency and standards
   - There are no different words, situations, or actions that mean the same thing.
   - Overall rating: 5/5

5. Error prevention
   - Whenever it is needed, there are good error messages. Moreover, the careful design prevents a problem from occurring in the first place. The system eliminates error-prone conditions and checks for them. When needed, the system presents a confirmation option before users commit to the action.
   - Overall rating: 5/5

6. Recognition rather than recall
    - Overall, the system minimises the user’s memory load by making visible objects, actions, and options users must remember some information from one part of the dialogue to another. Proper instructions for using the system are visible and easily retrievable whenever appropriate.
    - Overall rating: 5/5

7. Flexibility and efficiency of use
    - Accelerators — unseen by the novice user — may often speed up the interaction for the expert user so that the system can cater to inexperienced and experienced users. The system does not allow much freedom to the users to tailor frequent actions.
    - Overall rating: 3/5

8. Aesthetic and minimalist design
    - Dialogues do not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant information units and diminishes their relative visibility.
    - Overall rating: 5/5

9. Help users recognise, diagnose, and recover from errors
    - Error messages are expressed in plain language (no codes) and precisely indicate the problem (in most cases). Constructively they do suggest a solution.
    - Overall rating: 5/5

8. Help and documentation
    - Although the system can be used without documentation, it is necessary to provide help and documentation. The MEDCIN platform does not contain any distinct help area, with information accessible to search, focused on the user’s task, or lists concrete steps to be carried out.
    - Overall rating: 2/5

PILOT EVALUATION

During the MEDCIN project, three workshops were organised, including focus groups, presenting the platform in different stages of development for more than 100 users (students, academic staff, curriculum designers, and technicians). Precious feedback by giving online questionnaires was collected from end users and professionals thoughtfully with the possibility to implement the desired functions and functionality. One critical aspect that the evaluations showed was the use of the appropriate terminology. End users and professionals noted the need for more used language between the system and the real-life context. Some example responses received during this session are shown below:

— "If you can work more on the student interface, it could be helpful for them also. If they can see the whole curriculum and connections of different organs/systems of the body, they may learn better."

— "Visualizations are needed to support many different stakeholders. For example, a teacher may want to know if the ILOs, learning activities and examination are aligned in the course and how the learning outcomes and content relate to other classes in the curriculum. Just an example. Your platform provides a good start for being able to visualise this type of info in future."

— "Visualizations would be much desired to provide a better insight."

DEPLOYMENT

After thorough testing and feedback, the identified bugs and relevant new requirements were incorporated into the platform. The final version went through various stages of development. It was gradually made available to the individual teams on development, staging, and production servers that were part of the Masaryk University infrastructure. Access to the section of the MEDCIN platform that offered the curricula comparison module was only available in the production version after logging in. It provides three main functionalities – summary report, search by phrase, and course comparison. The whole module can be used for every imported curriculum to the MEDCIN platform in the proper standard-compliant format, or manual operations on input data are needed.

DISCUSSION

During the MEDCIN project, a best-practice methodology for achieving comparability of higher education quality across Europe was proposed, including an exemplar case study illustrating how these standards can be effectively integrated into the practice. This pilot implementation demonstrated how the

MedBiquitous health professions education and credentialing standards could effectively integrate into the courses. MEDCIN brought an innovative methodological background, including a web-based visualisation tool for a comprehensive evaluation and medical curricula map using proper data mining techniques and analytical methods. Users (faculty management, curriculum designers, and the academic community) can quickly identify and validate novel, potentially valuable patterns to help make the right decisions and build a well-balanced medical curriculum. Many medical and healthcare institutions have already described their curricula, but comparing curriculum content is still more challenging.

The MEDCIN project introduced a new computational system to compare curricula based on a standardised format, including data mining and natural language processing techniques. It allowed for drawing a comparison between two independent profiles of graduates. Using MedBiquitous standards ensured broader applicability beyond the partnership for MEDCIN's approach and model, disseminating more widely as an exemplar of best practices across Europe. The strength of the collaboration and the experience obtained allows MedBiquitous Europe to provide a more robust and coherent view of the European needs in developing data standards. The influence of this partnership on the discussion ensured that any future standards would have broader applicability and widespread use, allowing them to further progress towards widespread European directives such as the Bologna process. With the MEDCIN platform, users can parametrically describe and overview standard-compliant medical curricula and get comprehensive information for further expert verification. It also proved the efficiency of every individual method in practice on accurate institutional data.

The MEDCIN platform was successfully developed and operated by the project objectives. Basic information is still available on the official project website[5]. For technical and maintenance reasons, the MEDCIN web application, including all functionalities, was moved to the local network of the Faculty of Medicine of Masaryk University in 2022, where it can still be used for the medical and healthcare curriculum data comparison and mapping in consultation with the development team. Selected functionalities are continuously incorporated into the SIMUportfolio integration platform, which represents an innovative integrated system that makes it easier for students and teachers to learn and, as a result, improves students' knowledge and skills for practice.

[5] https://medcin.iba.muni.cz/index.php

EVALUATION OF THE AIMS OF THE CHAPTER

- To adopt and implement a proven methodological background addressing the current challenge faced by the health professions to evaluate and map medical and healthcare curricula.
  - Based on needs from Masaryk University (Czech Republic), Aristotle University of Thessaloniki (Greece), Karolinska Institutet (Sweden), and St George's University of London (United Kingdom), the MEDCIN project team prepared an in-depth review describing various MedBiquitous and AAMC (Association of American Medical Colleges) specifications for Health Professions Curriculum standards and finally decided to adopt three of them (Curriculum inventory, Competency framework, and Competency object) based on the given criteria.
- To systematically build an origin model using proper data mining techniques to compare standardised medical curricula.
  - The entity-relationship model successfully described the standards mentioned, which cover all crucial entities, relations, and attributes. The MEDCIN platform provided semi-automatic curriculum import and matching features to process, analyse, and visualise exciting patterns and clusters based on keyword similarity.
- To find a straightforward way to present data analytics interactive reports covering keyword-based exploration and similarity comparison on available data.
  - The MEDCIN curricula comparison module provided several approaches to explore and compare available parts of curricula in the form of interactive (filterable and sortable) tabular overviews, bar charts, pie charts, word clouds, and dendrograms. Fully standard-compliant features like summary overview, search by keyword, or similarity analysis helped explore curricula from qualitative and quantitative perspectives.

LESSONS LEARNED

This demonstration of international cooperation was based on the experience of the Faculty of Medicine of Masaryk University in the implementation of several projects in the field of curriculum description and mapping. The key element was the standardisation of curriculum building blocks that would then be applicable to other academic institutions where medicine is taught. The most challenging phase was understanding the domain, even though several partners had many years of experience in this area. Conducting a comprehensive search, identifying appropriate standards and implementing them enabled data collection from several European partners. Regular communication with the involved teams from different countries and overall project management was essential to the project's success.

Without a practical implementation of the theoretically described technical standards for medical education, it would not have been possible to collect and process data consistently. The outputs of the project have shown how similarities in selected parts of subjects taught in different faculties can be determined.

REFERENCES

[1] Nokkala T. The Bologna process and the role of higher education: Discursive construction of the European higher education area. In: Enders J, Jongbloed B (eds). Public-Private Dynamics in Higher Education. Expectations, Developments and Outcomes. Bielefeld: transcript Verlag; 2007. p. 221–245.

[2] Malkov SN. Customizing a functional programming language for web development. Comput Lang Syst Struct. 2010;36(4):345–51.

[3] Komenda M, Karolyi M, Vyškovský R, Ježková K. Towards a keyword extraction in medical and healthcare education. In: Ganzha M, Maciaszek L, Paprzycki M (eds). Federated Conference on Computer Science and Information Systems (FedCSIS). New York: IEEE; 2017. p. 173–176.

[4] Vaitsis CH, Spachos D, Karolyi M, Woodham L. Standardization in medical education: review, collection and selection of standards to address. MEFANET J. 2017;5(1):28–39.

[5] Balasubramaniam CH, Smoothers V. MedBiquitous Europe and its Role in Improving Technology Standards for Medical and Healthcare Education. Bio Algorithms Med Syst. 2010;6(11):7–8.

[6] Komenda M, Víta M, Vaitsis C, et al. Curriculum mapping with academic analytics in medical and healthcare education. PloS One. 2015;10(12):e0143748.

[7] Harden R M. Outcome-based education: the future is today. Med Teach. 2007;29(7):625–9.

[8] Nielsen J. Heuristic evaluation. In: Nielsen J, Mack RL (eds). Usability Inspection Methods. New York: John Wiley & Sons; 1994. ISBN 0-471-01877-5.

SECTION B

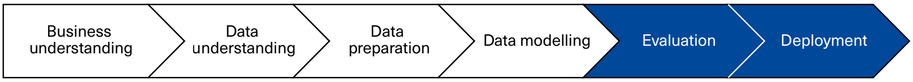
06

PROCESSING, ANALYSIS, AND VISUALISATION OF OBJECTIVE STRUCTURED CLINICAL EXAMINATION DATA

Martin Komenda, Daniel Barvík, Tereza Vafková, Petra Růžičková, Tereza Prokopová, Václav Vafek, Vojtěch Bulhart, Martina Kosinová, Petr Štourač

The case study was supported by the following grant projects: Masaryk University Strategic Investments in Education SIMU+ (CZ.02.2.67/0.0/0.0/16_016/0002416) from European Regional Development Fund and Masaryk University 4.0 (CZ.02.2.67/0.0/0.0/16_015/0002418) from European Social Fund.

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2016–now
Keywords	OSCE, first aid, interactive visualisation, SIMUportfolio
Research question	What is the role of a complex online analytical reporting system in a life cycle of objective structured clinical examinations in the medical and healthcare domain?
Type of result	Complex web application
Level of data processing	Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

The Simulation Centre (SIMU) of the Faculty of Medicine of Masaryk University is one of the largest (covering more than 8,000 square metres in total) and the most modern simulation centres in Central Europe. SIMU represents a unique teaching complex which combines theoretical and practical education and covers a comprehensive spectrum of simulation teaching methods. Great emphasis is put on developing students' soft skills, such as communication skills, decision-making skills, critical thinking, crisis communication and teamwork. SIMU focuses on the innovation of various medical programmes at Masaryk University by integrating advanced elements of simulation medicine into regular teaching. The SIMU team of experts provides the methodological, pedagogical and technical background for practical improvements in the General Medicine and Dentistry study programmes. The SIMU project addresses the following domains: (i) interactive clinical training, (ii) development of curriculum materials, (iii) standardisation of teaching and its continuous evaluation, (iv) implementation of objective structured clinical examination, (v) technology-enhanced learning [1].

The Objective Structured Clinical Examination (OSCE) is the gold standard for clinical skills evaluation such as patient communication, history taking, physical examination, specific procedures, prescribing, X-ray evaluation, electrocardiogram reading and many others [2,3]. As the entire process of the OSCE is time-consuming, a robust and systematic solution was needed. The integration platform called SIMUportfolio serves the faculty as support for learning, teaching, and faculty management. SIMUportfolio contains several specific modules, one of which is dedicated to the comprehensive management support of the OSCE. It aims to verify the student's clinical knowledge and skills and to provide objective and specific feedback.

Both the preparation and implementation of OSCE require a considerable investment from the faculty, including human resources and technical facilities. The examination is always scheduled for a whole day, with an exact arrangement of OSCE stations and times of individual students entering those stations. The process of OSCE preparation is relatively complex for a guarantor, who is expected to design a given examination with all of its parameters. Therefore, a dedicated module of the SIMUportfolio platform is used to support and automate some parts of this process (preparation of the station prototype, uncomplicated and systematic instructions for observers, preselection of rooms, planning the sequence of students accessing the examination to avoid collisions, automated evaluation of tests, archiving the examination results, and reporting services on available data) [4]. Finally, complex interactive reporting with the results of individual exams has been designed and developed based on collected data, which are stored in the central SIMUportfolio database.

AIMS

- To understand and adopt a comprehensive OSCE process according to the needs and requirements of MED MUNI.
- To develop a specific module for the arrangement of data describing a complete OSCE management.
- To design and develop an interactive reporting, where all crucial characteristics will be presented.

METHODS

METHODOLOGICAL BACKGROUND

The OSCE has been described in many valid research articles and conference papers [5–7]. To be able to fully understand how OSCE works in practice and how to implement it at MED MUNI, the SIMU team has taken part in several activities: passive as well as active attendance at international conferences (e.g. AMEE[1] or SESAM[2]), a broad review of published outputs on the Web of Science and Google Scholar, short internships and shadowing with European university partners, sharing experience in the OSCE domain among MEFANET partners covering all Czech and Slovak medical and healthcare faculties, as well as internal educational workshops. All OSCE development and management processes are based on the principles of objectivity and standardisation. The candidates move through a series of time-limited stations in a circuit to assess their professional performance in a simulated environment. At each station, trained observers evaluate and mark candidates against standardised scoring system[8]. All students and involved stakeholders (representatives of faculty management, guarantors of medical disciplines, teachers, methodologists, computer scientists, analysts, technicians) were informed about the OSCE implementation, which covers: (i) objectivity – all students pass through the same stations and the same situations, with the same evaluation criteria, (ii) structuredness – each station involves a specific task or an unambiguously defined scenario, (iii) focus on clinical practice – students' theoretical knowledge and practical skills are

[1] https://amee.org/

[2] https://www.sesam-web.org/

evaluated according to standardised evaluation criteria [9]. Detailed training of OSCE observers was also an integral part of the whole process of deployment. Training the observers on the methodological side (how to correctly assess the checklist items) and the technical side (how to work with the SIMUportfolio platform) was necessary.

PEDAGOGICAL BACKGROUND

The First Aid course covers the essentials of basic life support, developing students' soft skills, such as communication, decision-making, critical thinking, crisis communication and teamwork. Simulation elements with the necessity of pre-tests, post-tests and evaluations are used to meet the stated objectives of the subject. The primary learning outcomes mentioned above are tested through the OSCE principle. This required course was the first piece of the curriculum where OSCE had been fully implemented. It consists of three essential components: (i) online lectures, (ii) e-learning self-study material containing structured theory, instructional videos, photos, schemes, and algorithms, and (iii) practices. There are two essential prerequisites: the students have to come to the practice prepared (i.e. they have to achieve a previously defined theoretical basis), and the number of instructors must be adequate to divide the study group into teams of 4–8 students. This form of peer-to-peer education enables students to experience everyday clinical situations from different views, obtain feedback, and get acquainted with the principles of an objective evaluation already during the learning process [9]. Final OSCE exams are always planned at the end of the course. Three periods of complete runs were available (autumn 2020, autumn 2021, and autumn 2022).

The main reasons why we use the OSCE methodology at MED MUNI in the First Aid course are as follows:

- students' first contact with this type of assessment method,
- the opportunity to assess skills that an oral exam or test could not evaluate,
- standardisation and objectivity of testing,
- comparability of results,
- the possibility of assessing the achievement of learning objectives and outcomes.

TECHNICAL BACKGROUND

A technically oriented team of computer scientists, developers, analysts, and senior OSCE guarantors has long provided full support for OSCE observers (teachers who observe the OSCE exams) in implementing this innovative approach to assessing clinical competencies. The basis for the long-term and systematic sustainability of the technical solution is comprehensive know-how across the following areas: design, development and implementation of web-based applications, information extraction from databases, data analysis, interactive data visualisation, and interdisciplinary team management. A close and regular collaboration between the faculty management, guarantors, lecturers, OSCE observers, teaching technicians, and the development team has proved crucial. Development and the analytical kits provide a complete set of components, tools, features, and applications needed for further data arrangement, processing, analysis, and visualisation: PostgreSQL technology for the data layer in the SIMUportfolio platform, R programming language and Microsoft Excel for advanced data processing, transformations and mapping, and Looker (previously known as Google Data Studio) for interactive data visualisations.

RESULTS

BUSINESS UNDERSTANDING

To adopt the OSCE domain in detail, the SIMU team had to thoroughly understand the entire life cycle of this innovative testing method. By the decision of the faculty management and the OSCE guarantor, the OSCE examination has been designed as a summative assessment [10]. Its main goal is to evaluate students' learning at the end of a defined period (e.g. comprehensive course) by comparing it against a specific standard or benchmark. Summative assessments often have high demands and a high point value. The critical steps in OSCE setup and implementation included the overall organisation (assembling a complete team), detailed research on the usage of OSCEs in medical and healthcare curricula, attending dedicated conferences and workshops, analysis of available resources (human sources, OSCE room capacities), blueprinting (to choose domains, skill, and competences), mapping to actual courses (learning units in close relations with defined learning outcomes), defining OSCE stations, and setting up OSCE exams (instructions for observers and students, creation of structured checklist, room and station equipment, internal review and documentation). For a successful and sustainable implementation of OSCE, three successive phases were implemented: (i) OSCE training stations, (ii) OSCE pilots during workshops, conferences, and selected learning units, (iii) OSCE deployment in selected courses. Individual comprehensive phases were identified and deployed as specialised but dedicated submodules (see Table 1) within the SIMUportfolio platform. Four main features covering these phases, which fully cover all needs and requirements of faculty management, guarantors, and teachers, have been designed, developed, and implemented.

Table 1: Overview of OSCE modules in the SIMUportfolio platform

OSCE module	Description
Sketch	Allows the user to define stations (a checklist with a specific situation that needs to be done) and exams (basic information, students for the exam, exam time, selected station(s) etc.).
Execute	Allows teachers to conduct exams and assess students on a given day. The teacher selects students on the platform and evaluates them through a predefined checklist.
Report	Includes the real-time students and exam results for the given day and a global summary overview of all exams.
Stats	Advanced reporting with statistics and overviews for the entire learning period (typically one semester) is presented here using the Google Looker Studio.

In addition, various user roles have been specified that directly correspond to the above-mentioned phases and are controlled by the access control list of SIMUportfolio (see Table 2).

Table 2: Overview of OSCE user roles in the SIMUportfolio platform

OSCE user role	Description
Designer	Responsible for preparing scenarios at an OSCE station and exams in the Sketch module.
Observer	Responsible for running and evaluating the examination itself. During the OSCE exam, he/she observes and assesses students' reactions with predefined checklists.
Guarantor	Manages the entire design, planning and evaluation process. This person can also monitor the whole process thanks to an adapted view in the platform, which involves up-to-date information on the progress of students going through individual stations, their results, and the entire examination schedule.
Student	He/she is essential for the examination aiming to prove his/her knowledge of medical procedures in specific situations.

In addition to the above technical settings, a blueprinting procedure defined a set of themes as individual OSCE stations linked to previously identified learning outcomes. Specifically, there were three stations focused on cardiopulmonary resuscitation. An integral part of this is the setup of the length for each stage of a particular OSCE exam for a single student (communication of instructions, the exam itself, setting up the station). Figure 1 schematically shows the implementation of one 60-minute OSCE exam where five students rotate through one of three content topics.

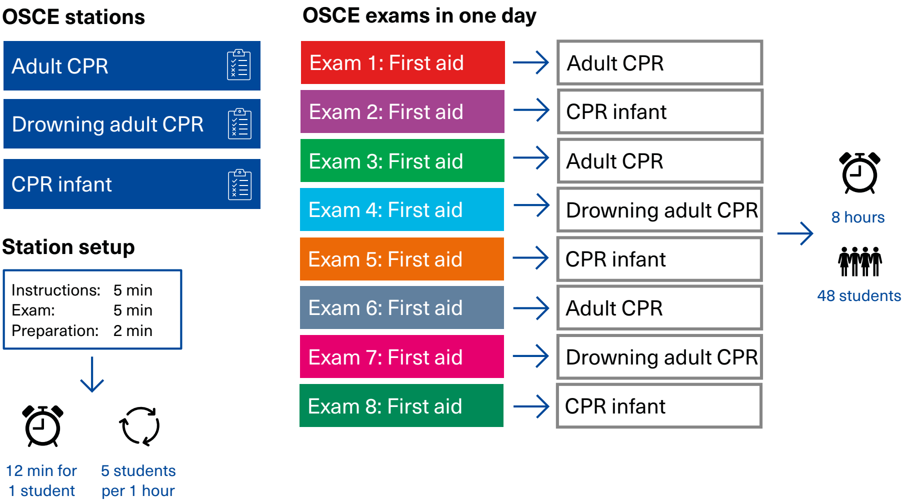


Figure 1: Schematic representation of OSCE stations and exams

DATA UNDERSTANDING

For the subsequent in-depth analysis of the OSCE data, it was necessary to appropriately design the database layout of all the essential characteristics encountered during the exam. In total, 16 standard and associative entities are linked to other objects in the SIMUportfolio platform. For analytical processing and accessible archiving of data, three export files were designed and implemented in practice, containing all information about each OSCE exam performed in a given period: (i) only the basic exam overview, (ii) complete summary information about OSCE (overview of OSCE observers, success rate of students, overview of stations etc.), (iii) a detailed analysis of all actions done by individual observers in SIMUportfolio during the given OSCE exam (see Table 3).

Table 3: Overview of OSCE exports in the SIMUportfolio platform

Export file	List of attributes
1 Basic export	student_uco,zkouska_vysledek,zkouska_id,zkouska_datum,zkouska_cas Student_id, exam_result, exam_id, exame_date, exam_time
2 Full export	zkouska_id, zkouska_nazev, zkouska_datum, student_uco, zkousejici_uco, zkousejici_prijmeni, zkousejici_jmeno, stanice_id, stanice_nazev, stanice_vysledek, stanice_pocet_bodu, stanice_max_pocet_bodu, komentar, ziskane_body_max_body, kategorie_hodnoceni, pracoviste, termin, vysledne_hodnoceni, prepsano, prepsano_komentar Exam_id, exam_name, exam_date, student_id, examiner_id, examiner_surname, examiner_firstname, station_id, station_name, station_result, station_number_of_points, station_max_number_of_points, comment, quotient_of_points, rating_category, workplace, attempt, final_result, change_result, change_result_comment
3 Observers log	zkouska_id, zkouska_nazev, zkouska_datum, stanice_id, stanice_name, student_uco, zkousejici_uco, zkousejici_prijmeni, zkousejici_jmeno, akce_cas_realny, akce_cas_stopky, akce_typ, akce_id_prvku, akce_skupina, akce_otazka, akce_hodnota_prvku, akce_je_zvoleno, zkouska_vysledek, stanice_vysledek Exam_id, exam_name, exam_date, station_id, station_name, student_id, examiner_id, examiner_surname, examiner_firstname, action_time_real, action_time_stopwatch, action_type, action_element_id, action_group, action_question, action_element_value, action_is_selected, exam_result, station_result

DATA PREPARATION

During the preparation phase, the primary focus was on the validation and overall accuracy of data exported from the primary SIMUportfolio database. The basis for data analysis was the OSCE report module, where the results of all exams and actions conducted by all observers are stored. First, data were explored, cleaned and validated manually and semi-automatically using contingency tables and one-dimensional graphs. The data cleaning process was then closely linked to the actual practical use of the data in analysis or subsequent visualisation. In particular, the file 3 Observers log, where all operations and user actions during the OSCE trial are located, required extensive editing to remove unnecessary or duplicate data. However, these are valid; they were extreme for the required output. Mapping such processed data to user roles and profiles was also integral in providing an accurate and personalised report for faculty management.

MODELLING

In designing and applying the algorithm with input-ready datasets (data preparation phase), the goal was to create comprehensive modules of the SIMUportfolio platform that would provide several critical views of the available data. The requirements list and specifications were made closely with all stakeholders (teachers, OSCE observers, OSCE guarantors, technical and development team, and faculty management). The result is two full-featured modules: the OSCE Report and the OSCE Stats.

The OSCE Report provides an individualised view of each OSCE examination conducted. It is easy to track the implementation date, the number of stations prepared, the number of students tested, and the exam's success rate. In addition, each station's progression is evident, including a preview of the item-structured checklist, optional text comments, and observer feedback. Finally, a detailed log of all interactions made by the observer during the trial is available in SIMUportfolio. In practice, this means that the start time, the marking or changing of each response (radio button, checkbox, Likert scale item), the textual comment and the final result, including the saved time, are stored (see an example of individualised log below).

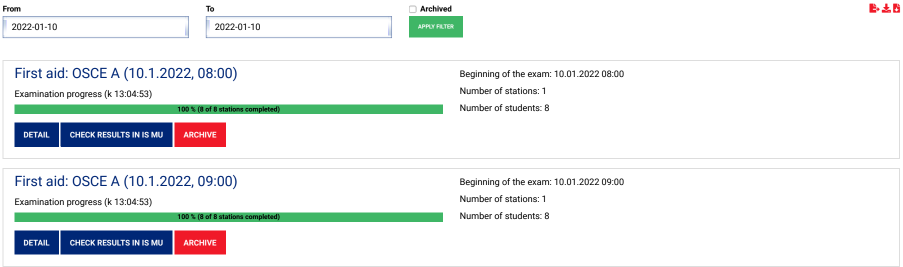
— COUNTER PLAY (time: 2022-01-10 08:26:43, stopwatch: 04:59)

— radios (time: 2022-01-10 08:26:44, stopwatch: 04:59): **itemId:** item0, **groupName:** Přístup k bezvědomému, **questionName:** Kontrola vlastního bezpečí (Safety), **value:** Ano, **checked:** 1

— checkbox (time: 2022-01-10 08:26:49, stopwatch: 04:54): **itemId:** item1, **groupName:** Přístup k bezvědomému, **questionName:** Oslovení (Stimulate), **value:** Oslovení (Stimulate), **checked:** 1

— checkbox (time: 2022-01-10 08:26:49, stopwatch: 04:54): **itemId:** item2, **groupName:** Přístup k bezvědomému, **questionName:** Zatřesení (Stimulate), **value:** Zatřesení (Stimulate), **checked:** 1

Using the date picker, the user can directly select any date or time interval in the application to display all exams performed. In addition, the three exports above are available over completely up-to-date data for further analysis, documentation, or archiving. Figure 2 shows a screenshot showing three OSCE exams from the First Aid course in the autumn of 2021 exam period.



From: 2022-01-10

To: 2022-01-10

□ Archived

First aid: OSCE A (10.1.2022, 08:00)

Examination progress (k 13:04:53)

100 % (8 of 8 stations completed)

Beginning of the exam: 10.01.2022 08:00
Number of stations: 1
Number of students: 8

First aid: OSCE A (10.1.2022, 09:00)

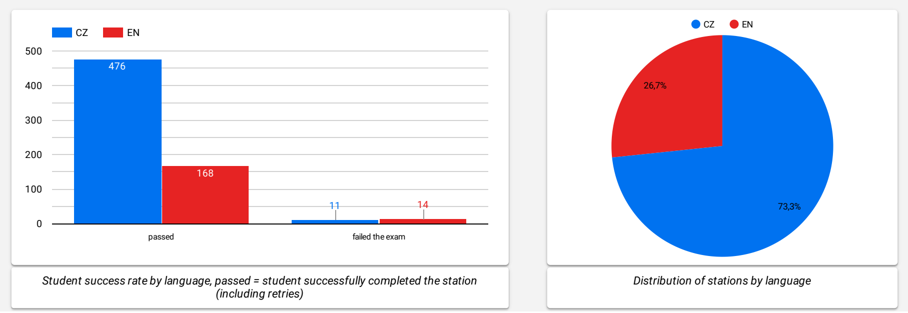
Examination progress (k 13:04:53)

100 % (8 of 8 stations completed)

Beginning of the exam: 10.01.2022 09:00
Number of stations: 1
Number of students: 8

Figure 2: Screenshot of the OSCE Report module presents an overview of available exams in the selected period

OSCE Stats delivers comprehensive, real-time interactive reporting on OSCE exam data over a given period, typically the spring or autumn semester. The complete visualisation (see Figure 3) is divided into full reports (basic overview, exam overview, observer overview, station overview, station ratings, and comments). Each view shows differently detailed outputs, ranging from an aggregated overview (aggregated numbers of students, exams, stations, observers, and overall pass rates) to detailed results for individual stations divided by language (Czech, English) and observer workplace (typically clinic or department at MED MUNI).



Student success rate by language, passed = student successfully completed the station (including retries)

Distribution of stations by language

Figure 3: Success rate and distribution of OSCE stations by language in the autumn of 2021 in the First Aid course

Beyond this set of reports, additional analytical outputs focused on specific OSCE issues are created; for example, a detailed analysis of the critical questions, which are mandated to meet the minimum criteria to pass the exam. Generally, due to the sensitivity of the information published, these reports are published in two modes:

— a Looker Studio visualisation embedded into the SIMUportfolio platform,

— a stand-alone Looker visualisation available only to selected MED MUNI users.

These reports are then used as objective material for optimising and improving individual OSCE stations for subsequent semesters, setting or refining threshold scores for passing the exam, administrative agenda related to observer rewards, and overall evaluation by faculty management.

EVALUATION

The evaluation phase was crucial because all analytical reports are primary decision-making support for course organisation and faculty management overview. Output evaluation was carried out at several levels: (i) validation of the three export files with the available individualised reports in the OSCE Report module and then comparison with the database dump containing relevant table structures and the data, (ii) validation of the interactive Looker reports with a set of crosstabs (contingency tables) and summary reports created for these purposes in MS Excel, (iii) internal manual testing and checking by testers and other members of the development team, (iv) final double-control by the First Aid course team. In close collaboration between these interdisciplinary teams, designing and implementing several summary visualisations was possible, including full exports over the available OSCE data.

DEPLOYMENT

All summary reports designed and published as an integral part of the OSCE evaluation are available directly in the SIMUportfolio platform in the OSCE Stats module for users with a given role according to a global access control list. Reports that contain detailed or more sensitive (but in line with General Data Protection Regulation), e.g. individual examiner scores, are deliberately prepared as stand-alone visualisations in the Looker application and are only available for the management of selected courses.

DISCUSSION

The first set of First Aid courses with OSCE in SIMU was launched in September 2020 with the vision to become the leading working place in medical and healthcare education in the Czech Republic by involving advanced trends and techniques. Very intensive and close cooperation with guarantors of the First Aid course provided interactive dashboards and data sets reflecting the current needs and requirements of the SIMU management. Within three years, most of the steps focused on data collection, processing and subsequent visualisation have become almost entirely automated. This means that every period (typically semester) it is necessary: (i) to manually process selected data sets in a minor way (data cleaning, transformations, and mapping), (ii) to set up the correct and valid input data for Looker reports, and (iii) to double-check format as well as content correctness of available data overviews (see Evaluation subchapter).

The SIMU management and the course guarantors then thoroughly analyse all information and modify the curriculum, specific stations or staffing if necessary.

EVALUATION OF THE AIMS OF THE CHAPTER

- To understand and adopt a comprehensive OSCE process to the needs and requirements of MED MUNI.
  - Thanks to the experience gained at the national and especially international level, the SIMU team has understood and fully adopted the comprehensive OSCE workflow. As a critical point, it should be noted that for a successful implementation, it is essential to appropriately combine the management requirements, the time, technical and spatial possibilities of the workplace, and the methodological background with recommendations for good practice.
- To develop a specific module for the arrangement of data describing a complete OSCE management.
  - OSCE can be implemented in paper or electronic form. If OSCEs are to be managed systematically, effectively, and in the long-term perspective, no other way than electronic is sustainable. SIMUportfolio fully covers all the complicated OSCE process features using the newly developed dedicated OSCE module. An undeniable advantage is the well-designed database model that contains all the essential characteristics describing individual OSCE. All data exports and static or interactive reports are always strictly based on the database built on top of this model.
- To design and develop an interactive reporting, where all crucial characteristics will be presented.
  - OSCE reports aim to provide basic and advanced overviews describing the progress of the OSCE assessment in a comprehensible way. The requirements specification for these reports, intended to support the optimisation of OSCE setup and delivery globally, was carried out in close collaboration with the team of the First Aid course. The subsequent choice of technology (using the business intelligence Looker tool) and the development of interactive visualisations were already purely technical issues. With a detailed knowledge of the OSCE methodological background, a well-arranged data layer, and a robust application to support the implementation of OSCE, it was possible to implement all the required reports and thus provide a tool as support for data-driven decision-making.

LESSONS LEARNED

The topic of objective structured clinical examination (OSCE) has become not only highly emphasised at the Faculty of Medicine of Masaryk University, but above all practically grasped. The enthusiasm and cooperation of the individual teams of the Medical Simulation Centre, together with the long-term support of the management, has resulted in the adoption of the theoretical principles into the local environment and the needs of the faculty, a brand new OSCE management module in the SIMUportfolio platform, thousands of tested students and available data on their evaluation and results. The complexity of the final domains of the CRISP-DM model, i.e. evaluation and deployment, became fully apparent, which in practice meant thorough scrutiny and optimisation of the resulting reports and the reflection of these outputs into practice in the form of modification of OSCE stations and exams.

The role of the online analytical reporting system became indispensable in evaluating the OSCE examinations in the past semester and obtaining valid and objective data for further modification of the whole OSCE examination scheme for the upcoming semester.

REFERENCES

[1] Komenda M, Schwarz D, Blažková J, Štourač P. Complex medical simulation centre at the Masaryk University integrates innovative teaching modalities. In: SESAM Virtual Annual Meeting 2021, 2021.
Available from: https://www.muni.cz/en/research/publications/1771697

[2] Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. Ann Surg. 1995;222(6):735–42.

[3] Rushforth HE. Objective structured clinical examination (OSCE): review of literature and implications for nursing education. Nurse Educ Today. 2007;27(5):481–90.

[4] Karolyi M, Ščavnický J, Růžičková P, Šnajdrová L, Komenda M. Design and Management of an Objective Structured Clinical Examination using the SIMU portfolio Platform. In: Lane HC, Zvacek S, Uhomoibhi J (eds). Proceedings of the 12th International Conference on Computer Supported Education - Volume 1: CSEDU. Setúbal: SciTePress; 2020 p. 269–276.

[5] Harden RM. What is an OSCE? Med Teach. 1998;10(1):19–22.

[6] Elshama SS. How to design and apply an objective structured clinical examination (OSCE) in medical education?. Iberoam J Med. 2021;3(1):51–5.

[7] Kumaravel B, Stewart C, Ilic D. Development and evaluation of a spiral model of assessing EBM competency using OSCEs in undergraduate medical education. BMC Med Educ. 2021;21(1):204.

[8] Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. Med Teach. 2013;35(9):e1437–46.

[9] Karolyi M, Růžičková P, Šnajdrová L, Vafková T. Technology Enhanced Assessment: Objective Structured Clinical Examination Supported by Simuportfolio. In: Lane HC, Zvacek S, Uhomoibhi J (eds). Computer Supported Education. Cham: Springer; 2021. p. 302–316.

[10] Joshi M K, Srivastava A K, Ranjan P, Singhal M, Dhar A, Chumber S, Parshad R, Seenu V. OSCE as a Summative Assessment Tool for Undergraduate Students of Surgery—Our Experience. Indian J Surg. 2017;79(6):534–8.

SECTION B

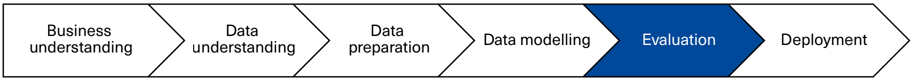
| 07 |

BUILDING CURRICULUM INFRASTRUCTURE IN MEDICAL EDUCATION

Martin Komenda, Jaroslav Majerník, Inga Hege, Andrzej A. Kononowicz, Adrian Ciureanu

The case study was supported by the following Erasmus+ grant project: BCIME – Building Curriculum Infrastructure in Medical Education – reg. no. 2018-1-SK01-KA203-046318, which was funded by the European Commission.

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2018–2021
Keywords	Curriculum structure, medical and healthcare education, data analysis, curriculum mapping
Research question	How to use apply a parametric and structured descriptive approach based on a standard-compliant model to deal with curriculum development, harmonisation, and reporting?
Type of result	Complex web application
Level of data processing	Descriptive statistics

DATA TO DOWNLOAD

INTRODUCTION

Today, there is no systematic solution based on proven pedagogical approaches and methodologies that can define, create, manage, and analyse the curricula of medical and healthcare institutions of higher education within a single robust system [1]. The BCIME project (Building Curriculum Infrastructure in Medical Education) delivered proven methodological background and a unified web-based platform for curriculum optimisation entitled EDUportfolio[1] at medical faculties in five European countries: Poland, Czech Republic, Romania, Germany, and Slovakia. Based on the local requirements of all involved institutions, the curricular descriptions of one common preclinical medical discipline (anatomy) and five different medical disciplines were completely mapped and optimised to reach modernised effects and educational practice owing to the application of the project outputs. Thanks to an open and modular approach of the platform's framework, the interdisciplinary study programmes were efficiently designed and managed, the duplicities in curricula were revealed and minimised, and missing components and teaching units were identified. The proposed open structure of curricular description and flexible methodology can be employed not only in medical and healthcare education, but also in any branch across higher education. Thus, it can be applied without restriction at other education institutions in partner regions.

The Faculty of Medicine at Masaryk University (MED MUNI) has already developed and run its curriculum management system (the OPTIMED platform), which was built on an outcome-based paradigm [2], which meant a fundamental paradigm shift in the curriculum design for many higher education institutions. It helped teachers, guarantors, curriculum designers and faculty management in the curriculum reengineering activities and provided a clear overview of the curriculum structure. Thus, inspired by MED MUNI piloting activities (using the EDUPortfolio platform), the BCIME project extended and transferred not only the knowledge, but also helped to effectively implement a proven curriculum management system, including a proper methodology, to structure various other curricula. Using this platform, selected individual study programmes (medical disciplines) were described by appropriate metadata including course attributes, learning units and outcomes, and links to a standardised vocabulary and essential terms. The project results assisted in the identification and location of potentially problematic areas (desirable or undesirable overlapping and missing topics) and, hence, helped to build a well-balanced overview for the subsequent global in-depth medical curriculum inspection.

[1] https://eduportfolio.iba.muni.cz/

AIMS

- To obtain a complete and valid specification of the tool for the creation and subsequent optimisation of the curriculum from stakeholders.
- To build a formal and general model suitable for any standard-compliant curricula.
- To organise the curriculum's components and explore their interrelationships to easily view, evaluate and maintain available content.

METHODS

From a methodological point of view, it was necessary to get exact and complete requirements and ideas about tools for curriculum development and subsequent optimisation across partner institutions. As these were five independent academic organisations, there were also different opinions on the critical features of the overall system. In addition, a detailed literature search was carried out to complement the themes obtained from the participating universities with experiences and outputs published in selected research journals.

NEEDS ANALYSIS

The first phase in the BCIME project was the needs analysis, where the specific individual institution's expectations and requirements from the complex curriculum management perspective were collected. The results of 31 feedback surveys extended by personal discussions from partner institutions in all five countries involved in the project were included in the needs analysis. These results showed a close connection to the curriculum development and harmonisation process. The BCIME team integrated all the relevant requirements into the curriculum management system, tested and implemented across partners' institutions. In addition, the variations in individual suggestions were carefully considered. If approved by all the partners, they extended the core features requested by all institutions, as reported in this survey. The curriculum system's key characteristics and features derived from the needs analysis included the following:

1. available online,
2. visual overview of the curriculum,
3. integration of different user roles,

4. export of curricula by course, study field, department, and faculty,
5. visual relations between various components of the curriculum,
6. possibilities to search by keywords,
7. integration of international recommendations,
8. possibility to modify reports and outputs according to the institutional requirements,
9. evaluation of learning objectives,
10. identification of redundancies in learning objectives,
11. outcome-based education compatibility,
12. complex reporting based on available curriculum building blocks.

Below is an example of evaluating one of the questions to check and control the curriculum's developed parts regarding students' competencies. The respondents (n = 12) claimed that it is very challenging to identify a set of competencies of what a student is expected to know, understand and/or be able to demonstrate after the end of the learning period, typically after the completion of individual courses. The median of the answers was 4 (More challenging), the mode of the solutions was 5 (Very challenging), mean was 4,36.

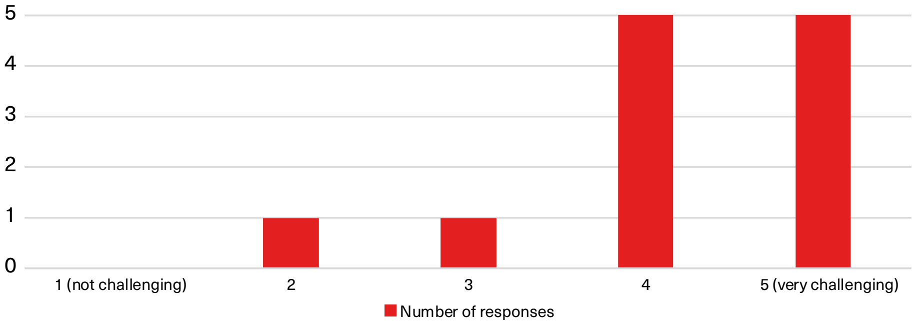


Figure 1: How challenging is it for partner institutions to identify a set of competencies of what a student is expected to know, understand and/or be able to demonstrate after completion of your courses? Scale: 1 = Not challenging at all to 5 = Very challenging

METHODOLOGICAL AND TECHNOLOGICAL TRENDS IN CURRICULUM INNOVATIONS

This part covers the compilation of recommendations and best practices to improve a platform for medical and healthcare curriculum management. It also helps to optimise curriculum innovations and mapping processes. The recommendations are based on the needs analysis that was used in combination with the expertise of partner institutions to compile optimal methodological material acceptable not only for curricula in medical education [3]. The twelve critical characteristics above were selected as a coding frame to structure the compilation of recommendations and best practices to improve the medical and healthcare curriculum management platform. Two primary sources of guidance were selected: (i) experiences collected by project partners in managing their curricula so far, and (ii) lessons learned on designing curriculum mapping software reported in the literature and published during the last five years. A list of search terms (Table 1) was proposed based on the literature collected on the topic before the review and our own experiences.

Table 1: Methodological and technical keywords in use in the literature search

Methodological/pedagogical keywords	Technical keywords
Curriculum map Curriculum development Curriculum management Medical curricula Healthcare curriculum Curriculum reform Curricular innovation Curriculum innovation	Software Management system Database Web-based Computer Online Online Mapping Tool Digital Electronic Information system Visual analytics Academic analytic Data mining Text mining Platform

Due to time constraints, the literature review was limited to the last five years, namely from January 2013 to January 2018. Moreover, immediate attention was paid to the literature query on the PubMed/Medline reference database, acknowledging this as a limitation of this study. This research resulted in valuable data describing the relevant domains that the project wanted to build on. The following list summarises the thematic overview detailed in one of the BCIME project deliverables (Intellectual Output II):

- Electronic curriculum management and mapping systems in use in the last five years
- Visual overview of the curriculum
- Integration of different user roles
- Export of curricula by course, study field, department, and faculty
- Visual relations between various components of the curriculum
- Possibilities to search by keywords
- Integration of international recommendations
- Possibility to modify reports and outputs according to the institutional requirements
- Evaluation of learning objectives
- Identification of redundancies [and gaps] in learning objectives
- Outcome-based education compatibility
- Complex reporting based on available curriculum building blocks
- Step-by-step manual on how to implement the requested changes
- Are those recommendations acceptable not only for curricula in medical education?

Combining the needs analysis and a thorough examination of published results, a final specification was created, based on which user stories were designed to describe the essential features of the system and the interaction with the user.

RESULTS

BUSINESS UNDERSTANDING

Among partners in the BCIME project, there was a common need for a uniform curriculum model providing a general and standardised way to describe education-related study building blocks and attributes using predefined parameters. To speed up and improve the long-term medical and healthcare curricula harmonisation process, the BCIME team adopted in-depth theoretical concepts supported by practical tools and systems. It brought an innovative and well-structured system for curriculum optimisation, which was easily applicable in practice. Based on a detailed needs analysis supported by the literature review, which generated a set of local institutional requirements related to curriculum organisation's goals, aspirations, and current features, BCIME defined a coherent and comprehensive framework in the form of a web application encompassing all necessary instruments and features for easy curriculum management. Five crucial target groups were identified including the main activities which they are responsible for: (i) Curriculum designers are experts who create the content of a particular learning unit, (ii) guarantors are experts who approve the content of a particular learning unit, (iii) heads of departments are senior experts who lead the particular department and supervise the entire teaching agenda, (v) teachers are employees affiliated with the faculty who can only browse the content, (v) students can browse the content and use all information related to the curriculum as supporting material in their study programmes. To thoroughly understand the curriculum structure and links between the curriculum building blocks and all descriptive attributes, a curriculum structure matrix (Table 2) was designed to demonstrate clearly how the different elements are interconnected.

Table 2: Curriculum structure matrix describing fundamental relations between building blocks and descriptive attributes

	Study programme	Medical discipline	Course	Learning unit	Learning outcome
Category of learning outcome	No	No	No	No	Yes
Assessment form	No	No	No	No	Yes
Range of learning unit	No	No	No	Yes	No
Type of teaching	No	No	No	Yes	No
Importance	No	No	No	Yes	No
Annotation	Yes	Yes	Yes	Yes	No
MeSH keywords	No	No	No	Yes	No
Important terms	No	No	No	Yes	No

DATA UNDERSTANDING

The entire data layer is based on the curriculum description, which makes it possible to define particular blocks in a parametric and thus structured way (e.g. for the study programme, medical discipline, course, learning unit, learning outcome). The data structure is compliant with international standards provided by the MedBiquitous association[2]. The BCIME team used several descriptive attributes to specify the curriculum in a structured form suitable for data processing (Table 3).

[2] https://www.medbiq.org/

Table 3: Example of curriculum descriptive attributes

Parameters	**Description**
Category	Program-level, Competency-level, Sequence block-level
Assessment form	Form of assessment of student's knowledge and skills
Duration of teaching	In teaching hours
Type of teaching	E.g. lecture, seminar, clinical practice, self-study, etc.
Importance	Initial brief explanation (why a building block occurs, what is its relevance and/or how it contributes to the learning outcomes)
Description	Summary of a building block along with educational goals and instructional description
Keywords	Free text based on standardised keyword form
Significant terms	Fundamental topic contained and explained during the teaching period in a tree structure
Study materials	Recommended literature or e-learning information source

A complete curriculum was designed from individual building blocks representing basic curriculum development units (Table 4).

Table 4: Components of curriculum building blocks

Components of building block	**Content/Value**
Study program	Medicine/General medicine (depending on partner country)
Medical and health discipline	Anatomy
Sequence block	Course, Module, Unit, Block, Clerkship
Event	Instructional or Assessment Session
Competence	Learning outcomes, competencies, learning objectives, professional roles, topics, classifications (measurable description of what students are to demonstrate in terms of knowledge, skills, and values)

DATA PREPARATION

The challenge of the BCIME project was to perform curricula innovations in teaching domains formalised with a detailed parametric description and entities adopted from the outcome-based education concept. In general, such innovations enhance the transparency and continuity of the environment where teachers, guarantors, curriculum designers, faculty management, and students work daily. The fundamental curriculum building blocks were formalised using the entity-relationship data model (Figure 2).

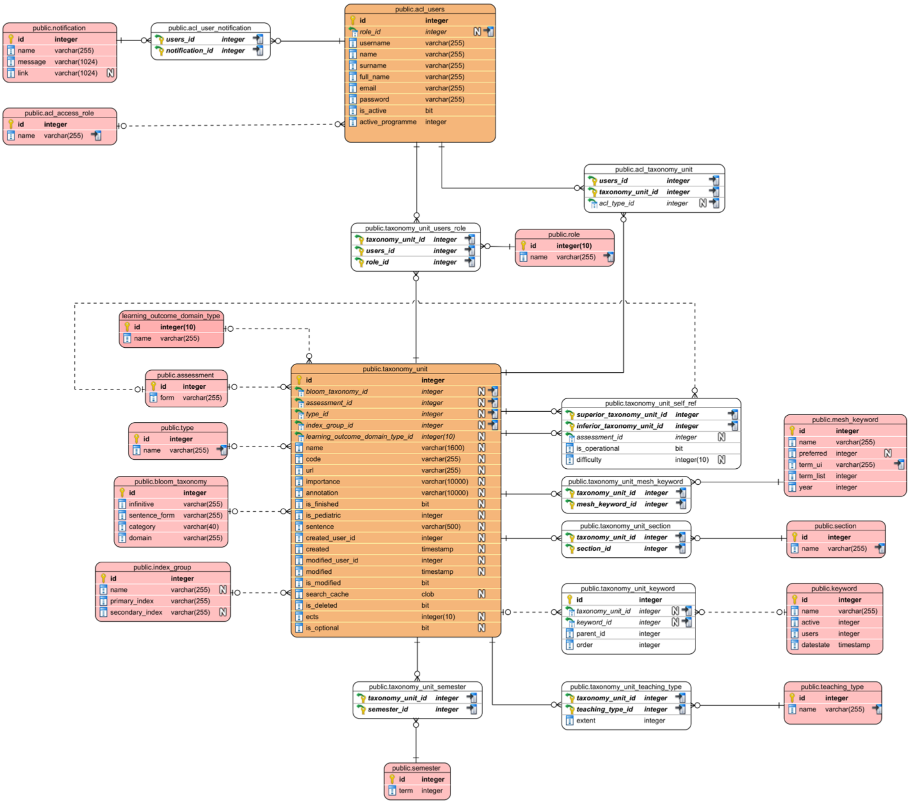
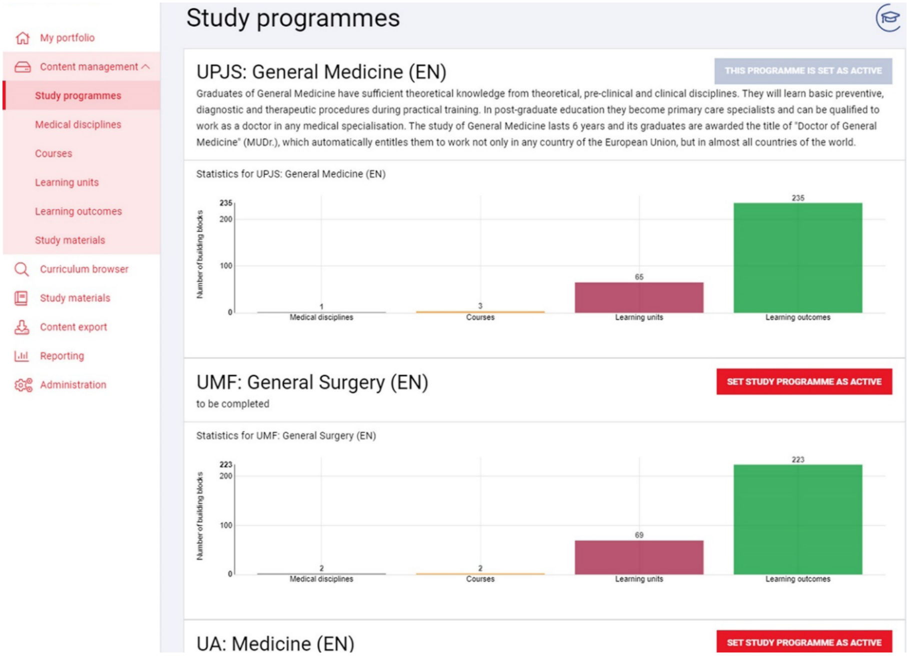


Figure 2: A database model presenting all fundamental curriculum building blocks

The data layer has been designed to be general enough to create and describe any object or link compatible with internationally accepted standards. The generality and robustness of this model proved helpful when data from partner institutions in parametric descriptions of their curricula were stored in a single database structure. The advantage was the unified import environment and the higher level structure at the application layer in the web application where data were accessible. In this particular case, the data processing phase means that from the primary data in heterogeneous and complex structures, a database was created to suit formal modifications and transformations of these data without affecting the information value and context, which enabled the processing of all relevant data in the form of entities and interrelationships.

MODELLING

As in many similar activities, modelling in the BCIME project was not specifically understood as the construction and application of a selected model. From a more abstract perspective, modelling can also be understood as the steps (using analytical and statistical methods) that lead to obtaining results over pre-processed data. The BCIME web application presented the available data from several different but interrelated perspectives. Below is an example of one of the imported curricula, showing a summary by core building block (Figure 3).



Study programmes

UPJS: General Medicine (EN)

Graduates of General Medicine have sufficient theoretical knowledge from theoretical, pre-clinical and clinical disciplines. They will learn basic preventive, diagnostic and therapeutic procedures during practical training. In post-graduate education they become primary care specialists and can be qualified to work as a doctor in any medical specialisation. The study of General Medicine lasts 6 years and its graduates are awarded the title of "Doctor of General Medicine" (MUDr.), which automatically entitles them to work not only in any country of the European Union, but in almost all countries of the world.

Statistics for UPJS: General Medicine (EN)

UMF: General Surgery (EN)

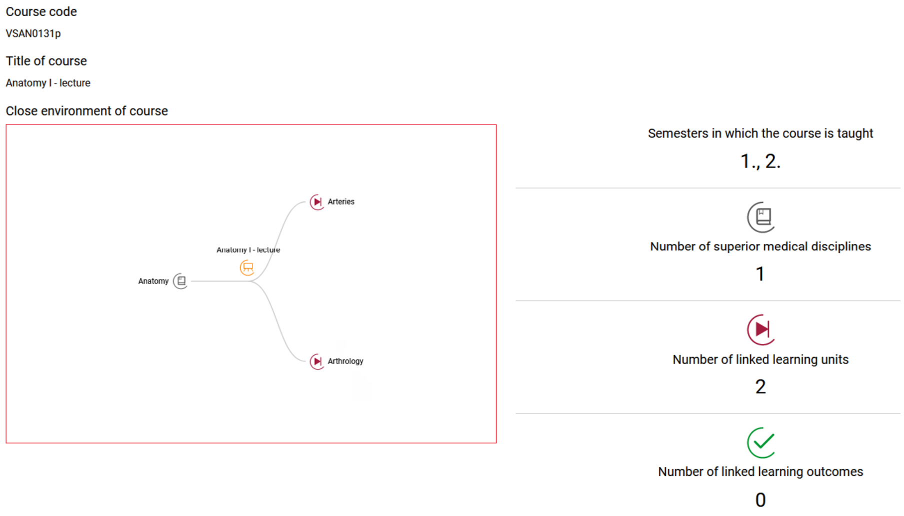
to be completed

Statistics for UMF: General Surgery (EN)

UA: Medicine (EN)

Figure 3: Curriculum development visualisation

A detailed overview can be intuitively used to visualise all interdependencies based on appropriately designed links between the individual objects for a better parametric description of the learning activities. Figure 4 illustrates a view of the teaching of Anatomy I course, which contains two teaching units.



Course code

VSAN0131p

Title of course

Anatomy I - lecture

Close environment of course

Semesters in which the course is taught

1., 2.

Number of superior medical disciplines

1

Number of linked learning units

2

Number of linked learning outcomes

0

Figure 4: Map of a particular course

Moreover, users can use summary reporting as a standalone module to see various aggregated statistics on curriculum data. Visualisation via bar charts represents a graphical overview of selected parts (learning units, learning outcomes, assessment forms, teaching range, as well as links between curriculum elements) and their level of elaboration. Vertical or horizontal multi-bar charts were built using the NVD3 JavaScript library for clear and interactive web visualisations.

EVALUATION

During this phase, the emphasis was mainly placed on in-depth quantitative and qualitative data analysis of available curriculum content, namely the learning outcomes and learning units, which have been reviewed and validated by content matter experts.

QUANTITATIVE ANALYSIS

Text-based descriptions of anatomy courses provided by each partner were exported from the BCIME application. Figure 5 shows the total number of learning units (red) and learning outcomes (purple), and it indicates a heterogeneous level of granularity in the mapping – a well-known challenge in curriculum mapping. For instance, Universitaet Augsburg in Germany defined no learning units because of the module-oriented structure of the entire curriculum.

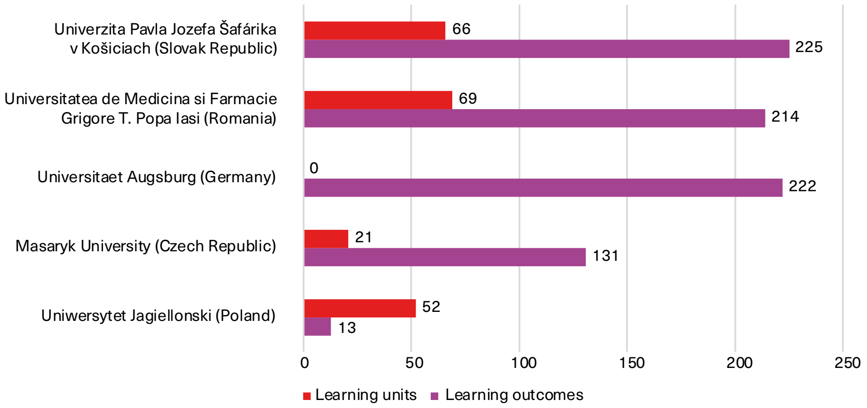


Figure 5: Number of learning outcomes in anatomy courses

Figure 6 shows a word cloud visualisation of the learning objectives. Terms used the most often include "explains", "structures", "systems", "describes," and "functions".

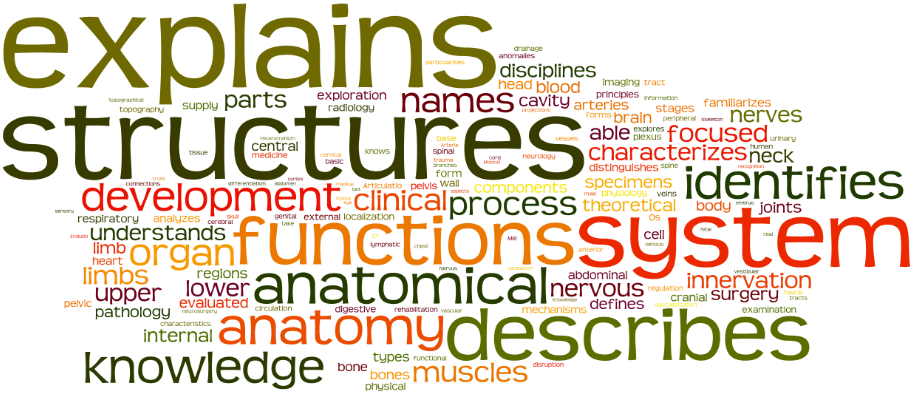


Figure 6: Word cloud of learning outcomes and learning unit descriptions of all partner anatomy curricula

QUALITATIVE ANALYSIS

Two experienced BCIME team members with a health profession education background coded all learning outcomes deductively and independently. Divergent coding items were solved by discussion, and consensus was reached in all cases. For coding, a guideline has been developed that covers categories mentioned in Table 5.

Table 5: Overview of coding categories

Category	Items
Learning category	Cognitive Psychomotor Affective
Bloom's knowledge dimension	Factual Conceptual Procedural Metacognitive
Bloom's cognitive process dimension	Remember Understand Apply Analyse Evaluate Create
MeSH chapter "Anatomy" and its following main categories	Body regions (A01) Musculoskeletal system (A02) Digestive system (A03) Respiratory system (A04) ...

Table 6 shows the frequencies of learning objectives action verbs in Bloom's cognitive process dimension for the anatomy curricula of all partners. For three partners (Masaryk University, Universitaet Augsburg, Univerzita Pavla Jozefa Šafárika v Košiciach), the "Understand" level is the most prevalent, whereas "Remember" is most common for Uniwersytet Jagiellonski and "Apply" for Universitatea de Medicina si Farmacie Grigore T. Popa Iasi. For all partners, the "Create", "Evaluate", and "Analyse" levels are less frequent.

Table 6: Bloom's most prevalent cognitive process levels for each partner are marked in bold

	Remember	Understand	Apply	Analyse	Evaluate	Create
Uniwersytet Jagiellonski (Poland)	5 (55.6%)	2 (22.2%)	2 (22.2%)	0	0	0
Masaryk University (Czech Republic)	27 (20.8%)	57 (43.8%)	31 (23.8%)	1 (0.8%)	13 (10.0%)	1 (0.8%)
Universitaet Augsburg (Germany)	2 (0.9%)	201 (90.5%)	6 (2.7%)	3 (1.4%)	9 (4.1%)	1 (0.4%)
Universitatea de Medicina si Farmacie Grigore T. Popa Iasi (Romania)	43 (20.6%)	63 (30.1%)	77 (36.8%)	24 (11.4%)	0	2 (0.1%)
Univerzita Pavla Jozefa Šafárika v Košiciach (Slovak Republic)	58 (25.9%)	133 (60.4%)	5 (2.2%)	25 (11.2%)	1 (0.5%)	2 (0.8%)
Total	135 (17.0%)	456 (75.4%)	122 (15.3%)	53 (6.7%)	23 (2.9%)	6 (0.7%)

WEB APPLICATION TESTING

The development and implementation phase also involved the testing procedures, while testing of the functionality of all web application elements was performed in three main steps.

1. The web application was filled with simulated content to determine the accessibility and functionality of all features. This approach proved helpful as it revealed many bugs and errors that have been fixed before filling the database of the platform with real content.

2. The functional and user acceptance tests were performed continuously and repeatedly. The individual developed and modified functions and features of the system were verified at the level of source code by the technicians involved in the project team (tested whether the system and its components work). Later, the user acceptance tests (tested how the system works) were performed by the curriculum designers, teachers, and researchers of the BCIME consortium to ensure that all the requirements had been met and that the web application operated as expected. The tests were oriented at different user groups, mainly curriculum designers and learners. A broad range of quality control test cases were designed and checked to verify that the developed curriculum management platform is in conformance with the requirements and thus acceptable to the users.

3. Users' feedback was invaluable to the acceptance and quality testing process. Here, the various scenarios were designed and evaluated using the user test cases to make the system testing processes easier and faster. The tests had to confirm that the curriculum management platform was ready for operational use and met the initial requirements. Also, the test scripts were developed to be repeatedly used with any system changes, including its updates and verification that no bugs remained after improvements and new features were implemented. The system functional testing was prepared to be performed by semi-automated and automated methods, depending on input data needed to be specified during the execution of a particular test. The individual user stories that were prepared to be used anytime required (after the system development is completed, the design changed, new functions implemented, changelogs applied, bugs repaired etc.) include several functional test scenarios (e.g. authorised log-in procedure verification, curriculum browsing by medical disciplines, or learning unit update). The individual test scripts were developed using Selenium open-source testing tools[3]. We decided to

[3] https://www.selenium.dev/

use Selenium for our testing tasks because it is compatible with various programming languages, testing frameworks, browsers, and operating systems. Being used by many developers in the automation testing of web apps, it was beneficial in creating our test scripts for exploratory testing. However, we also use these scripts in continuous regression testing and quick bug reproduction. The system bugs discovered by the testing scripts were noted and used to formulate the system's changelogs. These changelogs of all improved features and bug fixes have been reported to the development team and archived for later usage and documentation of changes that were done.

DEPLOYMENT

The individual partner institutions used one common and one individual instance of the EDUportfolio platform, which provides a tool for the description and visualisation of the medical curriculum. All instances were launched within the MED MUNI infrastructure. The unified graphical user interfaces, navigation elements, and functionalities enabled the involved faculties to define structured descriptions of selected parts of their curricula systematically and controlled. Dedicated teams at the partner institutions, thus were able to independently enter data into this system, perform an internal review, and, finally, an external examination by the guarantors of the medical discipline from another academic institution.

DISCUSSION

Using the outputs of the BCIME project, curricula designers across involved partner institutions were equipped with visual tools and feedback, allowing them to perform an effective administration of particular study blocks as well as to communicate with teachers and educators to adopt curricula to the recent trends and knowledge. Therefore, the whole process of curricula management was improved, visualised and simplified. The developed curriculum mapping application allows to transform a present clearly the descriptions of curricula for the students via the optimal distribution of courses, lessons, and topics. Following the student-centred approach, the curricula will integrate recent professional skills and, most probably, modern teaching methods based on new technologies. The students of all partner institutions will benefit from a higher quality of educational materials that international curriculum development teams can now produce. Based on the in-depth analysis in the mapping and evaluation phase, the BCIME team recommended describing curricula at a high level of learning outcome granularity. Such fine-granular outcomes are more specific, helpful for learners, and valuable for analysing a curriculum at institutional and faculty levels. Additionally, providing an overarching level of granularity under which the fine-granular outcomes are subsumed can help educators to get a quick overview of a curriculum.

EVALUATION OF THE AIMS OF THE CHAPTER

- To obtain a complete and valid specification of the tool for the creation and subsequent optimisation of the curriculum from stakeholders.
  - A systematic search of literature published in the last five years, enriched with the experiences of the five project partners in the form of a survey, resulted in a comprehensive overview of what and how is nowadays implemented or desired in curriculum mapping software.
- To build a formal and general model suitable for any standard-compliant curricula.
  - Based on the successful transfer of experiences with the complex process of curriculum description, review, and mapping, especially by MED MUNI, a general model for a fully standardised curriculum independent of the type of study programme has been proposed with the active involvement of all partners.
- To organise the curriculum's components and explore their interrelationships to view, evaluate, and maintain available content easily.
  - Based on a complete understanding of the curriculum model, the EDUportfolio platform supporting a proven methodological background allows not only essential descriptive characteristics to be visualised but also produces mapping diagrams and performs searches curricula over keywords or phrases to identify previously uncovered topics.

LESSONS LEARNED

A partner consortium of five European medical faculties was involved in adopting proven practices from the Faculty of Medicine of Masaryk University in developing descriptions of one part of the educational process. In this case, it was a demonstration of the transferability of experience and further development of a practice-tested procedure for curriculum data collection and subsequent analysis. The involvement of each of the partners brought new suggestions and requirements towards the local needs of each faculty; however, it was necessary to find consensus in the implementation of the whole project. The evaluation stage, where the qualitative and quantitative assessment of the results achieved was carried out, can be considered as crucial in this case.

The parametric and structured concept of curriculum description, which also fully takes into account international standards, contributed significantly to the design and fulfilment of the data basis for the follow-up processing and online interactive reporting over data of all partners involved.

REFERENCES

[1] Majerník J, Komenda M, Kačmariková A, et al. Development and implementation of an online platform for curriculum mapping in medical education. Bio-Algorithms Med-Syst. 2022;18(1): 1–11.

[2] Harden R M. Outcome-based education: the future is today. Med Teach. 2007;29(7):625–9.

[3] Kononowicz AA, Balcerzak Ł, Kocurek A,et al. Technical infrastructure for curriculum mapping in medical education: a narrative review. Bio-Algorithms Med-Syst. 2020;16(2):20200026.

SECTION B

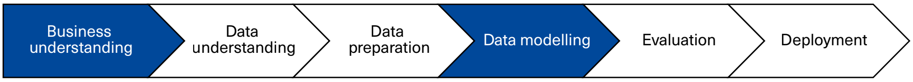
| 0 8 |

VISUALISATION OF TEXT-BASED DATA DESCRIBING BEST PRACTICES IN SOFTWARE DEVELOPMENT

Martin Komenda, Rudolf Ramler, Petra Růžičková, Doris Hohensinger, Michaela Kerberová, Mario Pichler

The case study was supported by the Interreg V-A Austria-Czech Republic programme within the project SESSAD (Sharing Experience in Scientific Software and Applications Development) (ATCZ292).

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2022–2022
Keywords	Scientific software, development, sharing experience, best practice
Research question	How to do inter-disciplinary best practice collection and visualisation in a domain of scientific software development?
Type of result	Static analytical report
Level of data processing	Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

In natural sciences, medicine and healthcare, digital solutions enable innovations and scientific advancements more than ever. Nowadays, with scientific software and applications, research, knowledge transfer and education can be made more accessible, more efficient and more available. Many applications are already available that support education and those that simplify daily duties or promote health literacy and prevention for each individual. Most prominently, since the beginning of the COVID-19 pandemic, software solution companies have played a central role in collecting and analysing data for visualising and predicting infections and developing, testing and distributing vaccines. Scientific software solutions are becoming increasingly large and complex, using the latest technologies such as artificial intelligence, big data, machine learning, image processing, simulation, and visualisation. At the same time, these solutions have to serve research with the highest standards regarding correctness, precision, performance, traceability, and explainability. Scientific software tools and applications cannot be developed by a single person or organisation anymore because there are so many specialisations to be focused on – data analysis, software engineering, testing, virtual reality etc. To fulfil all requirements, scientific software has to apply best software engineering practices and requires a joint effort across several teams and organisations and interdisciplinary collaboration.

The SESSAD project (Sharing Experience in Scientific Software and Applications Development) aims to share best practices and pitfalls between cross-border technically-oriented teams designing and developing scientific software solutions in the academic and non-profit space, specifically in medical education and simulation education. The project outputs will help the development teams to rapidly improve inter-regional cooperation and become the basis for collaboration within the Czech-Austrian network of development institutions.

Research and education institutions from the Czech Republic (Faculty of Medicine, Masaryk University – MED MUNI) and Austria (Software Competence Center Hagenberg – SCCH), which develop scientific software tools and applications independently or collaborate on the development with external providers, were involved in this activity. A series of workshops, shadowing activities, meetings and discussions were organised during this one-year project. The project output is a set of guidelines summarising best practices in international collaboration for institutions focused on scientific software development. The main benefits for both sides are: (i) sharing innovative know-how in scientific software design and development, especially in digital health, research, and technology-enhanced medical education, and (ii) building advancement by knowledge exchange and best practices in scientific software engineering across regions.

AIMS

- To design structured feedback forms mapping experience in software development based on cooperation.
- To identify appropriate target groups with expertise in a complex software development life cycle.
- To collect, process, and visualise text-based data describing particular best practices.

METHODS

Individual meetings, specialised workshops, and dedicated shadowing activities planned in this project led to identifying key and supporting topics in designing, developing, implementing, testing and operating web-oriented applications used for medical and healthcare education. In addition, attention was also focused on interactive discussions between the teams, real demonstrations of deployed web applications, as well as examples of good and bad practices and experiences in the complex development process. In addition to software applications, the graphics team also presented their outputs for MED MUNI projects, as each project is closely related to this domain and requires the creation of various graphic materials, both online and offline. Based on all of the above-mentioned activities, structured feedback for mapping best practice overview was designed by the SESSAD project team. The survey was implemented in practice as two related parametric questionnaires shown below. At first, attention was mainly focused on creating a best practice feedback form. It was created gradually during several meetings and workshops of both involved project partners. The first shadowing activity defined so-called issues in the design, development, implementation and testing of scientific software development. Subsequently, the primary structure of the feedback form and the tool for its creation were discussed. An online form is the best possible option for our purposes: it is easy to share, and data are stored via a Google spreadsheet in a structure that reflects the setting of individual questions, so subsequent processing is feasible.

The correct questionnaire design can affect both the quality and accuracy of the information obtained. That is why, when compiling the form, emphasis was also placed on the possible visualisation of data obtained in this way. The first draft of such a questionnaire was based on issues from scientific software development, evaluation of their importance, understanding and adoption by both project partners. In the subsequent discussion, the questionnaire was transformed from issues (that were very general) to specific best practices. In addition, a form with respondents' personal information was created, which was very important for evaluating the best practice feedback form. Both documents are described in detail below.

FORM A: PERSONAL INFORMATION

This form collected personal characteristics about individual members involved in the SESSAD project and possibly other partners of both institutions. Thanks to this form, it was possible to map personal answers from Form B (see below) to evaluate individual characteristics. The unique form items were as follows:

- First name
- Last name
- E-mail
- Professions
- Length of experience in the field

FORM B: BEST PRACTICE FEEDBACK FORM

This form helped to collect best practices from the fields of expertise of both technically oriented partner teams. It brought an insight into what motivated a particular team member to use this best practice, which domain they would assign it to and in which projects they had already used it. The form was used to evaluate best practices in scientific software development. The results served as a basis for creating a methodology for developing cooperation between cross-border development teams. The individual items of the feedback form were as follows:

- E-mail
- Title of best practice
- Description: a short description of selected best practice
- Links: related links
- Keywords: related to the topic
- Motivation: why it is profitable to follow this practice
- Requirements/prerequisites: anything needed before using the practice
- Level: generic, concrete or activity – what is the level of the chosen practice
- Application domain: the area in which the practice is used
- Main phase: the stage of development in which the practice is used

- Related literature, such as papers, books, articles, blog posts
- Projects in which the respondent applies the given practice
- Evaluation of the benefits (of using this practice) on respondents' projects and the reason for this evaluation
- Frequency of use of the selected practice
- Effort to implement the practice upfront
- Effort to apply the best practice in projects daily basis
- Organisation

RESULTS

BUSINESS UNDERSTANDING

An innovative approach to the project is sharing experiences and subsequent publication of results in software development. This is based on identifying key domains where members of the two institutions are strongly involved: advancements in the design, development, implementation and testing of web applications, as well as graphic design. These applications focus on supporting medical education and simulation-based learning, combining modern techniques with teaching and breakthrough technologies. The project team easily and comprehensibly obtained the required information to be processed into methodological guidelines through the designed and internally opposed feedback. The methodology provided a basis for developing cooperation between the partner institutions and other similarly oriented entities in interregional cooperation. The web presentation[1], which is one of the outputs of the project, contains thematically coherent sections with a clear focus on individual areas of development as well as project management (communication with the client, specification of the assignment, design phase, systematic development, user and automated testing, implementation and deployment of the product, operation and management, incorporation of feedback).

Each topic, as an integral part of the development cycle, was designed to present proven practices, working recommendations, and pitfalls and mistakes within each phase of the software development lifecycle. The exchange of know-how between partner institutions took place mainly in the programme area. A necessary prerequisite was the translation of selected insights and recommendations into internal processes within the individual stages of the design

[1] https://www.med.muni.cz/sessad

and development of software solutions. A fundamental idea of the project was that all activities planned in the framework of know-how acquisition and transfer of experience would be carried out on the current development projects of both partners. To successfully implement the experience data collection, it was crucial to understand the focus of the development teams, their deliverables and internal setup.

DATA UNDERSTANDING

The structure of feedback forms is shown in detail above. For the subsequent in-depth data analysis, it was necessary to perform an exploratory study. Data from both forms are mainly textual; the user was not restricted and could write any text string in the answer. In some questions, the respondent chose from a pre-given offer with the possibility of adding their response (e.g. form B: application domain, main phase) or rating the given practice (e.g. form B: evaluation of the benefits, frequency of use) on levels from 1 (low) to 5 (high). We have received: (i) 22 responses from the form with personal information (form A) and (ii) 53 responses from the best practice form (form B).

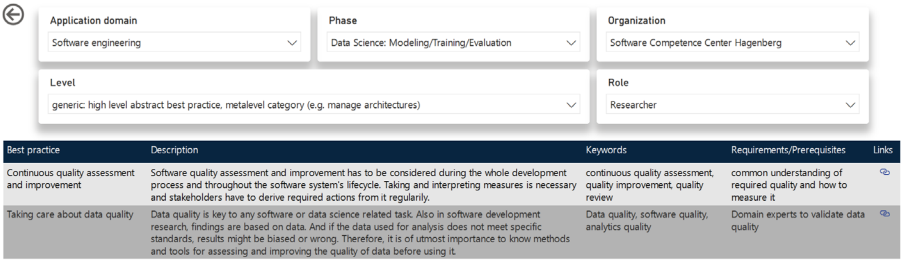
DATA PREPARATION

During the preparation phase, the main emphasis was on data validation from both forms. Since almost all form items were mandatory, there were no missing values. The inspection of both documents was carried out manually and semi-automatically using pivot tables and graphs. The data cleaning process took place mainly during validating e-mail addresses in both forms to map responses with individual personal information. Ultimately, individual best practices were mapped per user according to the completed e-mail in both forms.

MODELLING

DATA TABLE WITH A FILTERING FEATURE

The first in a series of deliverables was an interactive summary table with all recommendations and best practices across the MED MUNI and SCCH teams. Its primary purpose was to display basic information, including the title, brief description, keywords, prerequisites or input requirements. In addition, this aggregation had a set of filters that included the application domain, level (generic, concrete, activity), the phase of the development cycle the particular record falls into, the organisation, the professional role of the author, and direct link to the profile page, where the given description was shown in detail.



Application domain: Software engineering

Phase: Data Science: Modeling/Training/Evaluation

Organization: Software Competence Center Hagenberg

Level: generic: high level abstract best practice, metalevel category (e.g. manage architectures)

Role: Researcher

Best practice	Description	Keywords	Requirements/Prerequisites	Links
Continuous quality assessment and improvement	Software quality assessment and improvement has to be considered during the whole development process and throughout the software system's lifecycle. Taking and interpreting measures is necessary and stakeholders have to derive required actions from it regularily.	continuous quality assessment, quality improvement, quality review	common understanding of required quality and how to measure it	🔗
Taking care about data quality	Data quality is key to any software or data science related task. Also in software development research, findings are based on data. And if the data used for analysis does not meet specific standards, results might be biased or wrong. Therefore, it is of utmost importance to know methods and tools for assessing and improving the quality of data before using it.	Data quality, software quality, analytics quality	Domain experts to validate data quality	🔗

Figure 1: Summary data table with all recommendations and best practices

VISUALISATION OF APPLICATION DOMAIN

Another one of the many outputs was a clear visualisation of the different domains (e.g. medicine/healthcare, data science, industry) in which the MED MUNI and SCCH teams are involved in the long term, and where each recommendation was classified. A total of 166 items identified that all 53 recommendations and best practices were assigned to one or more domains. Given the multidisciplinary nature of both teams, this was not surprising. Software development, data analytics, and education had the most significant representation.

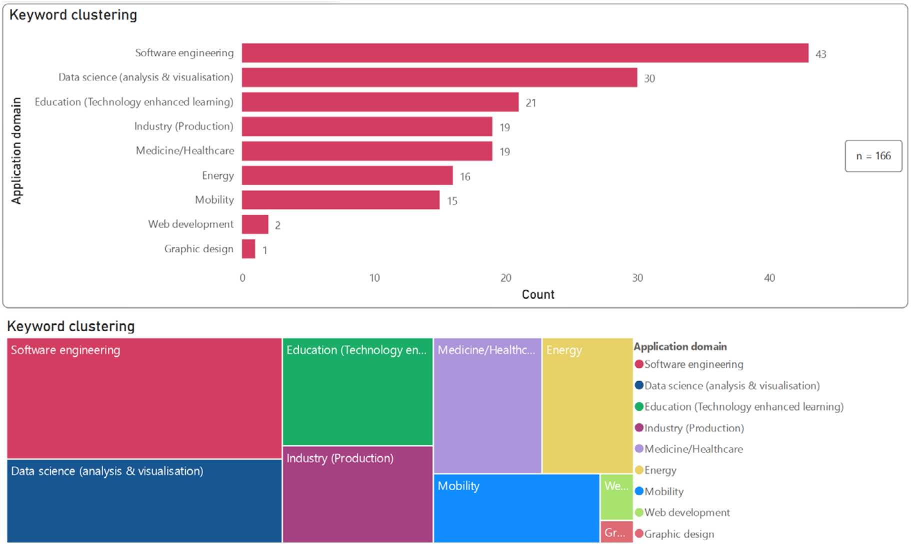


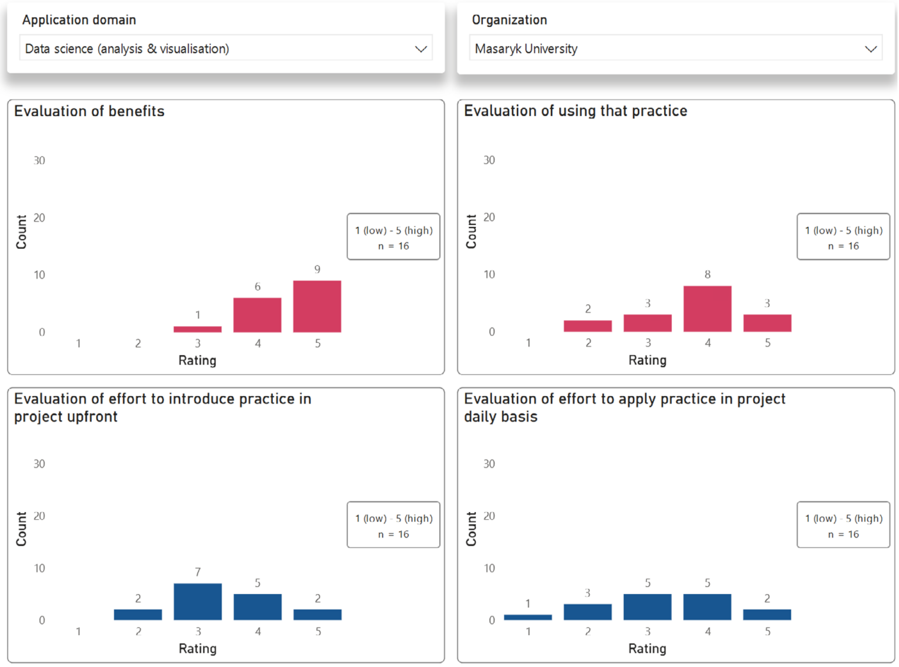
Figure 2: Cluster-based visualisation of application domains

DEVELOPMENT LIFE CYCLE EXPERIENCE PATH

Collecting, evaluating and visualising quantitative ratings of each recommendation on a scale of 1 (low/never) to 5 (high/always) was aimed at obtaining expert perspectives from MED MUNI and SCCH team members. The questions below focused on the practical application of the recommendations in real projects, including critical perspectives on the frequency and overall implementation efforts.

- How do you rate the potential benefit for your projects?
- How often do you use that practice?
- What is the effort to introduce the practice in your project upfront?
- What is the effort to apply the best practice in your project daily?

The visualisation below (Figure 3) contains the quantitative evaluations in the application domain of data science, which were entered only by members of the MED MUNI team (two various filters of available data were applied). In general, the assessment of the recommendations' overall benefits was optimistic. However, many experts need more support in implementing the suggestions.



Application domain: Data science (analysis & visualisation)

Organization: Masaryk University

Figure 3: Four quantitative questions overview

EVALUATION

In validating and evaluating individual text and graphical outputs presented in this case study, the primary input data obtained from the questionnaires and subsequent export were compared with the preprocessed data for the final visualisation using Microsoft Power BI. Due to a relatively small volume of the data, descriptive analysis in Microsoft Excel using contingency tables and graphs was sufficient for validation.

DEPLOYMENT

The complete outputs of the SESSAD project are published on the official website[2] in English, Czech and German. Reports from the attendance events and the project timetable are also presented there. The visualisations were produced in the selected business intelligence tool and embedded on the web in the iFrame form. This allows linking to the profile pages of each recommendation and the possibility of full-screen zooming or convenient sharing.

DISCUSSION

The main objective of the SESSAD project (Sharing Experience in Scientific Software and Applications Development) was to establish bilateral cooperation between cross-border development teams involved in designing and developing scientific software solutions in the academic and non-profit space, specifically in medical and healthcare education, data science, and simulation education. The methodological guidelines in the form of recommendations and best practices enabled the MED MUNI and SCCH teams to identify weak domains of their development life cycle. It also helped to emphasise inter-regional cooperation and became the basis for collaboration within the Czech-Austrian network of development institutions. A set of interactive text-based and graphic visualisations presented the main output of this project, including guidelines based on the evidence of best practices and recommendations in software development. These overviews on the available data reported global data table overview, application domains overview, word cloud, and summary of the Likert scale quantitative evaluations.

[2] https://www.med.muni.cz/sessad

EVALUATION OF THE AIMS OF THE CHAPTER

- To design structured feedback forms mapping experience in software development based on cooperation.
  - Two structured forms for collecting expert recommendations and best practices, including a quantitative assessment of real-world applicability, were designed in collaboration between two interdisciplinary development teams (MED MUNI, SCCH). The basis was a very intensive interaction between the team members to present the project, personnel, technical and methodological agenda in detail.
- To identify appropriate target groups with the given expertise in a complex life cycle of software development.
  - After implementing several face-to-face project workshops, target groups were specified and then actively involved in collecting recommendations. An interdisciplinary focus, the overall overview of complex design and development issues, and the field's specialisation and expertise are significant benefits.
- To collect, process, and visualise text-based data describing particular best practices.
  - The data collection process, through appropriately designed and structured questionnaires, provided a relatively easy stage of validation and subsequent visualisation of the results in the form of several different views of the available data. Each had another purpose, and together, they resulted into a comprehensive interactive online report of the recommendations.

LESSONS LEARNED

Cross-border cooperation brought not only the exchange of experience in the field of design and development of web applications but also data containing structured feedback from individual team members towards good and proven practices, as well as recommendations at different stages of the software development life cycle. Suggestions on what and how to collect, how to define relevant domains and how to design an online form were very closely related to understanding how both teams operate, what they do and what technologies they work with. Modelling and using business intelligence tools was the second time-consuming phase, but it produced original and insightful results.

The appropriately parametrically structured and pre-prepared data, which matched the content needs of the project, was interactively displayed using the robust Microsoft Power BI platform.

SECTION C

HEALTH INFORMATION AND STATISTICS

STAKEHOLDER OPINION

Radka Domanská

NATIONAL OPEN DATA COORDINATOR

Digital and information agency, Prague, Czech Republic

We often hear of data being described as the new oil of the 21st century. Data can indeed drive the ongoing digital revolution, improvements in decision-making, and cost savings in healthcare and many other areas. To do so, however, special attention and effort must be put towards collecting, describing, sharing, analysing, and effectively using it. Only then can we usher in an entirely new era in which we completely transform how diseases are understood, prevented, and treated. Data can reveal hidden patterns and correlations and help us make sense of complex problems. It can enable doctors and health professionals to create personalised treatment plans or to predict disease outbreaks and stakeholders to monitor and evaluate factors that could improve the efficiency of healthcare delivery.

The essential prerequisite for data processing is data availability. Especially in the healthcare field, a lot of sensitive patient data is collected and cannot be published without aggregation or anonymisation. This book proves how even health data can be provided as open data with this fact in mind while demonstrating how to describe data in a way that minimises misinterpretations and mistakes that might otherwise become fatal. Examples of data visualisations show how data can be shared with citizens or stakeholders who play a crucial role in healthcare and prevention decisions but may not know how to extract valuable information from data (or have time to do it) using statistical methods.

Sharing case studies, like the ones in this book, is a convenient way to inspire both public and private institutions to provide their data and to guide scientists or students towards processing data effectively. Identifying and understanding data-driven information will enable our society to make informed healthcare decisions that can save lives and enhance the quality of life for millions of people.

Václav Moravec
JOURNALISM RESEARCHER
Charles University, Faculty of Social Sciences, Institute of Communication Studies and Journalism, Prague, Czech Republic

When Luciano Floridi [1], the founder of the modern philosophy of information, characterises the contemporary advanced information society, he uses the metaphor of the mangrove, a fascinating community of shrubs and trees growing in river deltas where fresh water mixes with salt water. Using this analogy, he tries to draw attention to the fact that we do not live only in an online or offline world, but in a seamless onlife environment. In other words, we live in an infosphere that is analogue as well as digital, offline as well as online. So it has a special nature, like the mixing of fresh and salt water in which mangroves thrive so well. The contemporary infosphere, the imaginary 'mangrove community', is characterised by a significant amount of data, hence the terms datafication, data culture, data society, data economy, data capitalism, data elite and global data industry. In short, data and its accumulation are a key part not only of the political economy of the 21st century, but of society as a whole.

Sociologist and economist Jathan Sadowski [2] has identified five main ways in which value is extracted from data. First, data is used to profile and target people, for example in personalised advertising or news reporting, or to assess the financial risk and trustworthiness of a particular person. Second, data can optimise systems by identifying inefficiencies. Third, data is used to govern, manage and control the objects and entities that provide the data - thanks to a power relationship based on lack of knowledge. Fourth, data can model probabilities when, for example, city dispatch centres process a continuous stream of data to create simulations of events such as accident and disaster response. Finally, data is used to create digital products or to innovate existing ones, making them useful. Sadowski's model of transforming data into value can undoubtedly be applied to healthcare, as the following chapters illustrate.

The health crises of the early 21st century, led by the COVID-19 pandemic, have exposed another fundamental factor in the inexorable logic of the contemporary data society. Ignorance is encountered not only in the absence of relevant data, but also in the overload caused by too much data, too much information. Having too many data points at hand, and therefore "knowing too much", can have as debilitating an effect on an individual or organisation's decision-making as too little or no information. In the words of statistician Nate Silver [3], where there is too much data, there is also a lot of noise that is hard to separate from the signals, and where there is too much noise, organisational actors may end up with an even more intense sense of knowing too little or not yet knowing what is needed. As the historian Stefan Schwarzkopf [4] points out, data overload can trigger new forms of ignorance, such as the inability to generate and process information, as politicians, managers, doctors or journalists are overwhelmed by ever more incoming data, which in some cases even leads to an inability to make decisions or an unwillingness to engage further with more incoming data, regardless of its relevance.

The following chapters are a contribution to the cultivation of the contemporary infosphere so that the imagined "mangrove community" thrives, does not face ignorance from lack or excess of data, turns data into value with an emphasis on social responsibility, and does not see data as the preserve of the data elite.

REFERENCES

[1] Floridi L. Soft ethics and the governance of the digital. Philos Technol. 2018;31:1–8.

[2] Sadowski J. When data is capital: Datafication, accumulation, and extraction. Big Data Soc. 2019;6(1):2053951718820549.

[3] Silver N. The signal and the noise: Why so many predictions fail – but some don't. Penguin Books; 2015.

[4] Schwarzkopf S. Sacred excess: Organizational ignorance in an age of toxic data. Organ Stud. 2020;41(2):197–217.

SECTION C

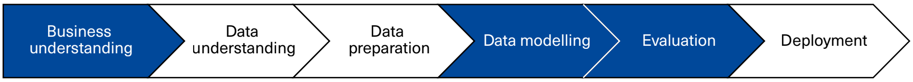
09

CANCER SCREENING PROGRAMMES: OPEN DATA AND VISUALISATIONS AS A SUPPORT TOOL FOR MONITORING AND EVALUATION

Martin Komenda, Renata Chloupková, Karel Hejduk, Petr Benáček, Ondřej Ngo, Lenka Šnajdrová, Ladislav Dušek, Ondřej Májek

This work was supported within the framework of the European Social Fund, Operation Programme Employment (grant no. CZ.03.2.63/0.0/0.0/15_039/0007216) and Operation Programme Employment+ (grant no. CZ.03.02.02/00/22_005/0000283).

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2017–now
Keywords	National screening centre, cancer care, cervical screening, breast screening, colorectal screening, National health information system
Research question	How to build a robust interface containing a complete information portfolio of selected screening programmes based on data from the National health information system in the Czech Republic?
Type of result	Interactive visualisation
Level of data processing	Advanced analyses, Open datasets

DATA TO DOWNLOAD

INTRODUCTION

Well-managed early detection programmes for selected diseases are an effective tool for reducing their morbidity or mortality. Therefore, early detection programmes are among the national health priorities of the Czech Republic, as declared in the national strategies Health 2020[1] and Health 2030[2]. The National Screening Centre (NSC)[3] has established and has been providing an umbrella system to support screening programmes in the Czech Republic throughout their life cycle (planning, pilot studies, programme implementation, programme monitoring, programme evaluation, and programme innovation) to ensure their maximum positive impact on population health together with high cost-effectiveness. Information support systems and health literacy development are also integral to these activities, forming an essential part of an organised screening programme. The 2003 EU Council Recommendation[4] states that the screening process and its results should be regularly monitored and that this information should be made available to the general public and the institutions involved in screening in a short period. The information system should therefore be capable of collecting, processing and evaluating data on all screening tests, additional diagnostic tests and final diagnoses. In the Czech Republic, this system includes monitoring the cancer burden in the population, the screening process through clinical data, and the screening process through administrative data.

Valid, complete and correctly interpreted data are undoubtedly an essential tool for evaluating and modifying existing screening programmes. Data also play an important role in deciding whether new programmes should be introduced. One of the objectives of the NSC is to build a unified and coordinated infrastructure for collecting and evaluating data from ongoing and newly introduced screening programmes in the Czech Republic. The conceptual system for evaluating screening programmes in the Czech Republic includes the involvement of a dedicated analytical team and tools for collecting, analysing and interpreting specialised data on early detection programmes for selected diseases. The consolidation and possible supplementation of data sources on the implementation of screening programmes are the basic building blocks for the publication of open datasets supplemented by various types of static or interactive reports, whose content and form meet the requirements of individual target groups.

[1] https://www.mzcr.cz/zdravi-2020-narodni-strategie-ochrany-a-podpory-zdravi-a-prevence-nemoci-2/

[2] https://www.mzcr.cz/category/programy-a-strategie/zdravi-2030/

[3] https://nsc.uzis.cz/

[4] http://data.europa.eu/eli/reco/2003/878/oj

AIMS

- To understand cancer screening issues focusing on breast, cervical, and colorectal cancer.
- To design pilot datasets describing coverage of the target population by selected screening examinations.
- To implement an interactive visualisation of published data.

METHODS

The creation of the NSC data basis helped to systematically implement the methodology for processing selected components of the National Health Information System (NHIS)[5], e.g. data from the National Registry of Reimbursed Health Services (NRHZS), the Czech National Cancer Registry and the National Registry of Reproductive Health are used. The procedure and conditions for the management of and access to these data are comprehensively regulated in Sections 70-78 of Act No. 372/2011 Coll. (on health services and conditions of their provision), as amended, and its implementing regulations, in particular Decree No. 116/2012 Coll. of the Ministry of Health on the transfer of data to the National Health Information System, or Decree No. 373/2016 Coll. on the transfer of data to the National Health Information System (effective from 1 January 2017). Data stored in the NHIS registries and other data sources are regularly updated at least once a year. The core indicators calculated from these data are updated for the previous year in the 3rd quarter of the current year (i.e. the data for 2022 is completed in the 3rd quarter of 2023).

Data on the demographics of the Czech population and clinical data from cancer screening providers collected by professional medical associations are equally important. The combination of all components creates a comprehensive information system that enables a comprehensive evaluation of all aspects of the performance, quality and cost of the screening process.

The National Screening Centre has built and is continuously developing the Data Portal[6], which is designed for integrated evaluation of screening programmes and provides access to analyses of the performance and quality of screening programmes down to the level of regional units. In doing so, it uses the two NHIS registries listed below: (i) NRHZS containing health insurance data on inpatients and outpatients, including complete data on reported

[5] https://www.uzis.cz/index-en.php?pg=nhis
[6] https://nsc.uzis.cz/data/

diagnoses, procedures and treatments, (ii) the Death Certificate as the primary source of information on each death. (It is promptly completed after the examination of the deceased by the examining physician, who, in addition to basic sociodemographic characteristics, also records the sequence of causes leading to death, since 1994 coded using ICD-10.) An integral part of the demographic population data is processed by the Czech Statistical Office[7], which contains the main demographic characteristics of the Czech population, especially the total population, detailed age structure, life expectancy characteristics and, for example, projections of the age structure of the Czech population. These input data were used to build a comprehensive information service that provides clear and correctly interpreted outputs and conclusions in accordance with the target group. The primary target groups include two types of users: (i) health service providers and commissioners who are at some stage involved in existing screening programmes (this target group includes county and municipal authorities staff involved in health services), (ii) staff in health promotion and disease prevention services who are involved in the establishment and coordination of screening programmes. Secondary target groups may include the general public (not only laypeople but also healthcare workers or journalists), with efforts to continuously raise awareness of screening programmes and develop information and health literacy. Therefore, selected types of outputs are presented in an educational and popularising form within the National Health Information Portal (NZIP)[8], which provides guaranteed and comprehensible information from the Czech healthcare sector.[9]

RESULTS

BUSINESS UNDERSTANDING

A thorough understanding of the issues, methodological procedures and knowledge of terminology is the basis for a correct and complete presentation of outputs over the selected data area. In the case of early detection programmes for selected diseases, these are the different phases of the life cycle of screening programmes (Figure 1). The different phases describe the entire process, from the assessment of the intention to develop a new screening programme, through the implementation and validation of a pilot project, to the global implementation of a national screening programme.

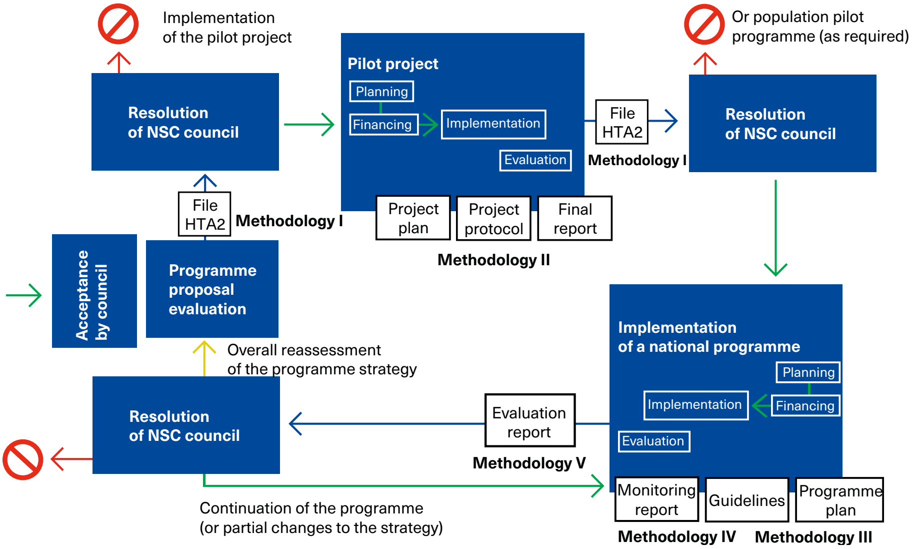
[7] https://www.czso.cz/csu/czso/home

[8] https://www.nzip.cz/

[9] This chapter describes the status of the portal as of 31 December 2022.

In addition to preparation and implementation, it is also necessary to understand how the quality of the screening programmes is continuously monitored and improved through ongoing monitoring of defined quality indicators. This control is carried out using so-called quality indicators, which are defined on the basis of European recommendations [1] for screening programmes and implemented in the Czech context.

It is also necessary to establish how to evaluate the impact of technologies in health screening and how to use them in health policy decision-making.



Implementation of the pilot project
Or population pilot programme (as required)
Pilot project
Planning
Financing
Implementation
Evaluation
Resolution of NSC council
File HTA2
Methodology I
Resolution of NSC council
Project plan
Project protocol
Final report
Methodology II
File HTA2
Methodology I
Acceptance by council
Programme proposal evaluation
Implementation of a national programme
Planning
Implementation
Financing
Evaluation
Overall reassessment of the programme strategy
Resolution of NSC council
Evaluation report
Methodology V
Monitoring report
Guidelines
Programme plan
Continuation of the programme (or partial changes to the strategy)
Methodology IV
Methodology III

Figure 1: Diagram showing the life cycle of a screening programme

DATA UNDERSTANDING

Before creating inertial visualizations, it is imperative to understand the issues that the data describe. Given our long experience with three national organised cancer screening programmes (breast /colorectal/cervical cancer screening programme) in the Czech Republic, we started with these. The first programme whose data was processed was the nationwide breast cancer screening (C50), which was officially launched in the Czech Republic in 2002. The aim of this programme is to regularly examine women without any signs of disease in order to reveal developing breast cancer at the earliest possible stage. The programme is aimed at women over 45 years of age, for whom screening is covered, at regular two-year intervals.

Another programme was colorectal cancer screening (C18–C20), which focuses on colorectal cancer prevention. This programme has been running in the Czech Republic since 2000, focusing on people over 50 years of age. For these patients, faecal occult blood test (FOBT) at two-year intervals (one-year interval for the highest risk group in the age group 50–54 years) and screening colonoscopy at ten-year intervals are covered[10].

The last screening programme was cervical screening (C53), which focuses on cervical cancer prevention. This programme was launched in the Czech Republic in 2008, targeting all adult women. The screening examinations (conventional cytology) are reimbursed at one-year intervals[11], women can also undergo high risk HPV test at ages 35 and 45.

DATA PREPARATION

Correct and error-free data preparation was essential in preparing specific open datasets. As mentioned above, the data sources are the central registries within the NHIS. The main registry is the NRHZS, which contains data from health insurance companies on both inpatients and outpatients, including complete data on reported diagnoses, procedures and treatments. Categories of interest were always selected from this registries according to their identification number, and these had to be mapped to the population data according to the Czech Statistical Office. This was followed by aggregation by age groups, sex groups, geographical units and years. Here, several obstacles had to be overcome; for example, the effect of inconsistency in the time series of demographic data between 2020 and 2021 due to the 2021 census. This problem was subsequently addressed by adjusting the input values of the calculations. For 2020, the population as of 1 July 2020 was used, and for 2021, the population as of 31 December 2020 was subsequently used. This was done to prevent artificial break in time series of coverage. However, a commentary has been added to the data description urging caution in interpreting changes between 2020 and 2021. It was also necessary to properly validate the NRHZS data - records without the insured's sex or district of residence were removed. In the case of the indicator coverage by population, it was also necessary to remove records that related to patients who died before the year for which coverage was assessed (e.g. if coverage is assessed for the year 2021, patients who died before and including 2021 were not included in the calculation).

[10] https://nsc.uzis.cz/data/index.php?pg=screening-rakoviny-kolorekta-c18-c20

[11] https://nsc.uzis.cz/data/index.php?pg=screening-rakoviny-delozniho-hrdla-c53

MODELLING

OPEN DATASETS

The individual aggregated data sets were prepared according to open data standards[12], which evaluate the effectiveness of individual screening programmes in terms of the coverage rate of the target population by a given examination (coverage by population). The outputs in the form of three open datasets are available as open data in the Ministry of Health catalogue[13]. In total, three datasets assess individual screening programmes.

Screening mammography: The coverage by population shows the proportion of women in the target population who have had screening mammography in the preceding 2 years. Data are always related to a specific geographic area and year.

- **Year**
- **Sex** – only women in this case, due to the focus of the screening programme.
- **Age** – five-year age categories, with the youngest age category being 45–49 years, due to the programme's focus.
- **District code and name** in the CZ-LAU1 format.
- **Region code and name** in the CZ-NUTS format.
- **Number of examinations** performed in a specific year, age group and geographic area. This is the total number of persons examined (in a two-year interval) from NRHZS data[14]. An insured person is taken to be an examined person if a procedure coded as 89178 (screening mammography in follow-up care) or 89,223 (screening mammography) is reported.
- **Population** – the total number of persons in the population of the Czech Republic according to the CZSO data[15]. A specific value for a given combination of variables is always listed.

Screening colonoscopy: The **coverage by population** shows the proportion of persons in the target population who underwent screening colonoscopy in the monitored ten-year interval or screening FOBT at two-year intervals (one-year interval for age group 50–54 years). Data are always related to a specific geographic area and year.

[12] https://opendata.gov.cz/standardy:technicke-standardy-pro-datove-sady-na-stupni-3

[13] https://data.mzcr.cz

[14] https://www.uzis.cz/index.php?pg=registry-sber-dat--narodni-registr-hrazenych-zdravotnich-sluzeb

[15] https://www.czso.cz/csu/czso/vysledky-scitani-2021-otevrena-data

- **Year**
- **Sex**
- **Age** – five-year age categories, with the youngest age category being 50–54 years, due to the focus of the programme.
- **District code and name** in the CZ-LAU1 format.
- **Region code and name** in the CZ-NUTS format.
- **Number of examinations** performed in a specific year, age group and geographic area. This is the total number of persons examined (in a proper interval) from the NRHZS data[16]. An insured person is taken to be an examined person if a procedure coded as 15105, 15107 (screening colonoscopy) or 15118, 15119, 15120, 15121 (screening FOBT) is reported.
- **Population** – the total number of persons in the population of the Czech Republic according to CZSO data[17]. A specific value for a given combination of variables is always listed.

Cervical screening: The **coverage by population** shows the proportion of women in the target population who underwent cervico-vaginal screening cytology in the monitored one-year interval. Data are always related to a specific geographic area and year.

- **Year**
- **Sex** – only women in this case, due to the focus of the screening programme.
- **Age** – five-year age categories, with the youngest age category being 15–19 years, due to the programme's focus.
- **District code and name** in the CZ-LAU1 format.
- **Region code and nam**e in the CZ-NUTS format.
- **Number of examinations** performed in a specific year, age group and geographic area. This is the total number of persons examined (at one-year intervals) from NRHZS data[18]. An insured person is taken to be an examined person if a procedure coded as 95198 or 95199 (cervical cancer screening) is reported.

[16] https://www.uzis.cz/index.php?pg=registry-sber-dat--narodni-registr-hrazenych-zdravotnich-sluzeb

[17] https://www.czso.cz/csu/czso/vysledky-scitani-2021-otevrena-data

[18] https://www.uzis.cz/index.php?pg=registry-sber-dat--narodni-registr-hrazenych-zdravotnich-sluzeb

— **Population** – the total number of persons in the population of the Czech Republic according to the CZSO data[19]. A specific value for a given combination of variables is always listed.

INTERACTIVE BROWSERS

Based on the above-described and published open datasets, business intelligence visualisations (interactive dashboards) were prepared to present screening data to experts and the informed public in a way that is easy to understand. The reports of all screening programs are based on a common, and therefore uniform, matrix screen layout, where sets of filters are arranged in a user-friendly format to refine the desired report. Specific graphical outputs, as well as tabular and map visualisations aggregate screening coverage results by geographic units (regions and districts), age, sex groups and years.

Each of the dashboards is built over four basic views of the data:

1. The main indicator (one of the main conclusions from the dataset) that characterises the overall population coverage of a given screening in the selected main target population (e.g., the colorectal cancer screening coverage of the target population in the two-year interval for 2021 is 26.9%). This information is used for quick comparison of this performance indicator and its potential comparison with established assumptions, including international ones20.

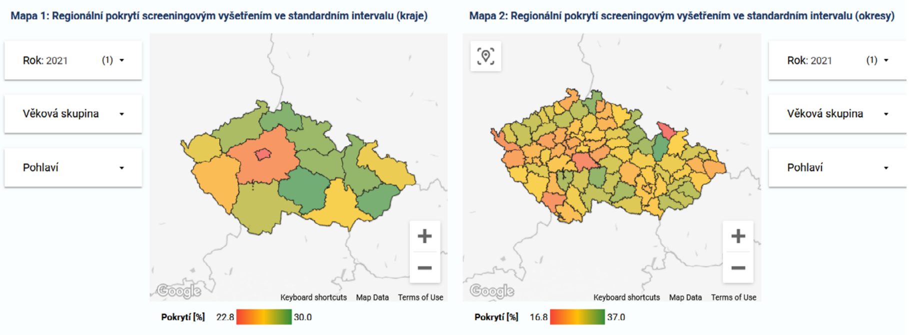
2. Bar charts are another form of visualisation that help the user quickly navigate a given programme's age and time trends. They are linked to individual filters that allow the user to view only the data in his/her area of interest.

3. Map visualisations are a third way of plotting the programme's success from a geographic perspective. The programme's success in each geographical unit is coded using a colour scale (Figure 2).

4. The table view is another data display that is one of the preferred forms of data presentation, primarily because of its simplicity.

[19] https://www.czso.cz/csu/czso/vysledky-scitani-2021-otevrena-data

[20] https://healthcare-quality.jrc.ec.europa.eu/ecibc



Mapa 1: Regionální pokrytí screeningovým vyšetřením ve standardním intervalu (kraje)

Rok: 2021 (1)

Mapa 2: Regionální pokrytí screeningovým vyšetřením ve standardním intervalu (okresy)

Rok: 2021 (1)

Figure 2: Example of map visualisations on the NSC Data Portal (available only in the Czech language)[21]

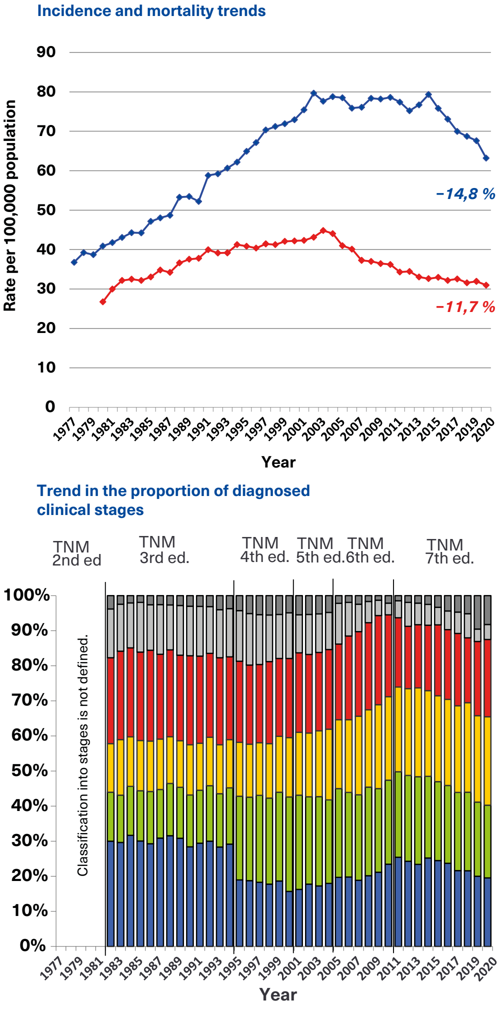
In addition to the interactive online visualisation, other static outputs (e.g. summary presentations published in PDF format or as an online presentation) are prepared from primary and open data to inform about the indicators in individual screening programmes. These reports show additional results achieved by each programme, focusing on key indicators based on geographic and demographic conditions (Figure 3).

EVALUATION

Validation is an integral and essential part of the life cycle of every deliverable. The dataset design and creation must ensure machine readability according to clearly defined open standards. The first step of validation was to check adherence to these standards. The resulting datasets were published in CSV format, so it was necessary to set up and check the correct encoding of the dataset, setting standardised separators and unit format. Once the structure checks were completed, the content checks followed. This was first performed during the creation of each dataset by the analysis team, followed by a check by the technical staff responsible for publication. This involved the preparation of a few simple summaries and contingency tables from the input data, looking for inconsistencies, unexpected results or non-standard values. If any unexpectedness was found, the set was returned for checking or completion. In creating the open data, another important element was the preparation of a metadata record that describes the descriptive attributes of the set and introduces the topic being processed. The metadata check is two-level: (i) content-wise, the description of the values and dataset is reviewed by the sponsor with emphasis on clarity and understandability; (ii) structural check verifies machine readability and consistency with the

[21] https://nsc.uzis.cz/data/index.php?pg=screening-rakoviny-kolorekta-c18-c20--online-analyza

primary input set. A similarly thorough check is performed when preparing interactive visualisations or static reports over the available data. Errors may arise, for example, during data transfer or by additional editing in the content or visual settings of the report.


Incidence and mortality trends

Trend in the proportion of diagnosed clinical stages

Source: Czech National Cancer Registry, IHIS CR – incidence; Czech Statistical Office – mortality

A marked decrease in the incidence and mortality of colorectal cancer has been observed in the long term. Colorectal cancer screening has a limited impact on the early detection of invasive cancers.

Figure 3: Static output describing cancer burden for C18–C20

DEPLOYMENT

The resulting datasets are published in the local open data catalogue of the Ministry of Health, from where they are synchronised to the national open data catalogue. More detailed information on these datasets is always available in the section of the screening programme on the NSC portal. The interactive reports, prepared using the business intelligence tool Looker Studio, are also embedded in the relevant sections of the NSC portal.

DISCUSSION

This data portal was developed as a tool to assess the quality, effectiveness, or equity of screening programmes. It provides access to analyses of the performance and quality of screening programmes down to the level of regional units. An integral part of the portal is publicly available datasets that allow the calculation (assessment) of key indicators of the quality of screening programmes according to basic socio-demographic characteristics - i.e., it is publicly available information that aims to make this assessment transparent. At the same time, datasets allow the creation of individual analyses, for example, for a specific assessment of the situation over time after the introduction of a health intervention. Preparation phase (programming and design) of the NSC portal was necessary to have a product that is easy to manage and update.

The future development plan includes (i) developing additional sections (expanding the portfolio of diagnoses that are displayed on the portal), (ii) expanding the datasets - not only simple ones for calculating quality indicators, but also complex ones for more complex analyses, and (iii) developing online visualisations for additional datasets.

In the future, the portal should serve as a comprehensive data source on the results of prevention and early detection programmes managed by NSC, and its content should be validated and reviewed by the NSC Council and the relevant Steering Committees for each screening programme underway or in preparation.

EVALUATION OF THE AIMS OF THE CHAPTER

- To understand cancer screening issues focusing on breast, cervical, and colorectal cancer.
  - The National Screening Centre, as a robust umbrella system in close cooperation with experts, has deeply understood and adopted selected cancer screening programmes and all related agendas.
- To design pilot datasets describing coverage of the target population by selected screening examinations.
  - Open datasets containing basic descriptive characteristics for selected screening programmes (breast, cervical, and colorectal) have been designed and successfully published. These outputs have set the way for the further publication of other results according to the priorities and needs of the National Screening Centre.
- To implement an interactive visualisation of published data.
  - The close collaboration of the individual teams has also resulted in an interactive data viewer built on top of the published datasets using business intelligence tools. This approach has proven to be functional and practical for other National Health Information System components.

LESSONS LEARNED

The development of a unified system of screening programmes helped to implement a comprehensive information service over available data from the National Health Information System. A set of static and interactive outputs was created over the pilot areas, now available on a separate dedicated portal. Through close collaboration across teams with connections to representatives from medical societies, an easy-to-understand data portal was developed as a valid and objective tool for assessing the quality, effectiveness and even equity of screening programmes.

The data has become a very important tool for evaluating and modifying existing programmes and for assessing the meaningfulness of introducing new programmes in an effort to maximise their impact.

REFERENCES

[1] European Commission Initiatives on Breast and Colorectal Cancer. Improving quality of care and reducing inequality in Europe [Internet]. [cited 8 Aug 2023]. Available from: https://healthcare-quality.jrc.ec.europa.eu/.

[2] Ngo O, Bučková B, Hejduk K, et al. Assessment of the impact of technologies in health screening. I. Methodological document of the National Screening Centre. Prague: Institute of Health Information and Statistics of the Czech Republic; 2020. Available from: https://nsc.uzis.cz/res/file/vystupy/hodnoceni-dopadu-technologii-ve-zdravotnim-screeningu.pdf.

SECTION C

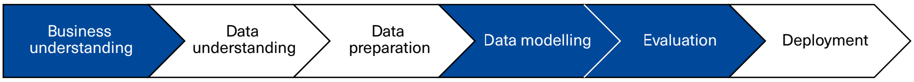
10

DATA AND ANALYTICAL BASIS FOR A MODERN MENTAL HEALTH CARE SYSTEM IN THE CZECH REPUBLIC

Martin Komenda, Jitka Soukupová, Hana Melicharová, Jiří Jarkovský, Ondřej Šanca, Vladimír Bartůněk, Petr Benáček, Lenka Šnajdrová, Ladislav Dušek

The case study was supported from the following grant project: Data and analytical basis for a modern mental health care system in the Czech Republic, ESF, CZ.03.2.63/0.0/0.0/15_039/0007755.

CRISP-DM CRUCIAL PHASES



Business understanding → Data understanding → Data preparation → Data modelling → Evaluation → Deployment

GENERAL INFORMATION

Year	2018–2022
Keywords	Mental health, psychiatric care, open data, interactive reporting
Research question	How to build a comprehensive information reporting tool supporting a broad mental health reform in the Czech Republic.
Type of result	Interactive visualisation
Level of data processing	Advanced analyses, Open datasets

DATA TO DOWNLOAD

INTRODUCTION

Mental health centres – as an intermediate link between outpatient and inpatient care – provide a wide range of health and social services to their clients, including long-term psychiatric care, early diagnosis and crisis support for acute mental health problems, social rehabilitation and social support and counselling. The primary motivation is to prevent or reduce hospitalisations and to help reintegrate people who were hospitalised for a long time back into their own community. The main target group is patients who have been diagnosed with a serious mental disorder, i.e. a serious disorder from the diagnostic categories F20–F29 (schizophrenia, schizotypal disorders, delusional disorders) and F30–F39 (affective disorders), or F42 (obsessive-compulsive disorder) and F60–F69 (personality and behavioural disorders in adults), who have been struggling with their disorder for a long time and whose functional impairment is relatively high. The services of the specialised centres are also intended for people under court-ordered protective treatment, which is provided in the form of specialised outpatient psychiatric care.

The project entitled "Data and analytical basis for a modern mental health care system in the Czech Republic" focused on multi-source evaluation of mental health care, analysis of data sources, validation, and definition of a national set of indicators. The implementation of these steps led to comprehensive optimisation and support for mental health care reform. The key activities of the project were divided into the following agendas: (i) Economic and operational agenda: in particular, the full-time capacities of all categories of health professionals, infrastructure and casemix of providers, the number of beds and their utilisation, management, time characteristics of care provided etc. (ii) Agenda of catchment areas of healthcare facilities in terms of assessment of access to care, migration of patients, continuity of outpatient and inpatient care, continuity of acute and follow-up inpatient care etc. (iii) Agenda of organisation, scope and needs of different types of care. Quantification of the need for care and knowledge of the real clinical burden, as well as knowledge of the risk stratification of patients requiring a particular type of care, is essential for assessing the availability and performance of psychiatric care. (iv) Agenda for identifying the future needs of different types of care in terms of assessing population trends, surveys focusing on the views and needs of management of providers, founders, regional and local authorities. An integral part of this was the implementation of representative statistical surveys, focusing on all providers of all forms of mental health care. The conceptual plans and needs of providers, regional and local authorities were surveyed, and care and needs were assessed from the point of view not only of health and social workers but also of patients' families, always on a representative sample. In addition, the project designed predictive models that mapped the capacity and needs of modern psychiatric care with respect to (i) regional epidemiological and clinical burden in different psychiatric diagnostic groups, and the resulting needs for the scope and type of psychiatric care, (ii) the staff capacity needed to provide different forms of psychiatric care, (iii) the number of different types of beds. Building a centrally managed data collection in the form of a registry and conducting observational surveys focusing on essential parameters of care was the basis for a completely new information system for psychiatric care provided by multidisciplinary teams. This registry mainly covers the following chapters: (i) characteristics of the health and social service providers involved, parameterisation of achievable performance and capacity, (ii) characteristics of patients admitted and classification according to risk and severity of their condition, (iii) development over time of patients' condition and assessment of intermediate outcomes of care, (iv) overall assessment of outcomes of care, including outcomes reported and assessed by patients themselves. The overarching platform is the Czech National Portal of Mental Health Care, which contains all of the project's supporting outcomes. This case study focuses on the area of psychiatric care primarily in terms of data processing, analysis, and visualisation.

AIMS

- To understand what agendas are necessary for the implementation of psychiatric care reform in the Czech Republic.
- To build a comprehensive information service over data on the state of psychiatric care, taking into account the target group (appropriate type and format of output over the available data).

METHODS

In order to systematically support the reform of mental health care services through information and communication technologies, it is necessary to understand in detail how this system of care works. Sources of reference data, including the methodological background for processing, interpreting, and providing the information service itself, are of key importance. By building a data base and analytical reporting that together map mental health care in the Czech Republic, a missing information tool has been created that allows (i) contracting authorities and providers of mental health care to map and evaluate this care, (ii) citizens who consume this care to search for information about the availability and parameters of care in their region or area, and (iii) family members and caregivers of the mentally ill to navigate the system of health care providers in their region or area.

To meet the set objective of the reform, the Czech National Portal of Mental Health Care was designed, developed and launched. Within this portal, maximally accessible data are published for different roles of visitors, mapping the availability and quality of services provided; in addition, comprehensive data on the distribution of services in all regions of the country are available. It is a robust platform offering a wide range of services, from traditional online news to interactive maps and interactive web reports to elaborate electronic publications and presentations. On a regional level, the portal offers a "logistical" map of accessibility to different mental health care services. For mental health care clients, the portal enables them to find adequate information about the scope and quality of services available near their home, thereby supporting informed decision-making about the consumption of these services. For contracting authorities and providers of mental health care, the portal makes available information on the specific current availability and utilisation of services offered, in addition to monitoring the coherence of services in their own region. For care providers, the portal is also a source of valid information on the type and quality of care in the region, on possible patient migration related to this care, and last but not least, it is also a space for publishing relevant information and news, as well as a means of communication and self-assessment.

Open data and interactive visualisation are built on data from the National Health Information System, which was created by an amendment to the Health Services Act, specifically Act No. 147/2016 Coll., amending Act No. 372/2011 Coll., on Health Services and Conditions of Their Provision (Act on Health Services), as subsequently amended.

RESULTS

BUSINESS UNDERSTANDING

Valid data sources are crucial for the design and preparation of descriptive characteristics mapping psychiatric care in the Czech Republic. If we want to monitor the state of quality of care, trends, strengths, weaknesses and future needs in the long term, we need nationally and regionally available reference data. The comprehensive integration of multi-source data, which was designed and implemented as part of a dedicated project, allowed not only monitoring of the current state of care, but also a critical assessment of weaknesses in its availability, logistics and capacity. This new form of access to comprehensive data has significantly helped to target mental health care in different regions in a more targeted way to minimise inequalities in access and to maximise the effect for the end consumer of these services. Thanks to the methodological and technical background that has been built up, it is now possible to process, analyse and publish various forms of reports and to develop an information service in the field of mental health care. This phase was characterised by the link between the absolute need for quality improvement of mental health care, the project challenge and all the professional teams involved. The overall correct understanding of the issue, as defined in the first phase of the CRISP-DM model, was based primarily on interdisciplinary collaboration and the attempt to address a conceptually straightforward problem through an across-the-board reform.

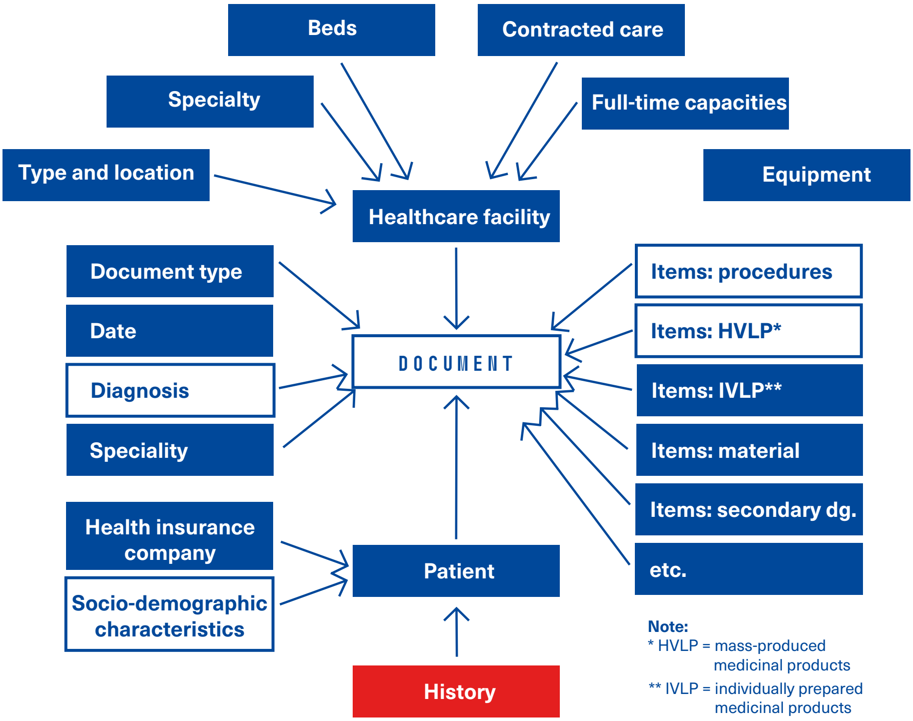
DATA UNDERSTANDING

This phase always introduces a thorough understanding of all data sources and input files that form the basis of all subsequent steps. In the case of this study, we are talking about two databases that are mutually linked to create the final outputs.

1. The National Registry of Reimbursed Health Services is an integral – and the most comprehensive in terms of data volume – part of the National Health Information System in the Czech Republic. This registry collects data reported by all health service providers to health insurance companies. The data collection in the registry is set up in such a way that it is not burdensome for health service providers. It makes use of the data collection already carried out by health insurance companies, which then submit the data to the registry. This saves a significant burden on providers, with only 7 health insurance companies contributing to the central database instead of more than 30,000 health service providers. Thus, the individual health insurance companies are the providers of data to the registry, transferring the data directly from their information system to the registry at three-monthly intervals, including a record of all exports made. Data collection is of course carried out in accordance with the Act on Health Services and other legislation. The complexity of the data from the National Registry of Reimbursed Health Services is documented in the diagram below (Figure 1).

2. Demographic data from the Czech Statistical Office, which are needed primarily for the design and implementation of advanced visualisation in psychiatry. These annual data describe the population in the Czech Republic overall and in smaller regional units, such as regions and towns, together with the population age structure. These demographic data are used primarily for epidemiological recalculations and comparisons of data across territories

governed by local authorities; the median population is used for analyses. The data are updated annually, always in the middle of the following year.



Note:
* HVLP = mass-produced medicinal products
** IVLP = individually prepared medicinal products

Figure 1: Data structure of the National Registry of Reimbursed Health Services

DATA PREPARATION

The retrieval, transformation and validation of the data submitted by health insurance companies, as well as the subsequent integration with data from other systems (for example, publicly available information on demography in the Czech Republic), takes place in a separate and secure internal infrastructure, which is primarily designed to process the primary input data and to make it available for subsequent analytical activities. This is also where the process of translating the transmitted identifiers of the insured persons into agenda identifiers of natural persons takes place. With regard to the necessary security, the dedicated servers for the management of primary data are accessible only to the authorised administrators of the National Registry of Reimbursed Health Services. An integral part of this process is a validation procedure aimed at verifying that the quality requirements of the transmitted data have been met so that they can be loaded into the data warehouse. The validation process uses a logical model for the distribution of controls (Figure 2) in the form of a five-level pyramid hierarchy with a direct link to specific stages of the data processing process. These phases are denoted by code abbreviations K0–K4, which make it easier to work with the control mechanisms and to navigate the statistical output reports.

- K0 are basic file structure checks – file names, column counts and names, and data types.
- K1 are database integrity checks – value ranges, primary and foreign keys.
- K2 are basic semantic checks at the level of a single data import – e.g., whether a patient has not been simultaneously admitted to two hospitals, whether a physician has not a weekly workload higher than 168 hours and the validity of values against data in the codebooks.
- K3 are checks comparing one import against other data – patient numbers against previous periods for the same health insurance company, the pattern of health services over the same period for other health insurance companies, and comparisons of expenditure against the annual reports of individual health insurance companies.
- K4 are content controls over the whole data warehouse – numbers of hospital admissions for all health insurance companies against NRHOSP, numbers of patients against CZSO demographic data.

Due to the large volume and variety of data, the validation process is carried out separately for each data processing stage and is given a high priority. The validation mechanism within each stage of data processing includes the following basic steps: (i) design and description of the controls, separately for each data sentence, (ii) setting up the control mechanism, (iii) execution of the control, i.e., accurate identification of the erroneous values, including assignment of error codes, (iv) recording the necessary information about the error found, (v) accompanying control mechanisms.

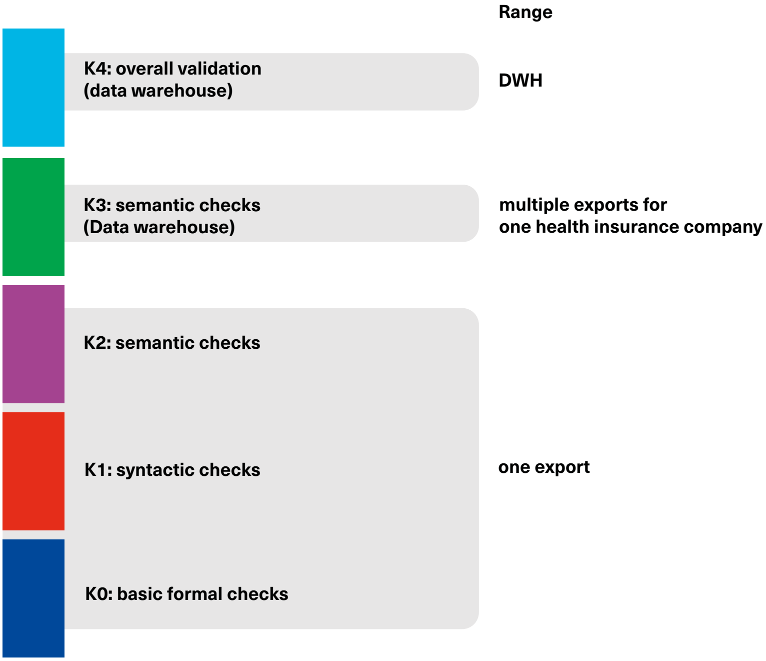

Range
K4: overall validation (data warehouse)
DWH
K3: semantic checks (Data warehouse)
multiple exports for one health insurance company
K2: semantic checks
K1: syntactic checks
one export
K0: basic formal checks

Figure 2: Model of control mechanisms

The output of the validation process is the identification of non-standard and erroneous values. Depending on the type of check performed, each found erroneous value is assigned a three-digit error code, which categorises errors in three levels of meaning and carries the basic information on which the correction mechanism is based.

The complex process of preparing data sources (Figure 3) consists of the following stages: (i) The preparation stage, which is a key part of the data preparation process, is the definition of the required patient groups; this stage ends with the preparation of the SQL scripts for data generation. (ii) Generation of data outputs, which takes place after the previous year's data has been closed (usually in the middle of the following year) using the SQL scripts prepared in the previous stage. (iii) Data transfer via a dedicated server to ensure maximum security; the purpose of this stage is to separate the individual databases and the web server. Due to changes in reporting and the introduction of new codes for diagnoses, procedures, drugs and medical supplies, the definitions of patient groups and the related SQL scripts are revised before the annual export.

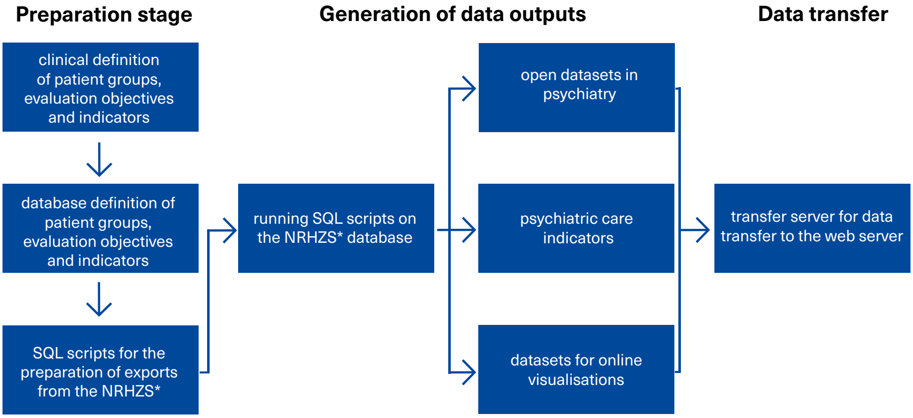

Preparation stage
Generation of data outputs
Data transfer
clinical definition of patient groups, evaluation objectives and indicators
database definition of patient groups, evaluation objectives and indicators
SQL scripts for the preparation of exports from the NRHZS*
running SQL scripts on the NRHZS* database
open datasets in psychiatry
psychiatric care indicators
datasets for online visualisations
transfer server for data transfer to the web server

Figure 3: Process of preparing data sources (* NRHZS = National Registry of Reimbursed Health Service)

MODELLING

OPEN DATA SETS

The structured data sets created in the previous stage have been converted into an open data format. Again, the basic requirements for open data output were met in terms of (i) ensuring factual content and correctness, (ii) choosing the appropriate level of openness at which the data set will be published, (iii) preparing a data schema specifying the structure in which the data set will be published (machine-readable JSON format), (iv) creating and publishing a record of the data set in the open data catalogue. After completion of all mandatory descriptive characteristics and another iteration of validations, each of the 11 data sets was published in the aforementioned catalogue[1]. The data sets below provide aggregated data on selected indicators of psychiatric care (incidence and prevalence, hospitalisation of people with mental health conditions, coverage of outpatient psychiatric services, antipsychotic medications for patients with dementia) based on patient numbers, stratified by selected diagnoses and year of recording.

1. Psychiatric care: Acute care
   — The data set provides aggregated data on patients admitted to acute care beds in psychiatric wards of hospitals.

[1] https://data.mzcr.cz/

2. Psychiatric care: Coverage by follow-up care

3. The data set provides aggregated data on coverage by follow-up care. It contains the numbers of patients who visited an outpatient psychiatrist after hospitalisation.

4. Psychiatric care: Long-term care

5. The data set provides aggregate data on long-term psychiatric inpatients. The data set is divided by age categories, i.e. into children's and adult's care.

6. Psychiatric care: Alzheimer's disease and unspecified dementias

7. The data set provides aggregated data on the numbers of patients with Alzheimer's disease and unspecified dementias. Patients under 50 years old are excluded from the dataset.

8. Psychiatric care: Outpatient care

9. The data set provides aggregated data on psychiatric patients who were treated on an outpatient basis.

10. Psychiatric care: Suicide attempts

11. The data set provides aggregated data on the number of suicide attempts, stratified by sex.

12. Psychiatric care: Rehospitalisations of people with mental disorders

13. The data set provides aggregated data on the number of rehospitalisations. The data are stratified into intervals according to the number of days since the first hospitalisation; the percentage of total hospitalisations is also available.

14. Psychiatric care: Mortality of people with mental disorders

15. This indicator tracks the mortality rate of persons with a history of hospitalisation for mental illness in the past five years, standardised to the general population. The standardised mortality rate is defined as the ratio of the actual number of deaths in a given year, in the population of persons aged 15 to 74 years with a history of hospitalisation for mental illness, to the expected

number of deaths in this population (i.e. the number of deaths in the general population if the age and sex distribution in the general population were the same as in the study population).

16. Psychiatric care: Suicidality of people with mental disorders

17. The data set provides aggregated data on suicidality of people with mental disorders. Suicide numbers are divided into intervals according to the number of days since the last hospitalisation.

18. Psychiatric care: Use of psychopharmaceuticals

19. The data set provides aggregated data on the use of psychopharmaceuticals, i.e. the percentage of patients who were prescribed medication in the total number of patients.

20. Psychiatric care: Incidence, prevalence, hospitalisation and outpatient care

21. The data set provides aggregated data on selected indicators of psychiatric care (incidence and prevalence, hospitalisation of people with mental health conditions, coverage of outpatient psychiatric services, antipsychotic medications for patients with dementia) based on patient numbers, stratified by selected diagnoses and year of recording.

INTERACTIVE DATA BROWSER

Interactive visualisations have been built over these published data sets, which are freely available to the general public, using business intelligence tools. The primary goal is to provide users with a platform within the Czech National Portal of Mental Health Care[2] where they can work with the available data and configure the reports themselves in an understandable yet dynamic way. Through simple filtering (e.g. year, diagnosis, sex or age category) and setting the view of the data (e.g. absolute numbers, per 100,000 population or relative structure), it is possible to very quickly obtain a very specific report according to the user's requirement. Figure 4 shows an interactive data viewer where the number of patients with psychotic disorders in 2021 is displayed in relative view concerning the population in the regions of the Czech Republic. It is straightforward to change the year, specify the diagnosis and select the gender or age category of patients.

[2] http://psychiatrie.uzis.cz

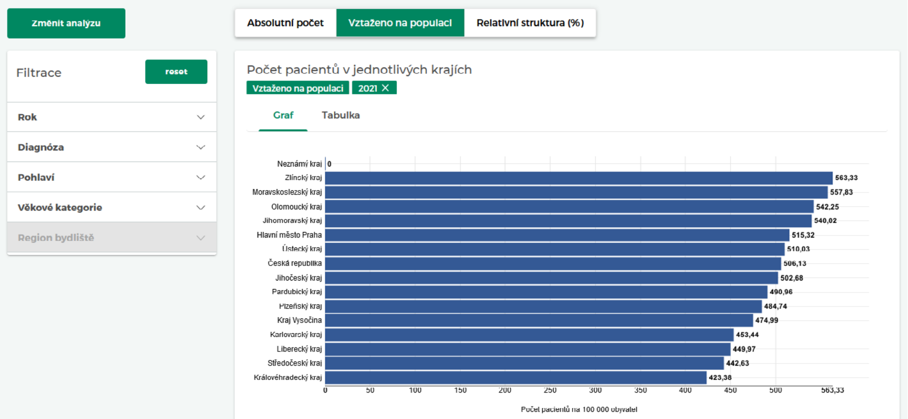


Figure 4: Psychotic disorders – frequency by region (only available in Czech)

Beyond the portal, an illustrative representation of the key information and conclusions that emerge from the published data sets has been developed in the form of infographics. Considering the target user group of lay and informed general public, this interactive visualisation has been published on the National Health Information Portal[3]. This is another in a series of interactive overviews as an example of innovation in the publication of psychiatric yearbooks, which are regularly published in static PDF format. The browser[4] provides alternative views and information on top of the yearbook data in a different format that is more interesting and accessible to a selected part of the target population. Figure 5 shows an overview of psychiatric care in an infographic, showing aggregated data on treated psychiatric patients by essential breakdowns.

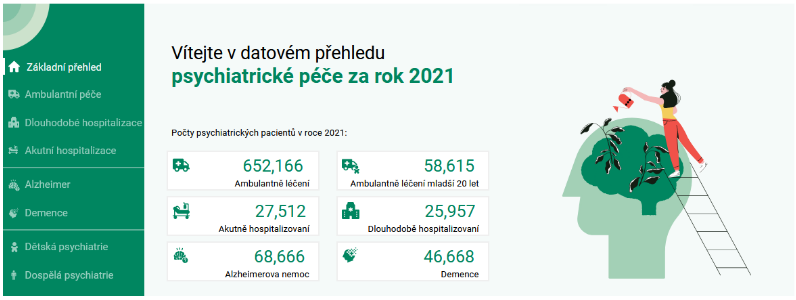
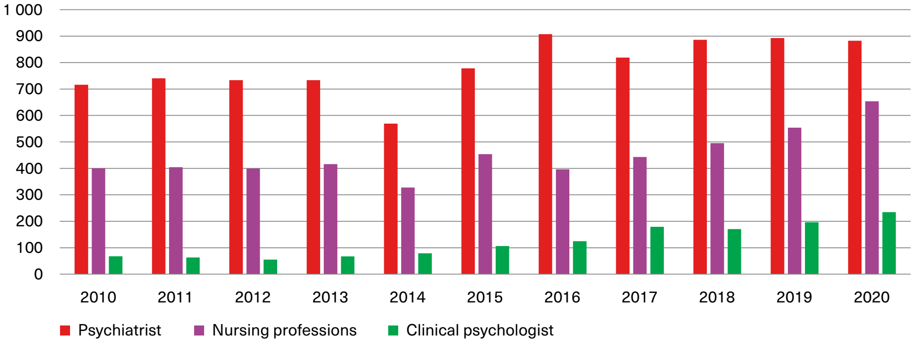


Figure 5: Homepage of the interactive browser of the Psychiatric Yearbook 2021 (available only in Czech)

[3] https://www.nzip.cz/infografika-dusevni-onemocneni-v-cesku
[4] https://psychiatrie.uzis.cz/cs/rocenka/

STATIC ANALYTICAL REPORTS

The next in a series of outputs over the available data are comprehensive analytical reports (Figure 6), which aim to provide a detailed basis for subsequent decision-making over valid and validated outputs. At the time of the publication of this book, systematic reform of psychiatric care in the Czech Republic was still underway. Therefore, the clear motivation for the above-mentioned analytical reviews was to answer the question of whether the ongoing reform is recognisable or where the pitfalls and weaknesses of the existing psychiatric care are. Specific examples are overviews of the number of beds in psychiatric hospitals, the state of availability of acute inpatient care, the number of long-term inpatients in psychiatric hospitals, the number of staff in psychiatric outpatient clinics and many others. The aforementioned overviews of the mental health care system can be verified by any user in open data and using advanced visualisations.



Data source: National Registry of Reimbursed Health Services, 2010–2020

Figure 6: Average annual full-time equivalents in psychiatric outpatient clinics

EVALUATION

The validation of the prepared data sets – and the reports that are linked to them – took place at several levels. For the prepared data sets, the validation was divided into two steps. In the first step, the content was validated in close cooperation with the team of guarantors who are responsible for the accuracy of the prepared data set. A second check by an analyst who does not have such extensive know-how in the field is very appropriate. In this way, formal errors and errors in content that are not visible to the expert can be detected. There may be exceptions or anomalies in the data that the expert is aware of, but which must be pointed out and validly described when published for the general public. The second step is technical validation. All data sets are adapted to comply with international standards for open data. Subsequently, a metadata record is prepared and checked, which must also be correct in terms of content and technical processing. A check for data correctness was carried out during the creation of each visual report. The prepared reports and interactive tools must fully match the prepared input in terms of data. For reports based on non-public data, the output is checked against the primary database. This step is taken care of by an analyst or tester, depending on the complexity of the prepared reports. When creating reports and visualisations, the visual appearance is also an important aspect, such as the colour scheme or the overall navigation on the page. The result must be user-friendly, engaging and easy to understand, and at the same time fit in with the concept of the entire portal. Accessibility must also be respected, with individual views being distinguishable and legible for people with disabilities (e.g. colour vision deficiency etc.).

DEPLOYMENT

The deployment and publication of all outputs over open data was part of the implementation of the Czech National Portal of Mental Health Care[5]. Thanks to the basic division into two target groups – i.e. patients (caregivers) and professionals (doctors) – it is possible to freely publish static and dynamic reports, which together form a comprehensive background for the production of representative and reference data in the context of optimising psychiatric care and supporting the optimisation of the process of psychiatric care reform in the Czech Republic. This communication platform is continuously maintained technically and methodologically to ensure its smooth operation and thus the availability of information on the mental health care system.

DISCUSSION

The project "Data and analytical basis for a modern mental health care system in the Czech Republic" has significantly supported the reform of mental health care services through the new data, analytical and information tools that comprehensively map mental health care in the Czech Republic. The fulfilled objectives were closely linked to the strategic and conceptual documents of the Ministry of Health of the Czech Republic, in particular the documents of the Strategy for the Reform of Mental Health Care, Health 2020, the Action Plan for the Evaluation of Health Indicators of the Czech population, the National Action Plan against Suicide, the National Plan for Dementia (Alzheimer) Patients, and the Strategic Framework Health 2030. The analysis of the results of the mental

[5] https://psychiatrie.uzis.cz/

health centres' activities has helped health insurance companies to subsequently contract health care. In addition to the intended target groups (i.e. providers and contracting authorities of health services, public administration staff working on social, family or health issues, and those most at risk of exclusion and discrimination due to their health condition), the project has involved doctors and a representative sample of the Czech population through the surveys carried out. This continues to improve the mapping of the quality of care for the people with mental health conditions. Specifically, patient awareness is increasing, the progress of mental health care reform is being monitored and evaluated, and a data base and information tools have been built to comprehensively map mental health care. Sustainability is ensured and data are updated in direct link to the National Health Information Portal[6].

EVALUATION OF THE AIMS OF THE CHAPTER

- To understand what agendas are necessary for the implementation of psychiatric care reform in the Czech Republic.
  - Based on the identification and prioritisation of needs – and in cooperation with the expert society – it was possible to clearly define which tools and outputs are necessary for the systematic reform and optimisation of psychiatric care.
- To build a comprehensive information service over the data on the state of psychiatric care, taking into account the target group (appropriate type and format of output over the available data).
  - The analytical outputs were prepared with the target group in mind so that their supporting information was as effective, easy to understand and clear as possible. The combination of an interactive browser, infographics, open data and static reports managed to cover users at different levels of IT knowledge and skills.

LESSONS LEARNED

The mental health care reform in the Czech Republic is a beautiful example of close cooperation between representatives of an expert medical society and technical teams. A thorough understanding of complex issues of mental health care in the Czech Republic enabled the construction of a robust data basis, on which the entire information system was subsequently built. It consists of static and interactive presentations of available data, as different target groups prefer

[6] https://www.nzip.cz

different views. Rather time-consuming was not only the modelling phase but also the subsequent validation, which took place in several iterations. Open data are among the important outputs, making source data available for interested parties to perform their own analyses of mental health care.

This activity has set up an ever-repeatable process of processing, visualising and publishing data that describe the state of psychiatric care in the Czech Republic. These data thus play a key value in further decision-making on reforms and innovations in this area.

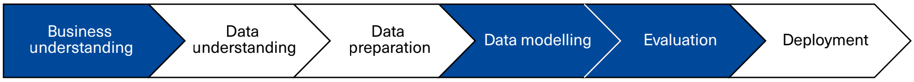
SECTION C

| 11 |

NATIONAL REGISTRY OF REPRODUCTIVE HEALTH: DATA INFRASTRUCTURE AND COMPREHENSIVE INFORMATION SERVICE

Jitka Jírová, Martin Komenda, Antonín Pařízek, Marian Kacerovský,
Vladimír Dvořák, Marek Ľubušký, Petr Janků, Jiří Jarkovský,
Lenka Šnajdrová, Ladislav Dušek

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2023–now
Keywords	Reproductive health, open data, neonatology, gynaecology and obstetric
Research question	How to design and implement a data infrastructure for comprehensive information communication of reproductive health topics.
Type of result	Interactive visualisation
Level of data processing	Advanced analyses, Open datasets

DATA TO DOWNLOAD

INTRODUCTION

The main theme of this chapter is a completely new information system that tracks all essential aspects of maternal and newborn care. This information system is being developed and updated under the full guarantee of professionals from clinical practice: specifically, it is a collaboration of the Czech Gynaecological and Obstetrical Society[1] and the Czech Society of Neonatology[2], while analytical and technical support is provided by the Institute of Health Information and Statistics of the Czech Republic. In terms of the quality of care for pregnant and parturient women, the Czech Republic has been long ranked among the best countries in Europe. This is evidenced, for example, by the very low numbers in key characteristics such as neonatal and maternal mortality. Indeed, the latest data from 2022 confirm the very above-average results of Czech perinatology, even in international comparison. Thus, we see very low maternal mortality (3.4 deaths per 100,000 live births) according to the international Unicef comparison[3], neonatal mortality (1.6 deaths in children aged 0–27 days per 1,000 live births) and early neonatal mortality (1.3 deaths in children aged 0–6 days in a given year per 1,000 live births), according to the international Eurostat comparison[4]. A number of process indicators focusing on treatment procedures and assessment of complications also show positive values. For example, the proportion of births terminated by caesarean section is below the EU average and has been slightly decreasing in the long term (according to Europeristat 2019 data[5]: Czech Republic: 24.5%, median EU countries: 26.0%).

Ensuring the availability of this care for all citizens of the Czech Republic and full coverage of costs from public health insurance are one of the important priorities of the Ministry of Health of the Czech Republic. In the area of control and evaluation of care, full digitisation of data collection and its gradual opening to the professional and lay public is planned. The National Registry of Reproductive Health, as an integral content part of the National Health Information Portal[6], covers key areas of this issue and documents the current state of affairs through various types of outputs.

[1] https://www.cgps.cz

[2] http://www.neonatology.cz

[3] https://data.unicef.org/resources/data_explorer/unicef_f/?ag=UNICEF&df=GLOBAL_DATAFLOW&ver=1.0&dq=.MNCH_MMR._T.&startPeriod=2016&endPeriod=2022

[4] https://ec.europa.eu/eurostat/web/main/data/database

[5] https://www.europeristat.com/index.php/reports/ephr-2019.html

[6] https://www.nzip.cz

AIMS

- To describe data sources and data collection process in collaboration with representatives of expert medical societies.
- To design and publish a comprehensive set of outputs above the available data describing the current state of the reproductive health agenda in the Czech Republic.

METHODS

The methodological communication background is mainly based on the global concept of sharing thematic agendas of the National Health Information System. It is based on the use of unified communication channels towards the lay, informed and professional public. The dedicated reproductive health section of the National Health Information Portal[7] contains statistical yearbooks and analytical reports, aggregated population and clinical time-series data, and thematically focused open datasets. It also includes a description of the data base and methodologies for data collection, review and validation, and refers to the applicable data collection guidelines.

RESULTS

BUSINESS UNDERSTANDING

The National Registry of Reproductive Health is the main data base. It is a population-based registry that consists of five modules: (i) assisted reproduction, (ii) newborns, (ii) abortions, (iv) parturient women (i.e. women in labour), (v) birth defects. The National Registry of Reproductive Health does not process personal data of women who have requested to keep their identities confidential during childbirth pursuant to Section 37 of Act No. 372/2011 Coll., on Health Services and Conditions of Their Provision, as subsequently amended. The care of pregnant and parturient women in the Czech Republic is guaranteed by a very dense network of healthcare facilities, which creates a desirable three-level system of care covering all types of care according to the risk of pregnancy and childbirth.

[7] https://www.nzip.cz/nrrz

ASSISTED REPRODUCTION MODULE

The purpose of the registry is to record all women who have been started on ovarian stimulation or monitoring for the treatment of sterility (their own sterility or the sterility of another woman in the case of oocyte donation) by the method of in-vitro fertilisation (IVF) or related techniques. Monitoring of IVF cycles provides the necessary information on the method, course, results and possible complications for the needs of health professionals, the Ministry of Health of the Czech Republic, health insurance companies and for international data reporting. The information obtained enables the evaluation of treatment procedures and is used for the management and improvement of care for infertile couples and for the implementation of state policy in the field of assisted reproduction and sterility treatment.

NEWBORNS MODULE

The purpose of collecting the required data is to provide the necessary information on perinatal care for the needs of health professionals, the Ministry of Health and for international data reporting. The information obtained is an important source for assessing the health status of newborns and is used for the management, evaluation and improvement of newborn care. The Newborns Module contains basic data on the immediate condition of the newborn after birth, its further health status, complications, treatment.

ABORTIONS MODULE

The purpose of collecting the required data is to provide data for assessing the quality of reproductive health care. The collection of data on abortion in the Czech Republic has become a long-standing tradition and an essential part of demographic and perinatological information on the Czech population. The anonymised data are submitted to the Czech Statistical Office on a quarterly basis for demographic statistics.

Data analysis provides a number of internationally recognised criteria for quality of care and quality of health and provides a necessary complement to other perinatological data, without which it is not possible to comprehensively assess the quality of reproductive health care.

PARTURIENTS MODULE

The purpose of collecting the required data is to provide basic information about the woman's reproductive history, the course of her pregnancy, childbirth and the newborn. Maternal follow-up is used to assess the health status of the woman in terms of quality of care. The information obtained is a valuable source of information for gynaecological and obstetric care and is an important tool for improving the care of pregnant and parturient women. Information from the registry is used to determine the conception and implementation of national health policy in the field of gynaecological and obstetric care and is also used for the databases of the World Health Organization (WHO) and the Organisation for Economic Co-operation and Development (OECD). In addition, it is provided to other international organisations in accordance with contractual obligations.

BIRTH DEFECTS MODULE

The purpose of the required data is to Registry prenatally and postnatally diagnosed birth defects in the population, which is currently one of the basic factors needed to assess the health status of the population and is an integral part of the evaluation of prenatal, perinatal and postnatal care. Monitoring the prevalence of birth defects contributes to the evaluation of early detection of birth defects. The information obtained is used to assess the health status and quality of the new population. The information is used as a basis for the development of national health policy in this area. They are used in the presentation of the data monitored abroad. The information is used for the World Health Organisation (WHO) and Organisation for Economic Co-operation and Development (OECD) databases.

The data collection process is described in summary form in Table 1 below. Data are entered into the Registry only electronically via data batches or by direct entry via an online form. Both of these inputs have built-in automatic basic checks for internal consistency of the data. In particular, the input checks the completion of mandatory items, the correctness of the provider and person identifier entry and the logical consistency of the fields filled in. In addition to the automatic checks, interim checks and a complex analytical check are performed before closing the data for the previous calendar year. The checks also include the settlement of duplicate records.

Table 1: Description of data collection on reproductive health in the Czech Republic

Module	What is reported:	Who is reporting:	When is the report made:	How is the report made:
Parturients	all births that took place in the Czech Republic	any provider in whose healthcare facility the birth took place or who carried out the first postnatal treatment of the parturient woman	after the end of the health service provided for each calendar month, by the end of the following calendar month	electronically: either by remote transmission or online via a web form
Newborns	all newborns born in the Czech Republic	any provider who provided health services to the newborn in connection with the birth, provided inpatient care to the newborn, or performed the first treatment of the newborn	after the end of the health service provided for each calendar month, by the end of the following calendar month	electronically: either by remote transmission or online via a web form
Abortions	termination of pregnancy by abortion, whether miscarriage or induced abortion, including termination of ectopic pregnancies	any provider who has performed an induced abortion or provided post-abortion care, or in whose healthcare facility an abortion was induced	after the end of the health service provided for each calendar month, by the tenth day of the following calendar month	electronically: either by remote transmission or online via a web form
Birth Defects	congenital developmental defects, genetic or rare diseases in the fetus, child or adult	any provider who diagnosed the defect	after the end of the health service provided for each calendar month, by the end of the following calendar month	electronically: either by remote transmission or online via a web form
Assisted Reproduction	any monitoring and/or treatment process aimed at achieving a woman's pregnancy by means of assisted reproductive techniques	any provider who has used assisted reproductive techniques and procedures	within 3 days of the date on which the relevant phase of treatment is started or the relevant procedure or examination is carried out	electronically: either by remote transmission or online via a web form

DATA UNDERSTANDING

The main data components of the NHIS, i.e. national registries that have been used for the analysis of the reproductive health agenda, include:

- The **National Registry of Health Services Providers** as a comprehensive Registry recording all types of health services providers and their basic characteristics. In addition to its own records, the registry enables analysis of time trends and dynamics in the number of providers. Data are updated monthly.
- The **National Registry of Healthcare Professionals** is a nationwide Registry of all health workers, i.e. doctors and individual non-medical and health professions. The Registry contains basic characteristics of the workforce such as age, sex, acquisition of relevant qualifications and place of work in the health sector. The data are updated monthly.
- The **National Registry of Reimbursed Health Services** contains data from health insurance companies as regards both inpatients and outpatients, including complete data on reported diagnoses, procedures and treatments; data are currently available for the period 2010–2022. The data are updated quarterly.
- The **National Registry of Hospitalised Patients** is a nationwide population-based registry that records inpatient admissions that were closed during the reporting period. Data are available from 1994 to 2022, with the full range of data monitored from 2007–2022.
- The **Death Certificate Information System** is the primary source of information on each death. It is completed by the examining physician immediately after the examination of the deceased, and records the sequence of causes leading to death (coded using ICD-10 since 1994) in addition to basic socio-demographic characteristics. Data are available until 2022.

Information on the demographic and epidemiological background of the Czech population, comorbidities and other patient characteristics was also needed, possibly as a source of information for complex analyses:

- demographic data (provided by the Czech Statistical Office),
- Czech National Cancer Registry,
- National Diabetes Registry,
- National Registry of Occupational Diseases,

— National Registry of Incapacity for Work,

— National Registry of Drug Addict Therapy.

DATA PREPARATION

At this stage, it was necessary to perform a thorough validation of all input data together with the necessary adjustments. The data required additions, transformations and aggregations so that no natural or legal person could be identified in subsequent processing or analysis. The published datasets must be fit for purpose as defined in the NHIS, and must be produced according to a standardised methodology. Examples of modifications may include flagging missing values, correcting erroneous records, unifying data types for descriptive attributes, aggregating groups of records across patient age categories, or calculating new variables.

MODELLING

A number of outputs were generated over the data described in the previous sections of the case study. For illustrative purposes, some of these are briefly described below to show the variation in the types of materials with respect to the target group.

STATISTICAL YEARBOOKS AND ANALYTICAL REPORTS

Yearbooks[8] are historically well-established and still widely used static outputs available in PDF format. For example, statistical data on demographic characteristics and maternal health during pregnancy and childbirth, possible complications, as well as obstetric and fetal and neonatal health data are published at regular intervals. The publication combines information obtained by gynaecologists, obstetricians and neonatologists, and possibly (in the case of stillbirth or early death) also by pathologists. Birth statistics are also compiled by the Czech Statistical Office (CZSO), but the data from the Registrys in the present publication are more detailed, more extensive and include more health characteristics. Although this type of output may seem outdated or unusable for machine processing, it undoubtedly continues to find its consumers.

Analytical studies[9] in the form of presentations with graphical and tabular outputs summarise the defined issues and briefly add the necessary and desirable interpretation of the conclusions. The three reports below have been published under the reproductive health agenda as of 31 May 2023:

[8] https://www.nzip.cz/kategorie/249-narodni-registr-reprodukcniho-zdravi-rocenky-publikace

[9] https://www.nzip.cz/kategorie/250-narodni-registr-reprodukcniho-zdravi-analyticke-studie

- Analytical summary of selected data
  - This report summarises key population and clinical data collected in the National Registry of Reproductive Health on mothers and newborns, including long-term time series.
- Selected reproductive health indicators and provision of maternal and newborn care in international comparison
  - This report assesses selected indicators of maternal and newborn health outcomes and quality of care. It is the first version of this type of analysis, for which internationally recognised indicators were selected, notably those of Eurostat, Unicef, the World Health Organization and Europeristat. Furthermore, the report addresses the issue of self-assessment of quality and safety of care with regard to the availability of the necessary input parameters.
  - Reproductive health indicators in the CZ-DRG system
  - The report describes the volume of acute inpatient care for pregnancy, childbirth, puerperium and newborns in relation to the CZ-DRG classifications[10].

One example of analytical summaries is a comparison of the coverage of gynaecological and obstetric care in each EU Member State together with the average of EU (Figure 1).

[10] https://drg.uzis.cz

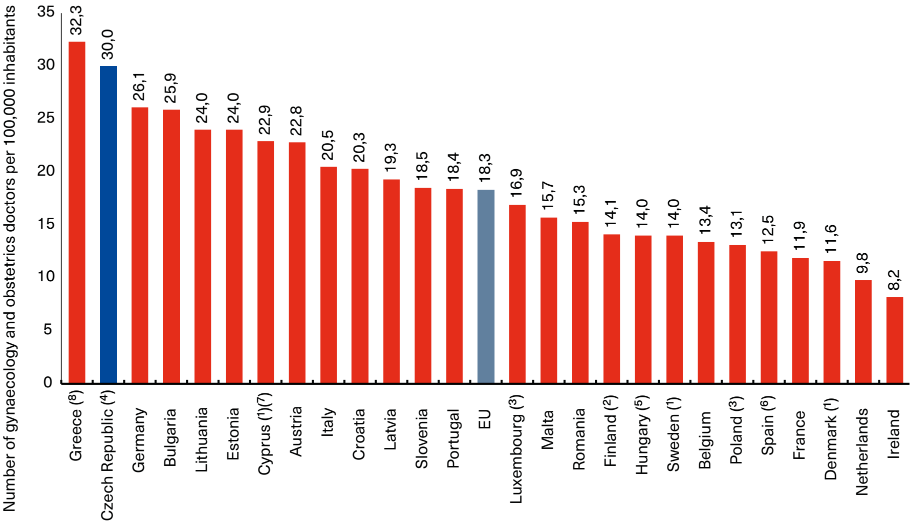


Figure 1: Number of gynaecology and obstetrics doctors per 100,000 inhabitants (data source: Eurostat, 2020). Note: 1 2019, 2 2018, 3 2017, 4 Estimated, 5 Break in time series, 6 Data on specialists refer only to physicians working in hospital, 7 Incomplete coverage for total physicians, 8 Except total, 2019

SUMMARY POPULATION AND CLINICAL DATA IN TIME SERIES[11]

This type of output is usually published in XLSX format and provides aggregated views of available data in a ready-to-use format. MS Excel allows the formatting to be defined, and therefore the resulting data summary is clearer for many users without the need for further processing. This differentiates it from open data and has a very significant added value for selected users.

- Selected indicators of healthcare in maternity hospitals in the Czech Republic
  - The primary objective of this dataset is to provide information on the volume and type of care in individual maternity hospitals in the Czech Republic. It is a thematically focused dataset, allowing in particular to monitor the development of specific parameters over time in one maternity hospital or to compare the values of parameters between maternity hospitals of the same type and volume of care.

[11] https://www.nzip.cz/kategorie/260-narodni-registr-reprodukcniho-zdravi-datove-souhrny

- Parturient women by region of residence of the mother 2000–2021
  - The report provides basic information on parturient women and the course of their births by region of residence. In addition to trends in the number of births, it also provides the number of complications during pregnancy, during labour, information on induction of labour, way of delivery and the total length of hospital stay of the parturient woman after delivery. The stratification criteria available in this analysis are maternal age, gestational age, parity and possible multiple pregnancy.
- Parturient women by region of residence of health services provider 2000–2021
  - The report provides basic data on parturient women and births in the Czech Republic overall and by region of health services provider. It thus makes it possible to monitor differences between mothers, but also regional differences in care. The differences in care and the concentration of risk cases between regions are mainly determined by the presence of specialised centres in the region.
- Newborns by region of residence of the mother 2000–2021
  - The report presents basic data on newborns in the Czech Republic and individual regions by mother's place of residence. In addition to basic trends in the number of births by vitality and location, it also presents antenatal and perinatal characteristics, information on postnatal adaptation, respiratory, nervous and sensorimotor system diseases, infections, mortality and length of hospitalisation. The stratification criteria available in this analysis are birth weight and gestational age.
- Newborns by region of residence of health services provider 2000–2021
  - The report provides basic data on newborn in the Czech Republic overall and by region by the location of health services providers. It thus allows not only to monitor the above characteristics of births, but also differences in care between regions. The differences in care and the concentration of high-risk cases between regions are mainly determined by the presence of specialised centres in the region.

THEMATIC OPEN DATASETS[12]

The primary goal of these outputs is to describe a narrowly defined thematic area and enable users to do their own analytical processing. It contains real, fully anonymised individual data on births and newborns in the Czech Republic. The

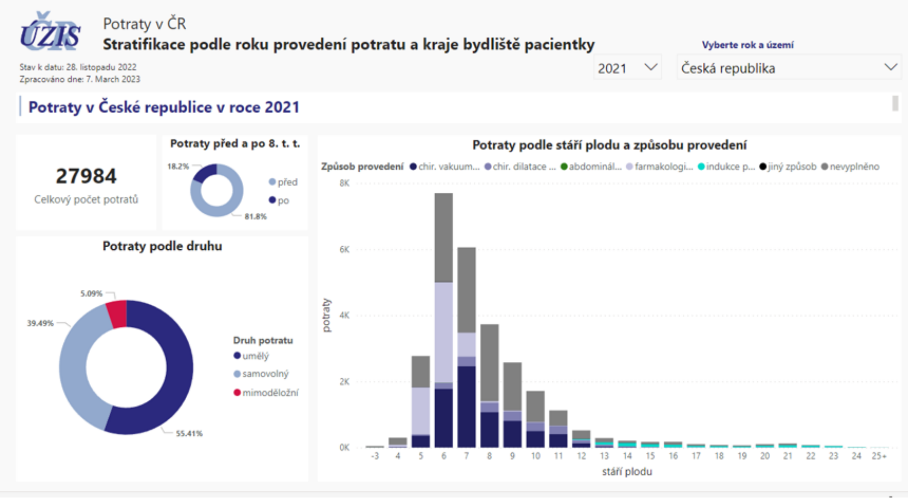
[12] https://www.nzip.cz/kategorie/252-narodni-registr-reprodukcniho-zdravi-otevrena-data

datasets are completely separate, and it is impossible to interlink them – in order to ensure data protection, in accordance with the current legislation. In the first phase (as of 31 May 2023), a total of 17 open data sets have been published; three examples are given below. Each dataset has its own entry in a local open data catalogue, meets technical standards and links to full documentation. This includes, among other things, a short annotation and objectives, the interpretation and informational value of the set, the information limits of the set, examples of use, the type of licence, the author collective and a statement on ensuring the protection of personal data and the identity of legal entities.

- Parturient women and screening tests in pregnancy
  - The primary objective of this dataset is to assess prenatal care in pregnant women. It is a thematically focused dataset, allowing in particular for the assessment of antenatal care use in pregnancy, as it includes a summary of available antenatal screening, week of antenatal care start and number of antenatal care checks. The dataset presents data for the period 2000–2021.
- Parturient women and caesarean sections
  - The primary objective of this dataset is to evaluate caesarean sections. The dataset includes a summary of basic information about the parturient women, such as age or previous history of caesarean section, and allows further summarisation of their deliveries by year of event and type of health services provider. It also lists the anaesthesia used for caesarean section and the type of caesarean section, which is defined by the start of labour and the date of determination of termination of pregnancy by caesarean section. The dataset presents data for the period 2000–2021.
- Newborns by basic socio-demographic characteristics
  - The primary objective of this dataset is to assess the main characteristics of newborns. The dataset includes an overview of basic data on newborns, allowing for summarisation by year of event, region of residence of mother, possible multiple pregnancy, vital rates, sex, birth weight, and gestational age. In particular, the dataset allows for demographic analysis of fertility and newborns over the long time series 1994–2021.

INTERACTIVE VISUALISATIONS

Another type of output that appeals to a target community other than data experts are easy-to-understand and easy-to-use visualisations that have been developed based on freely available datasets. In close collaboration between analysts and representatives of expert medical societies, aggregated dynamic reports are gradually created, with the user having the option of configuring or selecting the available attributes. These filters offer, for example, the selection of a specific year or regional unit (Figure 2).



Potraty v ČR

Stratifikace podle roku provedení potratu a kraje bydliště pacientky

Stav k datu: 28. listopadu 2022
Zpracováno dne: 7. March 2023

Vyberte rok a území: 2021 | Česká republika

Potraty v České republice v roce 2021

27984
Celkový počet potratů

Figure 2: Sample visualisation of a selected dataset using the BI tool (available only in Czech language)

EVALUATION

Data checking can be divided into two phases. The first is directed towards the primary data and its validation. A quantitative check is continuously carried out, where it is verified that all providers have submitted data for the past calendar month to all modules to which they are required to report. If a shortcoming is found, providers are promptly notified and asked to correct it. In addition, a more in-depth analytical review of the data is performed prior to closing the data for the past calendar year. The completeness of the data is checked over the records reported to the National Registry of Reimbursed Health Services, the National Registry of Hospitalised Patients, births and deaths (as reported by the Czech Statistical Office). Any drop in numbers compared to previous periods is also identified. Qualitative checks are carried out in each module. The most important checks in the Parturients and Newborns modules include interlinking. Each report on a parturient woman must be accompanied by a record on a newborn and vice versa. Checking the completeness of the Assisted Reproduction and Abortion modules against the National Registry of Reimbursed Health Services data is more difficult because some of this care is not covered by public health insurance. Therefore, only the part of the care that is included in the National Registry of Reimbursed Health Services is checked. Finally, the overall internal consistency of the records is subject to scrutiny. Similarly, the control of the Birth Defects module against the data of the central Registry of reimbursed services is only partial, due to the reporting method. Again, there are qualitative checks within the Registry, such as checks on the prevalence of specific diagnoses and diseases over time, between regions and by sex.

The second phase focuses on checking the outputs that are created over the primary data. All open datasets, aggregated datasets, analytical studies, yearbooks, and of course interactive visualisations and infographics are thoroughly reviewed in terms of language, form and content.

DEPLOYMENT

The deployment process in this case is very simple. Thanks to the existence of a unified communication platform in the field of Czech healthcare, health in general and health literacy, the information system on reproductive health was published on the National Health Information Portal[13]. This platform provides comprehensive technical support for content editing and will also be used to update and add other relevant outputs. The open datasets have been published with all formal and technical requirements in the local catalogue of open data in healthcare, which is linked to the national catalogue[14].

DISCUSSION

The comprehensive agenda on reproductive health was published as a complex information system that tracks all essential aspects of maternal and newborn care. The data for the period 2020 to 2022 confirm the high level of access to and quality of this care in the Czech Republic, which has not been disrupted even by the COVID-19 epidemic. Based on the unquestionable data provided by the information service, it can be concluded that the Czech Republic has never had better perinatal care outcomes in its history and is clearly among the safest countries in the world in terms of childbirth. Despite this, attention and priorities must continue to be given to further improving the quality and safety of care, and the newborns must not be forgotten alongside the parturient women. An extension of the set of indicators assessed for maternity hospitals is therefore in the pipeline, so that the indicators adequately cover neonatal care. Ensuring

[13] https://www.nzip.cz/nrrz

[14] https://data.gov.cz/datov%C3%A9-sady

full availability of care for mothers and newborns in all regions of the Czech Republic is a major priority of the Ministry of Health of the Czech Republic, which means guaranteeing adequate reimbursement from public health insurance. However, continuous quality control must not be forgotten, which is the subject of the Ministry's overall concept in cooperation with the management of expert medical societies. In terms of further development, it is also necessary to emphasise the need to standardise hospital information systems, which is a necessary condition for obtaining correct and objective data.

EVALUATION OF THE AIMS OF THE CHAPTER

- To describe data sources and data collection process in collaboration with representatives of medical expert societies.
  - Thanks to the complex system of national health registries, it was possible to clearly describe all the input data sources on which the reproductive health information system is built and to specify key aspects of data collection in relation to individual modules.
- To design and publish a comprehensive set of outputs above the available data describing the current state of the reproductive health agenda in the Czech Republic.
  - Under the guarantee of medical expert societies, it was possible to suitably compile typologically targeted outputs for individual groups of users so that they are maximally useful in terms of information delivery. They always consider the need and objectives of the take-home message, contain an interpretation of the conclusions and are further expandable according to specific requirements from real practice.

LESSONS LEARNED

The topic of reproductive health has been described in collaboration with expert medical societies so that a comprehensive set of outputs has been created, including static yearbooks, detailed analytical studies, data summaries and open datasets. The critical phases for such a comprehensive information service were domain understanding, modelling and evaluation. The available data have provided a tool replicable in practice, enabling an ongoing quality control of maternal and newborn care. To further develop this system, standardisation of hospital information systems is needed, which would contribute significantly to obtaining high-quality data.

Interactive visualisations and reports are being developed over open datasets that are available without any limitations to show the less advanced user the nature of data and key information contained, including correct interpretation.

ACKNOWLEDGEMENTS

The authors of this case study would like to thank:

— the leadership of the Czech Society of Neonatology,

— the leadership of the Czech Gynaecological and Obstetrical Society,

— the members of the analytical team and the webstudio of the Institute of Health Information and Statistics of the Czech Republic: Pohlová R., Zouharová A., Kučerová H., Menšíková K., Míšková M., Mužík J., Šanca O., Zelinková, H., Kerberová M., Gregor J., Vičar M., Bulhart V.

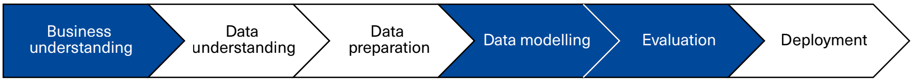
SECTION C

| 12 |

NATIONAL CARDIOLOGY INFORMATION SYSTEM: DATA INFRASTRUCTURE AND COMPREHENSIVE INFORMATION SERVICE

Jiří Jarkovský, Martin Komenda, Michal Vráblík, Miloš Táborský, Aleš Linhart, Jiří Pařenica, Klára Benešová, Jakub Gregor, Lenka Šnajdrová, Ladislav Dušek

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

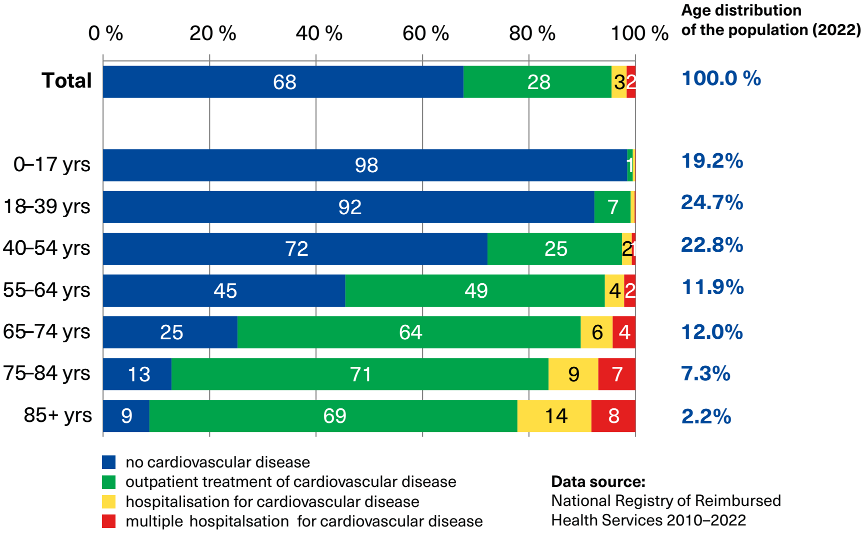
Year	2023–now
Keywords	Cardiology, cardiovascular disease, database, open data
Research question	How to design and implement a data infrastructure for comprehensive information communication of cardiovascular health plan in the Czech Republic?
Type of result	Interactive visualisation
Level of data processing	Advanced analyses, Open datasets

DATA TO DOWNLOAD

INTRODUCTION

CURRENT SITUATION

Cardiovascular diseases in the Czech Republic are one of the biggest challenges for the Czech healthcare system. The burden of cardiovascular diseases in the Czech population is enormous: in 2022 alone, 2.8 million patients were reported to receive healthcare in this area, and the trend is increasing. In addition, the medical history must be considered: in 2022, cardiovascular disease treatment was reported in 32% of the population over the last five years. This proportion increases with age, reaching 90% for those aged 75 and over (Figure 1).



Data source:
National Registry of Reimbursed
Health Services 2010–2022

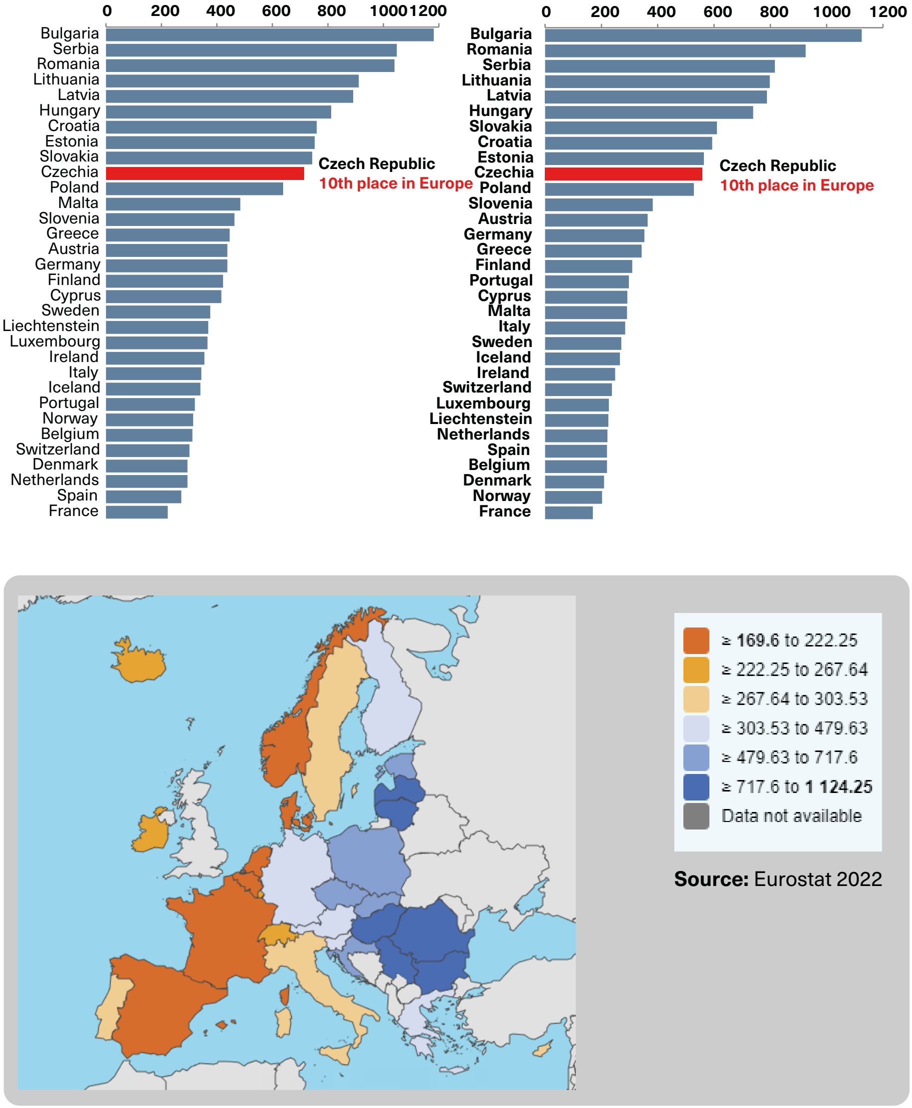
The occurrence of cardiovascular disease in a patient is defined by 1) hospitalization for a diagnosis I00-I99 (excluding I60-I69), Q20-Q29 in 2018-2022 or 2) reporting a diagnosis I00-I99 (excluding I60-I69), Q20-Q29 by outpatients healthcare (general practitioner, internal medicine), cardiologist, pediatric cardiologist) combined with the reporting of a drug from ATC group C (= cardiovascular system) in 2018-2022.

Figure 1: Burden of cardiovascular diseases of the Czech population in 2022

The occurrence of cardiovascular diseases will probably further increase in future due to adverse trends in population age structure (population above 75 years will double in the year 2050 in comparison to the year 2020) and the prevalence of obesity in the population (the 4th worst position among EU countries) and other lifestyle factors.

Cardiovascular diseases are the most important cause of mortality in the Czech Republic, with 39% of all mortality cases from cardiovascular causes in 2022. Although this proportion is decreasing over time (56% in 1994 and

49% in 2012) due to continuously improving treatment outcomes, the position of the Czech Republic among European countries does not change over time (Figure 2).



Source: Eurostat 2022

Figure 2: Mortality per 100,000 inhabitants caused by cardiovascular diseases in European countries

Nevertheless, the demographic and lifestyle problems are countered by quality healthcare with continuously improving treatment outcomes. Cardiovascular care is highly centralised, with coverage of all regions of the Czech Republic (Figure 3).

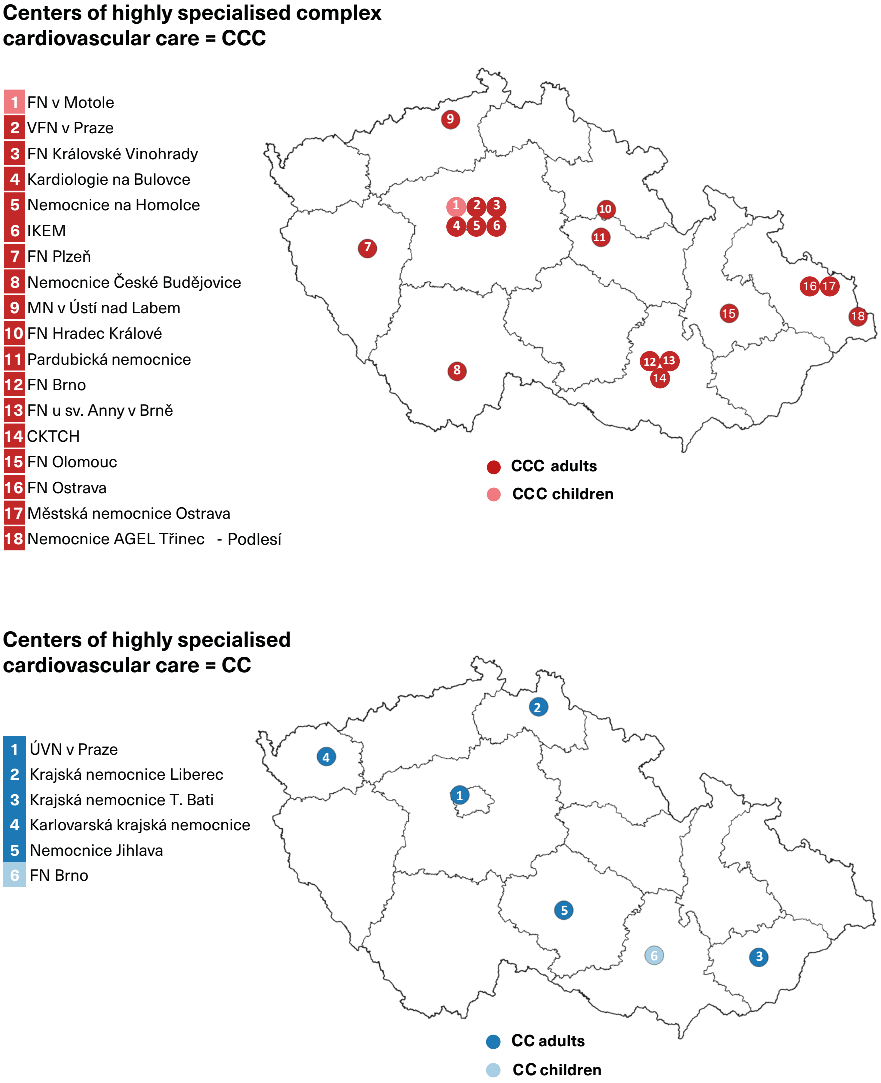
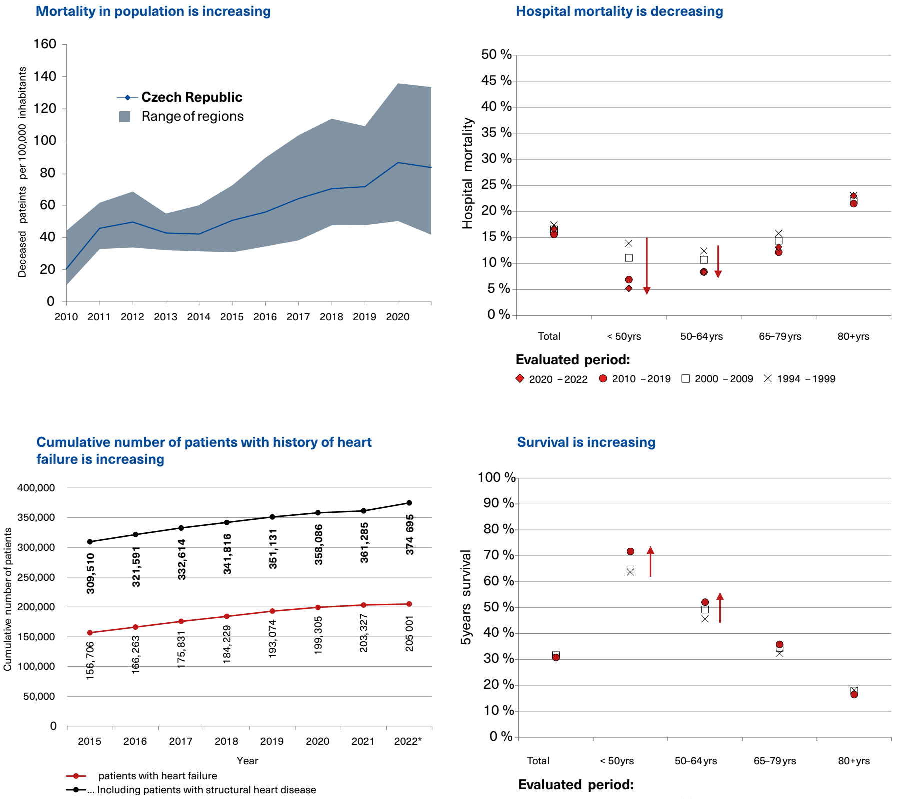


Figure 3: Centers of highly specialised complex cardiovascular care and highly specialised cardiovascular care (available only in Czech language)

The improved healthcare manifests in fewer acute hospitalisations, shortened length of stay, decreased hospital mortality, and improved long-term survival. The improvement of care combined with demographic changes affects the structure of cardiovascular diagnoses where acute coronary syndromes and ischemic heart disease shows a decreasing trend compared to, for example, heart failure and arrhythmias (with an increasing trend). Heart failure is an excellent example of these processes with future impact on cardiovascular care: demographic changes increase the number of cases and mortality in the population, whereas improved acute care decreases hospital mortality and increase long-term survival (Figure 4).


Mortality in population is increasing

Hospital mortality is decreasing

Cumulative number of patients with history of heart failure is increasing

Survival is increasing

Figure 4: Heart failure in time trends

The improved acute care and increased life expectancy brings increased need for palliative and end-of-life care for cardiovascular patients; it can be documented by the analysis of palliative teams in hospitals where 13% of all indications for palliative care were cardiovascular diagnoses.

The situation requires a conceptual approach, as required by the National Cardiovascular Health Plan; this is supported by the availability of data from the National Cardiology Information System, and is necessary for the description of the current situation in epidemiology, treatment, personal and other capacities of healthcare etc., predictive modelling, evaluation of results of measures taken, benchmarking, international comparison, and all other data requirements; all aspects of cardiovascular care through the course of human life have to be included (Figure 5).

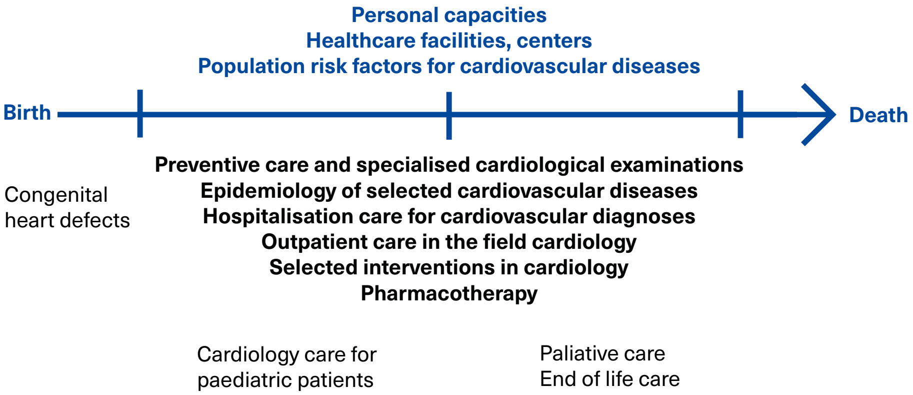

Personal capacities
Healthcare facilities, centers
Population risk factors for cardiovascular diseases
Birth
Death
Preventive care and specialised cardiological examinations
Epidemiology of selected cardiovascular diseases
Hospitalisation care for cardiovascular diagnoses
Outpatient care in the field cardiology
Selected interventions in cardiology
Pharmacotherapy
Congenital heart defects
Cardiology care for paediatric patients
Paliative care
End of life care

Figure 5: Required aspects of cardiological care for data coverage

NATIONAL CARDIOVASCULAR HEALTH PLAN

The creation of the National Cardiovascular Health Plan (NCHP) for 2023–2033 is a conceptual follow-up to the National Cardiovascular Health Programme of the Czech Republic from December 2013. Its purpose is to set the main strategic goals for maintaining and further improving the quality of preventive and therapeutic care of cardiovascular diseases; all this is based on the analysis of the epidemiological situation and trends of the last decade. The NCHP also includes the development of a system of quality indicators to monitor and further improve care, as well as the prediction of the epidemiology of cardiovascular diseases in relation to its current development and the prediction of the development of the population of the Czech Republic. An integral part of the NCHP is the implementation of information and material resources for the general population, enabling effective lifestyle changes and reducing exposure to risky behaviours and environments. The following are the key strategic objectives that the plan should progressively deliver over the period 2023–2033:

- Strategic objective 1: Availability of epidemiological data and analyses of quality of care indicators
- Strategic objective 2: Primary cardiovascular prevention
- Strategic objective 3: Availability and quality of care
- Strategic objective 4: Highly specialised and centralised care
- Strategic objective 5: Integration of cardiology care in the context of the whole healthcare system
- Strategic objective 6: Research and science

The design, development, implementation and management, including ensuring the updating and content expansion of the NCIS, fall under Strategic Objective 1, i.e. under the responsibility of the Institute of Health Information and Statistics of the Czech Republic (IHIS). The data basis is provided by the National Health Information System in the form of regularly updated and publicly available information that maps the basic epidemiology and resources of cardiovascular care. The general vision of the NHCP is to ensure that every resident of the Czech Republic knows how to prevent the development of cardiovascular disease and, if it develops, to ensure the highest possible quality of care and life, regardless of geographical location or stage of the disease.

NATIONAL CARDIOLOGY INFORMATION SYSTEM FOR THE NATIONAL CARDIOVASCULAR HEALTH PLAN

In accordance with the establishment of the National Cardiovascular Health Plan [1] for the years 2023–2033 under the supervision of the Czech Society of Cardiology[1] as the expert guarantor of cardiovascular care in the Czech Republic, the first version of the National Cardiology Information System was designed and published. It is based mainly on the data from the National Health Information System (NHIS) and its components (national registries maintained according to the applicable laws) supplemented by the data from standardised data warehouses of the DRG project reference hospital network. The main challenges of the newly built information platform include:

- Strengthening the computerisation and interoperability of partial data collection, introduction of standardised e-recording of cardiovascular disease diagnosis and treatment, and standardisation of hospital information system exports.

[1] https://www.kardio-cz.cz

- Full computerisation of data collection of the National Cardiology Information System, including real-time linking of reports from the laboratory segment.
- Completion of a predictive superstructure above nationwide sub-registries to strengthen predictions on economic and staffing needs, the impact of new technologies and drugs.
- Completion of a comprehensive information system for mapping patient trajectories in the health service system, identifying desirable and undesirable trajectories and evaluating measures for improvement.
- Developing a comprehensive information system for planning and evaluating end-of-life care for patients.
- Strengthening the publication of comprehensive indicators of access to and quality of care and departmental reference statistics.
- Implementation of the open data concept in cardiology and preparation for the EHDS (European Health Data Space).

The aim of the National Cardiology Information System (NCIS) is to maximise the use of existing data sources and minimise the administrative burden on healthcare providers and healthcare professionals. The NCIS is being built based on multi-source data integration covering all essential dimensions of assessment and segments of cardiac care across the life course. In terms of integration into the overall information architecture, which aims to disseminate valid and guaranteed information in the Czech healthcare sector, the NCIS is integrated as a specialised section within the National Health Information Portal[2]. The basic idea is to fully cover the key areas according to the strategic goals outlined in the National Cardiovascular Health Plan and to present them in an understandable form to the lay, informed and professional public.

AIMS

- To maximise the potential of using available data sources and incorporate them into the outputs of the National Cardiology Information System in close collaboration with representatives of medical expert societies.
- To design and publish a comprehensive set of outputs above the available data describing the current state of the cardiology in the Czech Republic.

[2] https://www.nzip.cz/nkis

— To provide data source for scientific research in the field of cardiology in cooperation with the Czech Society of Cardiology.

METHODS

METHODS OF DATA PROCESSING AND ANALYSIS

Primary data for analysis are stored in relational databases MS SQL Server and Vertica and processed using SQL language; data accessible to analysts are pseudonymised for the reasons of personal data protection. During the development of the NCIS, a new data layer was derived from the primary data combining the collected primary data, the clinical knowledge of the physicians, the knowledge of public health insurance data reporting and data linkage of information from the different NCIS registries. This derived data layer is the result of an iterative process between data analysts and experts and contains clinically relevant information that serves as the basis for all other outputs, i.e. analytical studies, data summaries and open data. A suite of statistical, automation and visualisation software such as SPSS, Stata, R, QGIS, ReportBuilder and PowerBI are used to deliver specific outputs.

UNIFORMLY CONTROLLED COMMUNICATION

The methodological communication background is mainly based on the global concept of sharing thematic agendas of the National Health Information System. It is based on a unified communication channel towards the general, informed and professional public. The dedicated cardiology section of the National Health Information Portal[3] contains statistical yearbooks and publications, analytical studies, data summaries and thematic open datasets.

RESULTS

BUSINESS UNDERSTANDING

The data base for the evaluation of cardiological care in the Czech Republic is being built by the Czech Society of Cardiology as a comprehensive and robust information system, which relies mainly on data from the NHIS and its components (national registries maintained according to the applicable laws). These data sources enable a comprehensive assessment of the cardiovascular status of the population throughout the life course and of cardiovascular healthcare from various aspects: from the epidemiology of cardiovascular diseases, the

[3] https://www.nzip.cz/nkis

healthcare provided, the characteristics of patients and their survival to the staff capacity and availability of healthcare.

DATA UNDERSTANDING

The following data sources (national registries as components of the NCIS) are mainly used for data reporting purposes:

- The **National Registry of Health Services Providers** as a comprehensive registry recording all types of health service providers and their basic characteristics. In addition to its own records, the registry enables analysis of time trends and dynamics in the number of providers. Data are updated monthly.
- The **National Registry of Healthcare Professionals** is a nationwide registry of all healthcare workers, i.e. doctors and individual non-medical and healthcare professions. The registry contains basic characteristics of the workforce such as age, sex, acquisition of relevant qualifications and place of work in the health sector. Data are updated monthly.
- The **National Registry of Reimbursed Health Services** contains data from health insurance companies as regards both inpatients and outpatients, including complete data on reported diagnoses, procedures and treatments; data are currently available for the period 2010–2022.
- The **National Registry of Hospitalised Patients** is a nationwide population-based registry that records inpatient admissions that were closed during the reporting period. Data are available from 1994 to 2022, with the full range of data monitored from 2007–2022.
- **National Registry of Cardiovascular Surgery and Interventions**
- **The Cardiac Surgery Module** records all cardiac surgery performed, with data available in a uniform format from 2007. The registry covers the activity of 100% of cardiac surgery centres.
- **The Cardiovascular Interventions Module** records all coronary and non-coronary catheterisation cardiovascular interventions performed. Uniform data are available since 2005.
- The **Death Certificate Information System** is the primary source of information on each death. It is completed by the examining physician immediately after the examination of the deceased and records the sequence of causes leading to death (coded using ICD-10 since 1994) in addition to basic socio-demographic characteristics. Data are available until 2022.

In addition to the main components of the NHIS (national registries), the following sources of information on the demographic and epidemiological background of the population of the Czech Republic, birth defects, comorbidities and other characteristics of patients can be used for analyses, or as a source of data for complex analyses linking individual sub-registries:

- demographic data provided by the Czech Statistical Office (CZSO),
- National Cancer Registry,
- National Diabetes Registry,
- National Registry of Reproductive Health,
- National Registry of Injuries
- National Registry of Joint Replacement,
- National Registry of Occupational Diseases,
- National Registry of Incapacity for Work,
- National Registry of Drug Addict Therapy.

The European Health Interview Surveys performed in the Czech Republic are also an important part of health statistics internationally. From the perspective of cardiology, the European Health Interview Survey and the European Health Examination Survey are relevant, assessing the prevalence of risk factors for the development of cardiovascular disease in the population.

The assessment of the burden, performance, outcomes and real costs of acute inpatient care in cardiology in the Czech Republic is based on a legally anchored and fully sustainable CZ-DRG system[4]. A major contribution of the Czech concept in this area is the long-standing reference network of hospitals that generate an annual reference database of all hospital admissions with high resolution of care content and cost items. For acute inpatient care, this system represents the interrelated classification rules, methodological procedures and algorithms, codebooks, information systems and software tools that are necessary to ensure correct functioning. The main contributions of the CZ-DRG are: (i) reflection of the real acute inpatient care provided, (ii) reflection of the real cost of acute inpatient care.

[4] https://drg.uzis.cz

DATA PREPARATION

At this stage, it was necessary to perform a thorough validation of all input data, together with the necessary transformations, registry linking and generation of derived, clinically relevant data. In addition, anonymisation and aggregation procedures were applied in the creation of data outputs and open datasets to prevent the identification of individuals or legal entities. Published datasets must be fit for purpose as defined in the NHIS and must be produced according to a standardised methodology. Examples of modifications may include flagging missing values, correcting erroneous records, unifying data types for descriptive attributes, aggregating groups of records across patient age categories, or calculating new variables.

MODELLING

A number of outputs were generated over the data described in the previous sections of this case study. To illustrate, some of these are briefly described below; the aim is to show the variation in the types of materials with respect to the target group.

STATISTICAL YEARBOOKS AND ANALYTICAL STUDIES

Yearbooks[5] are historically well-established and still widely used static outputs available in PDF format, possibly supplemented by data tables in the form of formatted outputs in XLSX format. For example, brief summaries of data from the National Registry of Cardiovascular Surgery and Interventions for a given period are published here at regular intervals. Although this type of output may seem outdated or unusable for machine processing, it undoubtedly continues to find its consumers.

Analytical studies[6] in the form of presentations with graphical and tabular outputs summarise the defined issues and briefly add the necessary and desirable interpretation of the conclusions. The six reports listed below forming the summary analytical study have been published as of 13 June 2023 as part of the NCIS agenda:

— Analytical summary of selected data
    — This report summarises key population and clinical data necessary to establish the goals and measurable indicators of the National Cardiovascular Health Plan. It covers all essential dimensions of cardiology care evaluation:

[5] https://www.nzip.cz/kategorie/259-narodni-kardiologicky-informacni-system-rocenky-publikace

[6] https://www.nzip.cz/kategorie/254-narodni-kardiologicky-informacni-system-analyticke-studie

    - Primary and secondary prevention of cardiovascular disease.
    - Evaluation of the treatment burden on providers.
    - Detailed evaluation of acute inpatient care with an emphasis on specialised and highly specialised care.
    - Special emphasis on the care of pediatric patients with cardiovascular disease.
    - Special emphasis on the evaluation of end-of-life care for patients with cardiovascular disease.
    - Indicators of access, outcomes and quality of care.
    - etc.

- Epidemiology of selected cardiovascular diseases
  - Case studies including reported diagnoses, procedures, treatment: hypertension, heart rhythm disorders, acute coronary syndrome, coronary artery disease, heart failure.
- Acute inpatient care for diseases and disorders of the circulatory system in the CZ-DRG system
- International comparison of cardiovascular disease prevalence, risk factors and lifestyle indicators
  - The presented analytical report is mainly focused on the comparison of the prevalence of cardiovascular diseases in the Czech Republic and in other countries of Europe and the world, as well as risk factors for their occurrence. The analysis is based mainly on data from the European Health Interview Survey and Eurostat[7].
- Cardiovascular prevention in the Czech Republic
  - Analytical report on prevention and risk factors for cardiovascular disease in the Czech Republic.
- Heart failure in the Czech Republic
  - Example of a detailed analysis of the epidemiology with a comprehensive definition in the National Registry of Reimbursed Health Services data, including reported diagnoses, procedures and treatment.

[7] https://ec.europa.eu/eurostat/web/main/data/database

DATA SUMMARIES[8]

This type of output is usually published in XLSX format and provides an aggregated view of the available data in a clear form. MS Excel allows the user to define the formatting. Therefore, for many users, the resulting data summary is clearer and without the need for further processing. This differentiates it from open data and has a very significant added value for selected users.

- Epidemiology of patients with coronary heart disease – prevalence of patients with a history of the disease
  - This report provides basic data on the epidemiology of patients with a history of coronary heart disease from the National Registry of Reimbursed Health Services data based on reported care. The data are stratified by sex, age, and region of residence; the results are expressed in absolute numbers, patient demographics, and per 100,000 population in each category.

- Cardiac surgery performed in the period 2007–2021
  - This report presents basic data on cardiovascular surgeries in the Czech Republic between 2007 and 2021 from the National Registry of Cardiovascular Surgery and Interventions, namely the Cardiac Surgery Module. The socio-demographic characteristics of operated patients, their cardiac history and history of previous interventions, preoperative examination and the status and characteristics of the cardiovascular surgery performed are summarised.

- Number of cardiology facilities by region
  - This report provides the latest data (May 2023) from the National Registry of Health Services Providers on the number of cardiology healthcare facilities broken down by specific specialty, type of health care facility and specialty centre status within the regions and districts of the country.

- Epidemiology of patients with heart failure – prevalence of patients with a history of the disease
  - This report presents basic data on the epidemiology of patients with a history of heart failure from the National Registry of Reimbursed Health Services data based on reported health care. The data are stratified by sex, age, and region of residence, and results are expressed in absolute numbers, patient demographics, and per 100,000 population in each category.

[8] https://www.nzip.cz/kategorie/258-narodni-kardiologicky-informacni-system-datove-souhrny

- Cardiovascular interventions performed in the period 2005–2021
  - This report presents basic data on cardiovascular interventions in the Czech Republic between 2005 and 2021 from the National Registry of Cardiovascular Surgery and Interventions, namely the Cardiovascular Interventions Module.

THEMATIC OPEN DATASETS[9]

The primary goal of these outputs is to cover a narrowly defined thematic area and enable users to do their own analytical processing. It contains real, fully anonymised individual data on selected topics of cardiovascular disease in the Czech Republic. The datasets are completely separate, and it is impossible to interlink them – in order to ensure data protection, in accordance with the current legislation. A total of six open data sets were published in the first phase (as of 13 June 2023); three examples are given below. Each dataset has its own entry in a local open data catalogue, meets technical standards and links to full documentation. This includes, among other things, a short annotation and objectives, the interpretation and informational value of the set, the information limits of the set, examples of use, the type of licence, the authors' collective and a statement on ensuring the protection of personal data and the identity of legal entities.

- Burden of cardiovascular diseases in the population of the Czech Republic
  - This dataset allows the assessment of the number of people in the population of the Czech Republic who have been identified with cardiovascular disease (CVD). Data on CVD prevalence are available from 2014 onwards; the user can stratify the data found in each year according to the demographic characteristics of the patients (sex, age category, region and district of residence).

- Deaths from circulatory diseases
  - The aim of the dataset is to provide the possibility to assess the frequency of deaths from circulatory diseases from 1994 to the last closed year. The user can stratify the data by person characteristics (age at death category, sex, residence at death, cause of death) and time (year) or by combinations of these.

- Epidemiology of heart failure
  - The dataset allows to assess the epidemiological characteristics of heart failure (incidence, prevalence, mortality) in the population of the Czech Republic in the years 2015–2021. The user can stratify the data found in

[9] https://www.nzip.cz/kategorie/255-narodni-kardiologicky-informacni-system-otevrena-data

each year according to the demographic characteristics of patients (sex, age category, region of residence). It is also possible to assess the number of heart failure patients treated in outpatient or acute inpatient care.

EVALUATION

Data checking can be divided into two phases. The first phase is directed towards the primary data and its validation. Given the combination of different data sources, the validation of primary data can be divided into three main areas.

1. For NHIS clinical registries, there is a continuous quantitative check to verify that all providers have submitted data to all modules to which they are required to report. If a shortcoming is identified, providers are promptly notified and asked to correct it. In addition, a more in-depth analytical review of the data is performed prior to closing the data for the previous calendar year. The completeness of the data is checked over the records reported to the National Registry of Reimbursed Health Services, the National Registry of Hospitalised Patients, births and deaths (as provided by the Czech Statistical Office). Any drop in numbers compared to previous periods is also identified. Qualitative checks are carried out in each module.

2. For the data of the National Registry of Reimbursed Health Services, data updates are carried out every three months, where the batch data submitted by health insurance companies are evaluated for consistency and trend changes according to a standardised protocol; in case of a detected shortcoming, health insurance companies are immediately notified and asked to correct it.

3. Mortality data form an integral part of the assessment of cardiovascular diseases; the data are generated in cooperation between the IHIS and the CZSO and are released for analysis only after the settlement of the causes of death, based on the death certificate.

The second phase focuses on checking the outputs created over the primary data. All open datasets, aggregated datasets, analytical studies, yearbooks and, of course, in the future, interactive visualisations and infographics are thoroughly checked in terms of language, form and content.

DEPLOYMENT

The deployment process in this case is very simple. Thanks to the existence of a unified communication platform in Czech healthcare, health in general and health literacy, the National Cardiology Information System was published on the National Health Information Portal[10]. This platform provides comprehensive technical support for content editing and will also be used to update and add other relevant outputs. The open datasets have been published with all formal and technical requirements in the local catalogue of open data in healthcare, which is linked to the national catalogue[11].

DISCUSSION

The comprehensive agenda of the National Cardiovascular Health Plan, supported by information and data support from the National Cardiology Information System, aims not only at cardiology care but also at strengthening a healthy lifestyle and early detection programmes. A newly published data audit has confirmed the high availability and quality of cardiology care in the Czech Republic. Patient care is guaranteed by a dense network of inpatient and outpatient facilities, which creates a desirable three-level system covering all types of diseases according to risk and urgency. In the most important parameters, such as hospital mortality, long-term survival of patients or the frequency and length of hospital admissions, the Czech Republic ranks among the best countries in Europe. However, further improvement in the quality of care is complicated by the high number of cardiovascular patients, which is increasing due, among other things, to the population's poor health and the widespread non-compliance with preventive measures. The main added value of the NCIS is and will be the complete coverage of all dimensions and segments of care that are key for cardiology – from preventive programmes to end-of-life care. In addition to the positive indicators of quality of care mentioned above, the NCIS also reveals weaknesses in the organisation of care, which will be targeted by the National Cardiovascular Health Plan. In particular, there is low participation in prevention programmes.

EVALUATION OF THE AIMS OF THE CHAPTER

- To maximise the potential of using available data sources and incorporate them into the outputs of the National Cardiology Information System in close collaboration with representatives of medical expert societies.
  - The new National Cardiology Information System efficiently processes all available and relevant resources and compactly provides data reporting

[10] https://www.nzip.cz/nkis
[11] https://data.gov.cz/datov%C3%A9-sady

    across domains such as epidemiology, population burden, capacity and needs prediction, performance and quality indicators, economics and reimbursement of care, prevention, organisation and access to care.

- To design and publish a comprehensive set of outputs above the available data describing the current state of cardiology in the Czech Republic.
    - Under the guarantee of medical expert societies, it was possible to suitably compile typologically targeted outputs for individual groups of users so that they are maximally helpful in terms of information delivery. They always consider the need and objectives of the take-home message, contain an interpretation of the conclusions and are further expandable according to specific requirements from real practice.

LESSONS LEARNED

The National Cardiovascular Health Plan, under the leadership of an expert medical society, provided the essential background for developing the National Cardiology Information System. Data from components of the National Health Information System were appropriately supplemented by clinical registries in selected cases. Static and interactive reports, together with open data, fully cover all essential dimensions of assessment and segments of cardiology care, from preventive programmes to end-of-life care.

Open data as the basis for all published outputs has validated the "open data first" approach, whereby an open dataset is first designed and reviewed, and only then are further outputs prepared over this source. A major added value is the replicability of all descriptive and analytical outputs, and thus the acquisition of feedback from the general lay and informed public.

ACKNOWLEDGEMENTS

The authors of this case study would like to thank:

- the leadership of the Czech Society of Cardiology,
- the members of the analytical team and the webstudio of the Institute of Health Information and Statistics of the Czech Republic: Ondřej Šanca, Anna Klechová, Anna Zouharová, Kateřina Menšíková, Hana Kučerová, Vladimír Bartůněk, Hana Zelinková, Marek Lichter, Šárka Daňková, Michal Vysoudil, Jitka Jírová, Markéta Bartůňková, Tomáš Pavlík, Zbyněk Bortlíček,

Daniel Klimeš, Petr Klika, Svetlana Drábková, Michal Vičar, Michaela Kerberová, Vojtěch Bulhart.

REFERENCES

[1] Linhart A. The National Cardiovascular Plan 2023–2033. XXXI. Annual Meeting of the Czech Society of Cardiology. Brno, 13–16 May 2023.

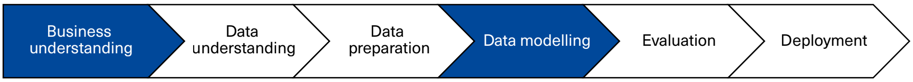
SECTION C

13

ADVERSE EVENTS REPORTING SYSTEM: DEVELOPMENT OF THE PLATFORM FOR HEALTHCARE PROVIDERS AND LAY PUBLIC

Andrea Pokorná, Dana Dolanová, Veronika Štrombachová, Denisa Macková, Petra Búřilová, Jan Mužík, Michal Pospíšil, Martin Komenda

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2020–now
Keywords	Adverse events, safety healthcare, management, methodology
Research question	How to create and implement a new tool supporting patient safety through the nation-wide strategy for adverse event control in inpatient healthcare facilities?
Type of result	Interactive visualisation
Level of data processing	Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

Reporting adverse events (hereinafter AE/AEs) is important for increasing the quality of care for patients. This is an international problem in all medical facilities that provide patient care. Many experts and researchers worldwide are dealing with this topic to find the best way not only for AE monitoring but especially for preventing harm to patients and promoting safety guidance and pathways used in clinical practice.

Healthcare-associated adverse events result in reduced patient safety and possible patient harm. The World Health Organization (hereinafter WHO) made an international classification for Patient Safety and published it in 2009 [1]. The WHO defines an AE as an incidence that occurs during medical care in healthcare facilities[1] [2]. AEs harm patients and can cause much damage, for example, injuries, suffering, pain, and disability, and can result in death [1,3–5]. The optimal quality of care is compromised by the increasing number of AEs, which are one of the most frequently occurring problems in healthcare facilities [3]. Because AEs increase the cost of treatment in healthcare facilities, it is essential to have strategies that can minimalise their impact, introduce policies aimed at combating AE, and support the development of a culture for patient safety [6].

The AE reporting system includes an adverse event, near miss, investigation, analysis, and feedback. The patient safety system is recommended, and healthcare providers are encouraged to report all AEs, near misses, or sentinel events to a hospital system and on the national level of the reporting system [7].

Systematic adverse events monitoring grounded under the Institute of Health Information and Statistics was built in 2014 as the Adverse Events Reporting System (hereinafter AERS) [8]. Initially, 80 inpatient healthcare providers began reporting and monitoring adverse events according to the AERS uniform methodology. Since 2018, the Adverse Event Reporting System has been included in the statistical survey programme of the Czech Republic. Decree No. 373/2017 Coll. of MZ CR (Ministry of Health of the Czech Republic) on the Programme of Statistical Surveys for 2018 imposes a reporting obligation to submit data on the number of adverse events reports to AERS to all inpatient healthcare facilities.

Statistical surveys of the Ministry of Health (hereinafter MZ CR) are part of the Programme of Statistical Surveys in the Czech Republic, compiled following Act No. 89/1995 Coll., on the National Statistical Service as amended. The central system for reporting AEs is the first system that monitors the reporting

[1] Categories of AEs adopted for the Czech National Adverse Event Reporting System (AERS): Pressure ulcers, Falls, People‘s behaviour, Accidents and unexpected injuries, Clinical performance/intervention, Technical problems, Medication / i.v. drugs, Unexpected deterioration in clinical condition, Medical devices / equipment, Clinical administration, Documentation, Sources/management of the organisation, Transfusions / blood derivatives, Diet / nutrition, Medicinal gases, Other AE.

of adverse events at all inpatient healthcare facilities in the Czech Republic. The prerequisite is the registration of AEs based on a uniform methodology, identification of risks, settlement of AE, and their systematic prevention using the potential of representatives of individual healthcare providers (hereinafter HCPs) and local know-how, through the preparation of uniform documents to rationally manage human and material resources (SHNU, 2023). The system is based on the historical, scientific development of patient safety, combining evidence-based resources and relevant literature with respect to the Czech legislation.

The main aim of the AERS is to collect and analyse data on the prevalence of AEs in inpatient HCPs in the Czech Republic and to establish effective preventive and corrective measures. The National AERS web portal[2] (hereinafter the NAERS portal), as a complex web platform, aims to provide a communication expert-oriented platform for the HCPs in identifying risks associated with providing health services, preventing adverse events, their recognition and classification, and uniform assessment at the local level to enable monitoring and reporting at the central level.

AIMS

- To systematically monitor the occurrence of adverse events (AEs) in clinical practice and subsequent data transmission to a central system based on developed methodological materials (e.g. guidance, video tutorials), which are available to the professional and lay public on the NAERS portal.
- To provide an expert-oriented communication platform for HCPs in identifying risks associated with providing healthcare services, preventing adverse events (AEs), their recognition and classification, and a uniform assessment at a local level to enable monitoring and reporting at the central level.
- To bring comprehensive analytical reports based on available data with the possibility of "benchmarking" in different categories of HCPs (internal data visualisation for authorised persons only).

METHODS

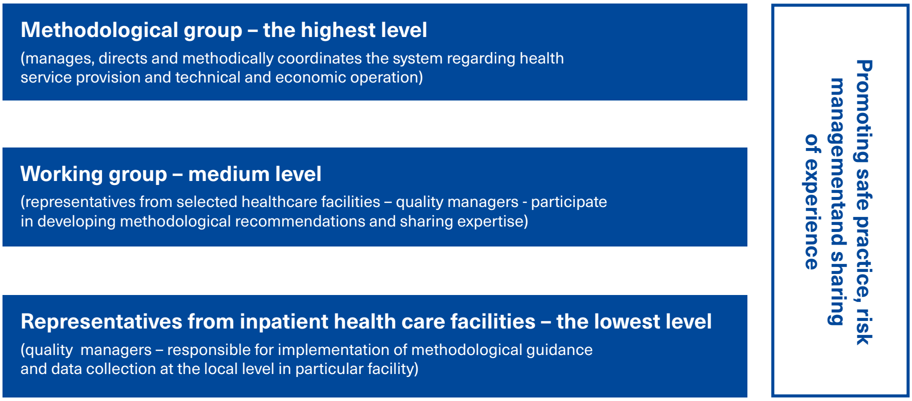
The methodology of monitoring AEs is based on the Recommendation of the Council of the European Union on patient safety, including the prevention and control of healthcare-associated infections (hereinafter HAI) of 9 June 2009 [9].

[2] https://shnu.uzis.cz/

The AERS records adverse events occurring in all inpatient care units based on a uniform methodology for identifying risks, dealing with AERs, and their systematic prevention, using the potential of representatives of individual inpatient care units and local know-how.

Within the AERS, methodological guidance are developed to assist healthcare professionals in clinical practice in interpreting the requirements for entering adverse event data at the central level. They include general information on describing and entering adverse events into local AE monitoring systems (for each type of AE). The AERS is designed as a rating system for evaluating anonymised aggregated data and comparing facilities with each other; moreover, it is a tool for unifying the process of AEs evaluation and identification based on the use of objectifying scales to assess patient status, enabling risk management, and provide a basis for the development of new recommendations for the prevention of adverse events according to uniform recommendations and safe practice.

AERS is established on close collaboration between methodological experts and health professionals from clinical practice, and AERS management is multilevel based (it includes three levels – see Figure 1).



Methodological group – the highest level
(manages, directs and methodically coordinates the system regarding health service provision and technical and economic operation)

Working group – medium level
(representatives from selected healthcare facilities – quality managers - participate in developing methodological recommendations and sharing expertise)

Representatives from inpatient health care facilities – the lowest level
(quality managers – responsible for implementation of methodological guidance and data collection at the local level in particular facility)

Promoting safe practice, risk managementand sharing of experience

Figure 1: Multilevel system of AERS coordination

The main coordinating body of the AERS (the highest level) is a methodological group consisting of experts with academic, research and clinical backgrounds, as well as health statisticians (biostatisticians). It manages, directs, and methodically coordinates the system regarding health service provision and technical and economic operation. The methodological group fulfils the purpose of methodological and organisational preparation and implementation of the Adverse Event Reporting System to ensure records of AEs affecting the quality of healthcare services and patient safety. Recording serious adverse events at the local level is the responsibility of individually authorised personnel in a particular healthcare facility. Responsibility for developing and updating methodological recommendations for monitoring and identifying risks and adverse events and implementing follow-up measures when an adverse event occurs is shared with a particular group of healthcare providers' representatives – the working group of AERS (medium level). The working group brings together representatives of individual HCPs who are locally focused on promoting the quality of care in healthcare facilities. The working group aims to participate in developing methodological recommendations and especially to share local know-how and expertise with all other stakeholders in AERS. Representatives from inpatient healthcare (the lowest level) providers must submit aggregated data on adverse events in each healthcare facility to the Adverse Event Reporting System. To cooperate with the methodological group, follow their instructions and recommendations, and improve the quality of care provided regarding notification of changes, recommendations, sharing of experience, etc. All staff of the HCPs has free access to methodological materials, and authorised representatives of the HCPs have access to internal documents available on the NAERS portal (i) As the AERS was newly developed and implemented at the national level, it was necessary to set up the basic principles in the development of methodological recommendations: creation based on recognised and recommended guidelines for the provision of health services issued not only domestically, but also and especially abroad Evidence-Based Practice (hereinafter EBP). (ii) Review based on recommended guidelines for providing health services, recommended working practices, and standards, enriched by own experience from practice. (iii) Implementation at the local level, considering Evidence-Based Healthcare (hereinafter EBHC) and respecting the local workplace capacities.

TECHNICAL BACKGROUND

To use the central Adverse Event Reporting System, it is necessary to have a local adverse event registration system in place (electronic or otherwise organizationally ensured and operated at the local level). Providers of inpatient healthcare services are obliged to provide the sending of aggregated data in the form of an electronic report L (MZ) 3-01 – on the number of AEs reports for central evaluation through the Central Repository of Reports, which is part of the Unified Technology Platform eREG.

Aggregate data collection is carried out once a year. The data are validated in two ways: (i) immediately at the time of submission of the L (MZ) 3-01 statement to eREG, the data are checked by the AERS´s methodologist (member of the Methodological group) at the time of approval of the statement; (ii) a second validation of the data is carried out by the AERS analyst (statistician) based on the established data-checking criteria. The validated data is then analysed and submitted for processing and publication by a special unit as part of the Institute of Health Information and Statistics – IHIS (so-called webstudio) to be presented graphically at the NAERS portal. Data on adverse events detected and reported are visualised in absolute numbers or as relative values („rates") per 1,000 patients. The AEs per 1,000 patients indicates how many AEs there would be if 1,000 patients were monitored in a given healthcare facility, thus allowing a comparison of the frequency of AEs in different groups of patients observed in concrete facilities.

RESULTS

BUSINESS UNDERSTANDING

The history of AERS is relatively short. Systematic monitoring of AEs was initiated at the national level in 2014, first in a pilot operation involving so-called directly managed hospitals by the Ministry of Health of the Czech Republic (MoH). With regard to the legislation in force, on 4 April 2017, the management of the MoH decided to include the monitoring of AEs in the Statistical Survey Programme of the Czech Republic for 2018. AERS is implemented as a prevalence data collection on the number of AE reports at the central (nationwide) level. Although much attention is paid to the actual data collection on the number of AEs, its primary and unquestionable objective is methodological cultivation (prevention and measures to increase patient safety and eliminate adverse events in clinical practice). The duty to monitor AEs results from the Decree on the Statistical Programme of Surveys for the relevant year (currently Decree No. 466/2020 Coll., on the Statistical Programme of Surveys for 2022). The adverse events "Falls "and "Pressure Ulcers" are mandatory monitoring values for all inpatient HCPs. Other AEs are monitored based on the decision of the management of individual inpatient care providers or based on the type of care (acute care, long-term care, etc.). As part of the methodological support for data collection, educational video tutorials have been created, focusing mainly on the correctness of completing the reporting template (L (MZ) 3-01). Over the years, methodological documents have been continuously developed, cultivated and updated, an overview of which is given in Table 1.

Table 1: Overview of AERS methodological outputs

Type of methodological materials	Year of production/ update
Methodological materials for individual adverse events. Instructions for data collection. (Bulletin of the Ministry of Health of the Czech Republic No. 7/2018)[3]	2018–2023
Taxonomic - definitional dictionary – the basic structure of terminology with explanatory terms and synonyms[4]	2018–2023
Methodological materials for inpatient care: Update of all methodological materials for each AE in two years (1× general methodology, 13× full methodology, 11× short methodology, 9× prevention algorithm, 7× corrective action algorithm)[5]	2017, 2019, 2021/2022
Methodological materials for the pilot project of Home Care Agencies[6]: General methodological materials (AEs Monitoring Methodology, Data Submission Instructions, Report Template). Methodological materials by the type of AE (falls, Pressure ulcers)	2017–2020

The methodological instructions are intended to assist healthcare professionals in clinical practice in uniformly interpreting the requirements for the entry of adverse event data at the local level and their reporting at the central level in a single aggregated form (based on unified terminology defined in the taxonomic dictionary). The main methodological document for the Adverse Event Reporting System is the Methodology for Monitoring Adverse Events in Inpatient Healthcare Providers, issued in Bulletin No. 7/2018 of the Ministry of Health of the Czech Republic. In the individual sections of the methodological materials are available: (i) methodologies including general information on the entry of adverse events – e.g. taxonomic dictionary, instructions for reporting aggregated data centrally to the Central Reporting Repository (CRR), (ii) methodological guidance with specific requirements for each main AE type. The required parameters that should be recorded are precisely described in the relevant sub-documents for ease of orientation and clarity.

The guidance documents on the main adverse events are implemented in several documents that are consistent in content and vary in scope:

1. **The full version of the methodological materials**, containing the following sections: definition of AEs; epidemiology - incidence and prevalence according to foreign sources and according to information from the central

[3] https://shnu.uzis.cz/cs/metodicke-materialy/obecna-metodika/
[4] https://shnu.uzis.cz/res/file/metodicke_dokumenty/shnu_taxonomie_2022_final_na_web.pdf
[5] https://shnu.uzis.cz/cs/metodicke-materialy/obecna-metodika/
[6] https://shnu.uzis.cz/cs/metodicke-materialy/domaci-zdravotni-pece/

reporting system, description of the items to be monitored and notes on their entry (explanation, description as in the taxonomy) and conclusion. The full version of the methodology is extensive. It should be available, especially to new entrants during the adaptation process and/or for workers returning after a long absence from work (e.g. after a long illness).

2. **Shortened version of the methodological materials**, containing the following sections: definition of AEs; epidemiology – incidence and prevalence according to international scientific sources and according to information from the central reporting system; checklist for checking preventive procedures before the occurrence of AE; checklist for checking immediate measures after the event. Its purpose is to provide clear information for rapid intervention. It should be kept at workplaces as an accessible document for quick reference in stressful situations. It is deliberately prepared concisely always to be available (e.g. sealed in foil and posted in the nurses' or doctors' office, examination room, or outpatient clinic).

3. **Algorithm of preventive procedures** related to a specific AE (Preventative algorithm) – a simple and clear tool for implementing preventative measures - again, it should always be available (e.g. closed in foil and posted in the nurses' or doctors' office, examination room, or outpatient clinic).

4. **Algorithm of immediate corrective actions** related to a specific AE (Corrective Action Algorithm) – a simple and clear tool for implementing corrective actions after the occurrence of an AEs – again, should always be available (e.g., closed in foil and posted in the nurses' or doctors' office, examination room, or outpatient clinic).

Methodological video tutorials (educational video tutorials) focused on the Adverse Event Reporting System concept, especially on completing the reporting template L (MZ) 3-01 on the number of adverse event reports for central evaluation. Educational videos are developed into individual sequences for better orientation in specific issues, but also as a more user-friendly format of methodological instructions, which visually and comprehensibly guide the user through the entire process of submitting data on the number of adverse events reports for central evaluation with the impact on minimising the input of erroneous data and the time burden for the representatives of the HCPs (Introduction to the Adverse event reporting system, the NAERS portal, Entering the registry and selecting/selecting the template L (MZ) 3-01 report, Completing and submitting the fulfilled template L (MZ) 3-01 report, Common errors in completing the template L (MZ) 3-01 report[7]).

DATA UNDERSTANDING

During its operation, the AERS has enabled the implementation of methodological documents in clinical practice and, of course, the collection of data on the number of AEs based on uniform internationally accepted terminology and methodology. It is a robust system with nationwide statistics. In total, 10,248,741 patients were followed up in AERS, and 528,827 AEs were reported between 2018 and 2022 (data for 2023 are currently unavailable as the data are collected for the previous finished year). The following table (Table 2) summarises the number of monitored patients in specific healthcare facilities and the number of reported adverse events for each year.

Table 2: Overview of the number of reported adverse events for the period 2018–2022

Year	Number of monitored patients (Pts)	Number of inpatient healthcare providers (HCPs)	Number of Adverse Events (AEs)
2018	2,706,998	408	105,509
2019	2,856,355	430	106,914
2020	2,320,850	435	101,030
2021	2,364,538	429	104,516
2022	2,545,319	426	110,858

We are aware that the increasing trend of reported AEs does not imply a poorer quality of care but a better ability of HCPs to recognise and report AEs. Since the interpretation of information on the number of AEs is challenging and can lead to misunderstanding of the results; the methodology team has been working on an online tool that would allow authorised persons of individual HCPs to analyse the data, aggregate them into categories of care providers, and also to capture trend data for each of the periods of interest, generate a summary report and develop the interactive visualisation system.

DATA PREPARATION

Aggregate data collection on adverse events at the central level is based on completed reports (reporting template: L (MZ) 3-01) from individual inpatient

[7] https://shnu.uzis.cz/cs/metodicke-materialy/obecna-metodika/

healthcare providers. The data are submitted annually in the structure following the categories of AEs:

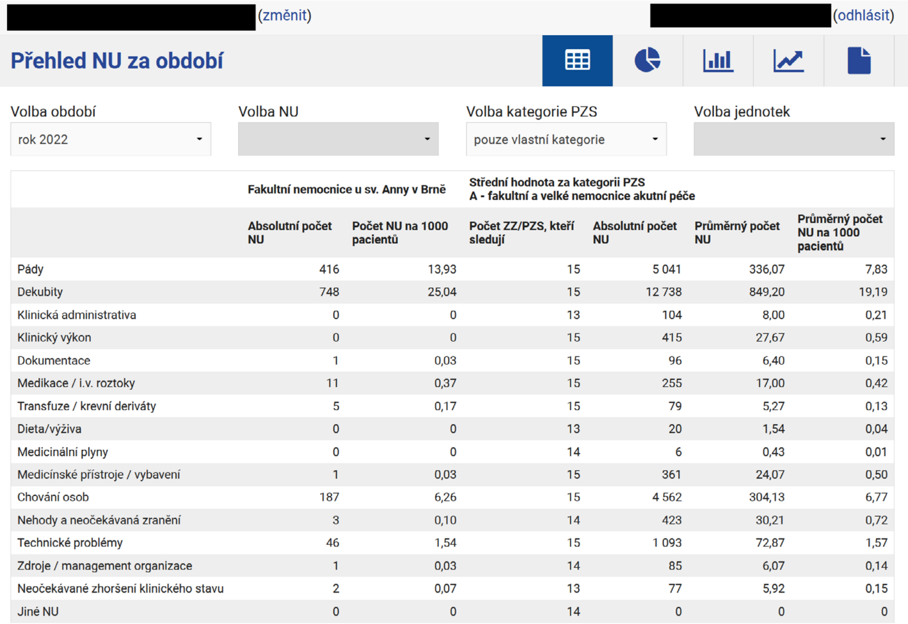
- Pressure ulcers
- Falls
- People's behaviour
- Accidents and unexpected injuries
- Clinical performance/intervention
- Technical problems
- Medication / i.v. drugs
- Unexpected deterioration in clinical condition
- Medical devices / equipment
- Clinical administration
- Documentation, Sources/management of the organisation
- Transfusions / blood derivatives
- Diet / nutrition
- Medicinal gases
- Other AEs

Methodology guidance and documents provide precise definitions for each adverse event category and learn whether it is a mandatory entry. An example is the definition of a fall-type adverse event: a patient accidentally falls (slumps) from his/her bed (or chair, wheelchair) to the floor. This is an unintended event where the person falls to the ground or another lower surface (a witness is present) or self-reports such an event (if it happened without witnesses). A situation caused by deliberate movement cannot be considered a fall [10]. In accordance with up-to-date data collection methodology, 154 AE characteristics (11 mandatory and 143 optional) were monitored in 2022 [11]. The collected data are then validated and transformed into a structure that allows for different views of adverse events, for example, by AE category, type of healthcare provider or period.

MODELLING

The internal system is available on the NAERS portal. The access is only for authorised persons with the mandate on behalf of the HCP who have access to data analysis for all AERS data collection periods (including the pilot period for counting the pilot data collection period in the case of the HCPs involved). The data are presented in tables and graphs with the possibility of different filters (all categories, custom categories, choice of period, choice of the absolute numbers of AEs / conversion of AEs per 1,000 patients). Each HCP-authorised person can generate a summary report according to their requirements for its content. There are several options for data interpretation and visualisation (see figures below). We are presenting all the data anonymously.

Figure 2 is an example of a summary table showing the values of a specific health facility/healthcare provider and the overall mean value in the same category of healthcare providers. The types of adverse events (health facility/healthcare provider-specific and total) are shown in absolute and relative numbers per 1,000 patients. The following figures include the black designation as it is necessary for the possibility of data anonymisation. The conversion of adverse events per 1,000 patients shows how many AEs there would be if 1,000 patients were followed up in a given HCP, thus allowing a comparison of the frequency of AEs in different groups of patients followed up.



[redacted] (změnit) [redacted] (odhlásit)

Přehled NU za období

Volba období: rok 2022

Volba NU:

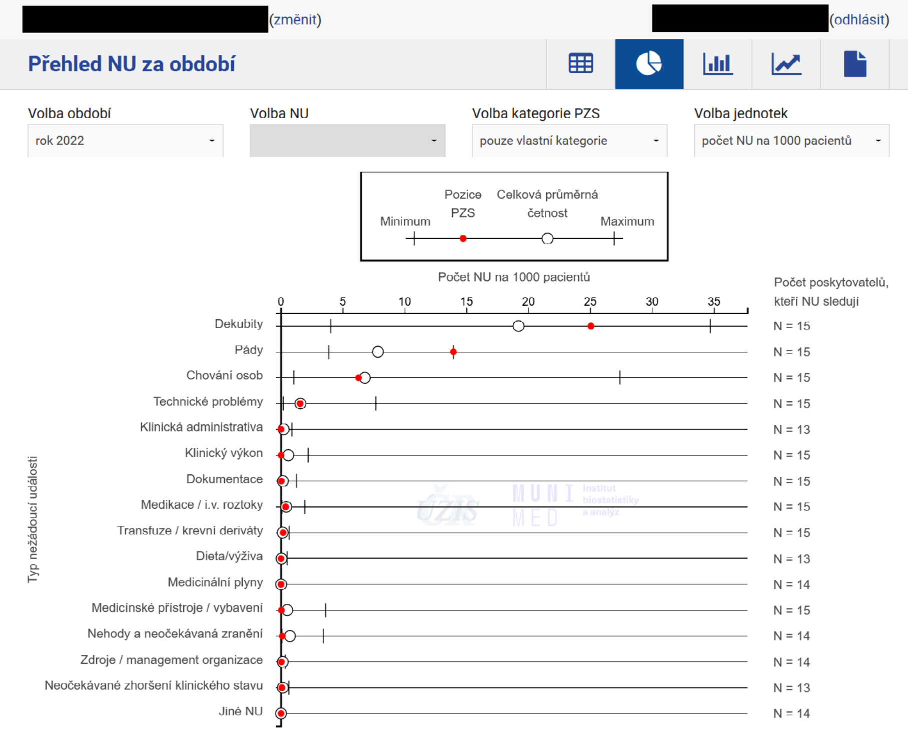
Volba kategorie PZS: pouze vlastní kategorie

Volba jednotek:

	Fakultní nemocnice u sv. Anny v Brně		Střední hodnota za kategorii PZS A - fakultní a velké nemocnice akutní péče			
	Absolutní počet NU	Počet NU na 1000 pacientů	Počet ZZ/PZS, kteří sledují	Absolutní počet NU	Průměrný počet NU	Průměrný počet NU na 1000 pacientů
Pády	416	13,93	15	5 041	336,07	7,83
Dekubity	748	25,04	15	12 738	849,20	19,19
Klinická administrativa	0	0	13	104	8,00	0,21
Klinický výkon	0	0	15	415	27,67	0,59
Dokumentace	1	0,03	15	96	6,40	0,15
Medikace / i.v. roztoky	11	0,37	15	255	17,00	0,42
Transfuze / krevní deriváty	5	0,17	15	79	5,27	0,13
Dieta/výživa	0	0	13	20	1,54	0,04
Medicinální plyny	0	0	14	6	0,43	0,01
Medicínské přístroje / vybavení	1	0,03	15	361	24,07	0,50
Chování osob	187	6,26	15	4 562	304,13	6,77
Nehody a neočekávaná zranění	3	0,10	14	423	30,21	0,72
Technické problémy	46	1,54	15	1 093	72,87	1,57
Zdroje / management organizace	1	0,03	14	85	6,07	0,14
Neočekávané zhoršení klinického stavu	2	0,07	13	77	5,92	0,15
Jiné NU	0	0	14	0	0	0

Figure 2: Summary overview of AEs reported by a particular HCP in a specified period (only available in Czech)

Figure 3 graphically displays the adverse event values of a specific healthcare facility/healthcare provider compared to the overall mean (overall average AE per HCP), minimum and maximum values in the same category (there could also be selected all HCPs categories) of healthcare providers.



[redacted] (změnit) [redacted] (odhlásit)

Přehled NU za období

Volba období: rok 2022

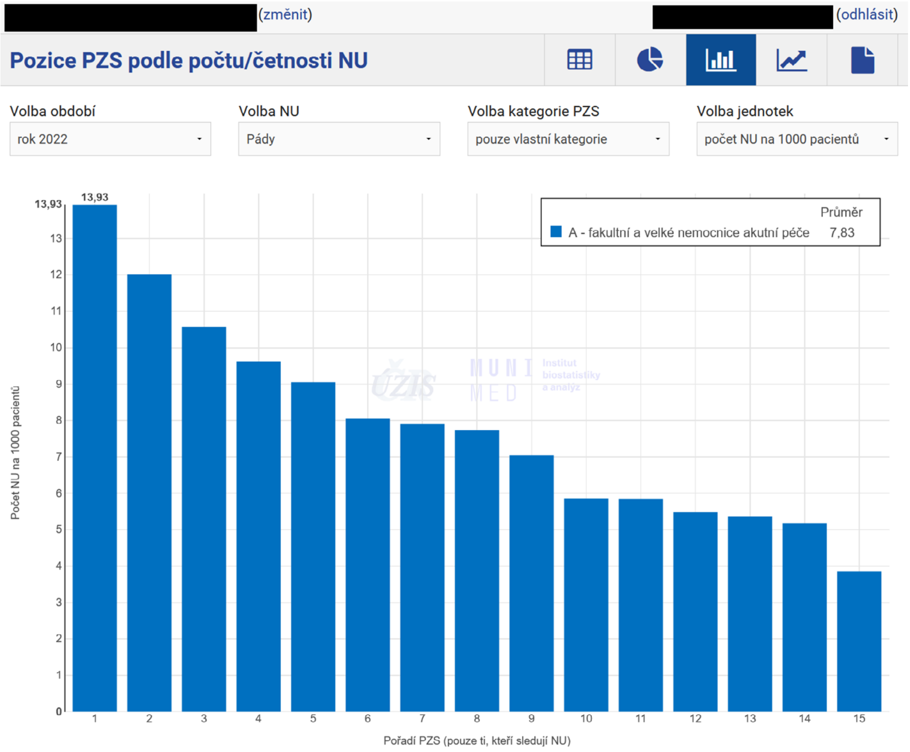
Volba NU:

Volba kategorie PZS: pouze vlastní kategorie

Volba jednotek: počet NU na 1000 pacientů

Figure 3: Summary overview of AEs reported by a particular HCP in a specified period – graph (only available in Czech)

Figure 4 shows the values of a selected adverse event of a specific healthcare facility/healthcare provider compared to the values of other HCPs in all categories of HCPs (different filters can be selected). This graph focuses on Falls in the categories of faculty hospitals and large hospitals; an average number of 7.83 per 1,000 patients is shown.



[redacted] (změnit) [redacted] (odhlásit)

Pozice PZS podle počtu/četnosti NU

Volba období: rok 2022

Volba NU: Pády

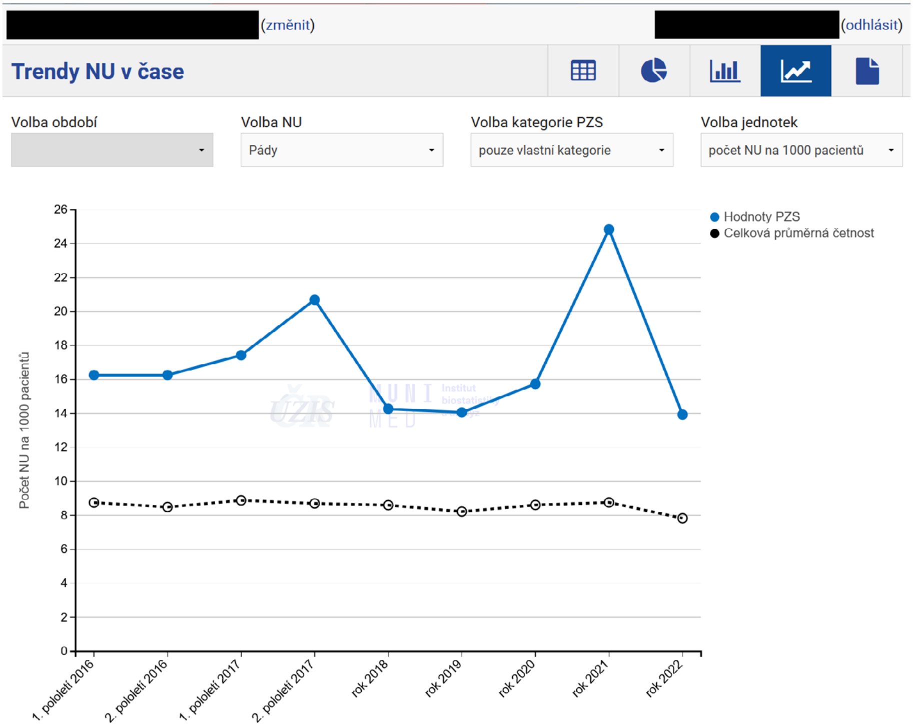
Volba kategorie PZS: pouze vlastní kategorie

Volba jednotek: počet NU na 1000 pacientů

13,93
13,93
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Počet NU na 1000 pacientů
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Pořadí PZS (pouze ti, kteří sledují NU)
Průměr
A - fakultní a velké nemocnice akutní péče 7,83
ÚZIS
MUNI MED
Institut biostatistiky a analýz

Figure 4: Position of a particular HCP by the number of AEs – with the filter "only the HCP's own category" applied (only available in Czech)

As individual inpatient healthcare providers also need to know continuous values and trends over the past period to compare the evolution of AEs prevalence, there is the possibility of filtering the data over a multi-year period.



[redacted] (změnit) [redacted] (odhlásit)

Trendy NU v čase

Volba období:

Volba NU: Pády

Volba kategorie PZS: pouze vlastní kategorie

Volba jednotek: počet NU na 1000 pacientů

Figure 5: AE trends over time - a particular HCP, with the filter "only the HCP's own category" applied (only available in Czech)

Figure 5 graphically displays the evolution of the values (trend) of the selected adverse event of a specific healthcare facility/healthcare provider over time compared with the overall mean values (overall average frequency of AEs per 1000 patients) in the same category of healthcare providers. The blue line presents the values of concrete-selected HCPs, and the black line represents the average frequency. The figure shows AE Falls with the data since 2016, as this particular HCP was involved in the pilot phase.

The analytical output of the data transmitted through the report (reporting template) L (MZ) 3-01 may also be a data summary report. Each healthcare provider can generate a summary data report in the internal part of the website. The data summary report is an output of the analytical tool of the individualised analyses of the Adverse Event Reporting System managed by the Institute of Health Information and Statistics of the Czech Republic. The purpose of the report is to provide a specific healthcare facility (or a specific healthcare provider) with comprehensive information on the number and frequency of adverse events reported to the system and their status, among other HCPs involved in monitoring adverse events within the AERS.

Monitoring adverse events is an important policy issue: the safety of the care provided is essential not only for governing bodies and care providers themselves but also for patients. For this reason, anonymised AE count outputs have been prepared for the general public. The user has access to anonymised data on the number of reported AEs with the possibility of using various functionalities (overview of AEs of all categories of HCPs, Trends of AEs over time, Position of categories of HCP by number of AEs, Comparison by type of AEs/HCPs). It is necessary to stress to the lay users that the reported AE counts depend on the quality of AE reporting. The data presented are not for comparison or even for the creation of "rankings" of quality of care and risks at the national level but mainly for the cultivation of processes to support the quality of care at the local level of individual healthcare providers and the ability to monitor the risks of care. Several options are similar for display in the internal part of the interactive visualisation for the lay population but always in an aggregated and anonymous form[8]. The publicly available data should increase awareness of potentially risky areas in healthcare provision among the lay public and motivate them to find more information about the possibility of preventing an adverse event occurrence. We hope that through this publicly available data and methodological materials, we could also positively influence the lay population's health literacy. This part of the NAERS portal is also intended to serve future healthcare professionals to find interesting information about adverse events.

EVALUATION

To ensure valid data and correct interpretation, all descriptive statistics and analytical outputs in the form of static reports (PDF files) and interactive web visualisations (the NAERS portal) are always thoroughly checked by a team of senior analysts, the portal development team and a team of methodologist and expert guarantors. Validation by the methodology group is based on mutual interaction (mostly phone consultations and online meetings), serving not only to refine the guidelines for data submission but also to check the consistency of the submitted data compared to previous periods, especially in the area of staff capacities, which are reported and important for subsequent detailed analyses of the prevalence of AEs. Another validation mechanism is the comparison with staff capacity reports in the form of economic reports and data also submitted for HCPs on an annual basis. In the case of significant deviations from previous years, the change in case mix of patients, change in staff structure, change

[8] https://shnu.uzis.cz/cs/analyzy/

in providers structure and change in focus of services provided are validated (Staffing Statements E (MZ) 2-01, E (MZ) 3-01, E (MZ) 4-01). The primary goal is to minimise errors and misunderstandings in the data and to present the conclusions clearly in accordance with the methodology.

DEPLOYMENT

Outputs from the Adverse Event Reporting System are mainly disseminated through the NAERS portal. Within the development cycle, deployment is the last phase, preceded by in-depth manual and automated testing and optimisation of the new releases in the development and stage server environment. The IT infrastructure of the Institute of Health Information and Statistics provides a full-fledged background for the portal's continuous development, integration and evolution according to the AERS methodological team's current requirements. In particular, interactive visualisations and open data domains are constantly improving.

DISCUSSION

The Adverse Event Reporting System is a new approach to patient safety and quality of care improvement in the Czech Republic, which has been developed over the last eight years. The system is founded on scientific evidence-based resources in light of actual clinical practice situations and the current legislation. As a part of the system, an online repository of methodological materials (written guidance, algorithm, and video guides) and a storehouse of long-term monitored data about the adverse event prevalence within inpatient healthcare facilities were built. The data are available both for internal use to improve the quality of care and safety regulation at the local level for particular healthcare providers as anonymous data for the lay public in accordance with the requirements of the MZ CR central governing bodies and in relation to evidence-based healthcare.

The adverse event reporting system certainly contributes to quality improvement and still has potential for future development. In the future, it is planned to expand the adverse event reporting to other AE categories, such as HAIs and work-related injuries in health professionals, improve and enrich the possibilities of interactive data visualisations, and raise awareness of AEs among the general public.

EVALUATION OF THE AIMS OF THE CHAPTER

- To systematically monitor the occurrence of adverse events (AEs) in clinical practice and subsequent data transmission to a central system based on developed methodological materials (e.g. guidance, video tutorials), which are available to the professional and lay public on the NAERS portal[9].
    - Coordination, organisation, and provision of administrative and management activities, especially in monitoring the occurrence of adverse events (AEs) in clinical practice and solving partial research and development tasks concerning the quality of patient safety care has been done successfully. Methodological procedures to support the quality of care provided - the creation of methodological materials (methodologies, guidance, video guides) available to the professional and lay public on the NAERS portal4, methodological support for reporting of AERS according to uniform terminology, strengthening of educational processes of target groups of health care workers and application of preventive measures were developed. The activities aim is mainly to improve the quality of care and the safety of services (e.g. by focusing on "nursing-sensitive" areas and epidemiology of selected disease states and diseases with impact on the quality of care and safety of services provided).
- To provide an expert-oriented communication platform for HCPs in identifying risks associated with providing healthcare services, preventing adverse events, their recognition and classification, and a uniform assessment at a local level to enable monitoring and reporting at the central level.
    - Information campaigns for target groups have been provided - operation of the NAERS portal as a communication expert-oriented platform for HCPs in the identification of risks associated with the provision of health services, prevention of adverse events, their recognition and classification, and uniform assessment at the local level, which will enable monitoring and reporting at the central level. Various methodological materials and results of individual data collection analyses in AERS and other project activities to promote quality of care have been regularly published and presented.

[9] www.shnu.uzis.cz

- To bring comprehensive analytical reports based on available data with the possibility of “benchmarking” in different categories of HCPs (internal data visualisation for authorised persons only).
    - A complex NAERS online portal has been developed for data collection among all involved/participating healthcare providers and especially for methodological support. It includes static analytical reports, dynamic and interactive data visualisations and benchmarking among healthcare providers in the Czech Republic. Moreover, the AERS methodological team ensures and coordinates strategic analyses to evaluate the quality of health care (analysis of data in AERS with the possibility of “benchmarking” between the different categories of HCP - internally available data - internal data visualisation, analysis of personal and staff capacity indicators of the healthcare system of the Czech Republic, further cooperation on the analysis of NHS data with a focus on the promotion and protection of public health with a special emphasis on the evaluation of the quality of care, identification of errors in the so-called “quality of care”, and the analysis of the quality of care.

LESSONS LEARNED

The central system is conceived as an evaluation system serving primarily: (i) to enable shared learning of individual healthcare providers in the unified method of risk identification, evaluation, and settlement of AEs, (ii) to evaluate anonymised aggregated data on reported main types of AEs, (iii) to compare devices with each other. The work of the Adverse Event Reporting System team (and the system itself) primarily ensures that AEs are recorded and processed not only at the local level (i.e., in a given HCP). The function of the system is not only to provide the results of data analysis but also to investigate them (causes and consequences) to help improve healthcare at the local HCP level as well as the nationwide level and to help health professionals provide care more safely, based on the purposeful exchange of experience and expert knowledge.

Benefits for HCPs:

- Collecting data on the number of reported AEs is particularly important from the point of view of AE prevention and improving quality of care in clinical practice.
- Correct identification of AE and appropriate categorisation of AE is also essential (beware of confusion between cause and consequence of AE).

- The importance of continuous methodological support from the AERS team.
- The usefulness of interactive visualisations by individual HCPs – positive feedback from HCP representatives.

Benefits for the public:

- Publicly available data enables information sharing/problem-solving and peer learning between providers.

REFERENCES

[1] World Health Organization & WHO Patient Safety. Conceptual framework for the international classification for patient safety version 1.1: final technical report January 2009 [Internet]. World Health Organization; 2010 [cited 26 Jul 2023]. Available from: https://apps.who.int/iris/handle/10665/70882.

[2] World Health Organization. Patient safety [Internet]. World Health Organization; 9 March 2019 [cited 26 Jul 2023]. Available from: https://www.who.int/news-room/facts-in-pictures/detail/patient-safety.

[3] San Jose-Saras D, Valencia-Martín JL, Vicente-Guijarro J, et al. Adverse events: an expensive and avoidable hospital problem. Ann Med. 2022;54(1):3157–68.

[4] Kruk ME, Gage AD, Joseph NT, et al. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. Lancet. 2018;392(10160):2203–12.

[5] Makary MA, Daniel M. Medical error – the third leading cause of death in the US. BMJ. 2016;353:i2139.

[6] Al-Mugheed K, Bayraktar N, Al-Bsheish M, et al. Patient Safety Attitudes among Doctors and Nurses: Associations with Workload, Adverse Events, Experience. Healthcare (Basel). 2022;10(4):631.

[7] Fujita S, Seto K, Hatakeyama Y, et al. Patient safety management systems and activities related to promoting voluntary in-hospital reporting and mandatory national-level reporting for patient safety issues: A cross-sectional study. PLoS One. 2021;16(7):e0255329.

[8] Pokorná A, Štrombachová V, Mužík J, et al. National Portal of Adverse Events Reporting System [Internet]. Prague: Institute of Health Information and Statistics of the Czech Republic; 2016 [cited 20 Jul 2023]. Available from: https://shnu.uzis.cz.

[9] Council of the European Union. Council Recommendation of 9 June 2009 on patient safety, including the prevention and control of healthcare associated infections (2009/C 151/01) [Internet]. Council of the European Union; 2009 [cited 26 Jul 2023]. Available from: https://eur-lex.europa.eu/LexUriServ/%20 LexUriServ.do?uri=OJ:C:2009:151:0001:0006:EN:PDF.

[10] Pokorná A, Mužík J, Štrombachová V, et al. Aggregated data collection on adverse events on a central level: guidelines for data reporting [Internet]. Prague: Institute of Health Information and Statistics of the Czech Republic; 2022 [cited 22 Jul 2023]. Available from: https://shnu.uzis.cz/res/file/metodicke_dokumenty/pokyny_pro_predavani_dat_l301_help_23_verze2022.pdf.

[11] Adverse events in 2022 [Internet]. Prague: Institute of Health Information and Statistics of the Czech Republic; 2023 [cited 22 Jul 2023]. Available from: https://shnu.uzis.cz/res/file/analyzy/shnu_data_2022_vysledky.pdf.

SECTION C

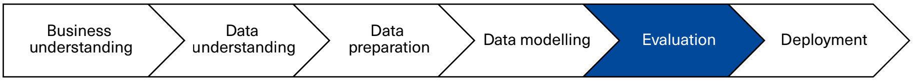
| 14 |

NATIONAL HEALTH INFORMATION PORTAL: TEXT-BASED OPEN DATA

Martin Komenda, Štěpán Svačina, Bohumil Seifert, Barbora Macková, Alena Šteflová, Milan Blaha, Vojtěch Bulhart, Ladislav Dušek

The case study was supported by the following grant project: Raising awareness through the National Health Information Portal, Operational Programme Employment (Reg. No. CZ.03.4.74/0.0/0.0/15_025/0016090), which was funded by the European Social Fund and the national (state) budget of the Czech Republic.

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2020–now
Keywords	National Health Information Portal, health literacy, guaranteed information, open data
Research question	How can selected parts of the National Health Information Portal content be published as open data for further text data analysis?
Type of result	Complex web application
Level of data processing	Descriptive statistics, Open datasets

DATA TO DOWNLOAD

INTRODUCTION

The National Health Information Portal[1] (NZIP) has a long-term goal of ensuring reliable communication between individual actors in healthcare, citizens – patients, public administration representatives and expert medical societies: (i) public administration communicates necessary information from the Czech healthcare to the public through NZIP, (ii) the public provides feedback through NZIP on which topics are not sufficiently covered by information, including evaluation of the quality of already published content in the form of articles, external sources and index terms. An integral part of NZIP is an interactive healthcare map where the nearest provider can be located, facilitating the connection of potential patients with health problems to their doctor/specialist. The main guarantors of the content are the Ministry of Health of the Czech Republic, the National Institute of Public Health, the Czech Medical Association of J. E. Purkyne, and the Institute of Health Information and Statistics of the Czech Republic. The long-term functioning of NZIP is supported by an extensive team, which brings together expertise across medicine, informatics and marketing (management, development team, author team, guarantor team, review team, analytics team, data team, promotion and PR team).

NZIP was launched on 23 July 2020, when a press conference was held in collaboration with the Ministry of Health of the Czech Republic. As of 22 May 2023, the portal contained 1,425 descriptive articles, 327 external recommended resources, 5,608 index terms and 39,469 records on healthcare professionals or healthcare facilities. The content of NZIP is divided into basic sections on selected life situations, prevention and healthy lifestyle, information on diseases, and recommended websites. All of these sections provide relevant and guaranteed information from Czech healthcare.

As part of the unification of published information from the data reporting area based on the National Health Information System (NHIS), a completely new section is currently being prepared on NZIP, where major reports, statistical summaries, analytical reports and news will be published and referred to. Departmental reference statistics and open datasets from specific NHIS registries, which are continuously published in the National Catalogue of Open Data[2], will be an integral part of this. In line with the overall national open data strategy, the health sector has moved to a new open data catalogue platform with full support for the current DCAT-AP standard [1], including the interconnection with NZIP in the form of a completely new module.

[1] https://www.nzip.cz

[2] https://data.mzcr.cz/

The primary ambition of NZIP is to avoid information fragmentation and ensure correct interpretation of the guaranteed materials through a single communication platform. This is primarily intended for promoting and using health services in line with the strategic health promotion and strategic objectives of the Ministry of Health of the Czech Republic. NZIP systematically and over the long term supports the increase in the confidence of the general and professional public in health information, emphasises the importance of health literacy of the population and facilitates access to available online resources across the board, thus contributing, among other things, to the promotion of healthy lifestyles and preventive programmes guaranteed by the state.

AIMS

- To understand the role and goals of the National Health Information Portal (NZIP) in continuous health literacy.
- To distinguish different types of published content concerning the design of open datasets.
- To design the structure and content of open datasets to maximise information value.

METHODS

TARGET GROUPS

The target groups include both the general public and the informed public, i.e. citizens of the Czech Republic in the broadest sense of the word, interested in health and healthcare information. Since the topic of health and healthcare is relevant at every age, the project will affect almost all citizens of the Czech Republic. The assumption is that NZIP will become a valid and transparent source of information on a healthy lifestyle, healthcare procedures and methods, the network of healthcare facilities and their quality parameters, the possibilities of health protection and promotion, prevention in general (as well as specific prevention programmes), diseases, care programmes for the chronically ill, etc. Within the individual sections of NZIP, targeted and addressed support is provided in particular to groups of persons at increased risk, such as unvaccinated children and their parents, chronically ill patients, persons with rare diseases, persons with low health literacy, persons threatened by specific epidemiological situations, or persons requiring palliative care. This list is not exhaustive and continuously expands based on the particular needs of the population in question. In addition, representatives of individual patient organisations, which bring together patients or family members of patients suffering from particular illnesses and defend their interests, including the right to information and accessible quality healthcare, are a well-defined subgroup of the target population.

Another key group is made up of the organisational units of the state and their employees, in particular the Ministry of Health of the Czech Republic (primarily the Department of Public Relations, the Section of Public Health Protection and Promotion, as well as the Section of Healthcare, namely the Department of Healthcare and the Department of the National eHealth Centre). These people will offer a platform to founders of healthcare facilities (regions, municipalities etc.) that individual facilities can use to reach out to the public. All published content will be fully guaranteed and therefore entirely trustworthy. NZIP also adds significant value to other ministries outside the health sector, such as the Ministry of Education, which plays a crucial role in health education and awareness. Local authorities and their employees, as major healthcare facilities founders and healthcare providers in the Czech Republic, should not be overlooked. Here we are talking mainly about regional authorities' employees from the region's healthcare administration. Important players are also state contributory organisations through directly managed organisations of the Ministry of Health of the Czech Republic and their employees involved in the management and PR and the public communication departments, regional public health authorities (in individual regions) and institutes of public health (especially the National Institute of Public Health).

RECOMMENDATIONS FOR AUTHORS

The articles for NZIP always respect the agreed thematic delivery plan, which is set quarterly by the Steering Committee. This plan respects an official opinion of the Ministry of Health of the Czech Republic intended for the public, which is supported by the author organisation (e.g. the National Institute of Public Health, the Czech Medical Association of J. E. Purkyne, etc.). When preparing the content, the authors and the editorial team proceed from a generally accepted source text, e.g. a decree, a valid clinical guideline or a peer-reviewed publication. The NZIP articles are not of the nature of professional publications. The author is assumed to be familiar with the information due to his/her professional background. The articles are prepared for the general public, so it is desirable to avoid drawing attention to external websites too often. In an attempt to maximise the uniformity of the author's submissions, which are then editorially processed into a form that can be published on NZIP, a methodology has been developed that includes, among other things, basic guidelines for expert guarantors and content creators:

1. Do not use foreign words or words taken into Czech from foreign languages (including the overused term "implement"). If you cannot avoid foreign words, explain them in parentheses after the foreign word.

2. Use short and simple sentences.

3. Do not use hyphens excessively. They sometimes enrich the text, but it does not look good if there are too many.

4. When expressing intervals (e.g. "400–500 eggs"), use the word "up to" instead of a hyphen.

5. You can use pictures to illustrate the text. However, you should always indicate the source of the image.

6. Omit metaphors, abstract concepts and abbreviations that may cause reading difficulties or confusion.

7. Work with those for whom the text is intended (Do they understand it? How would they write it?).

8. Try to avoid piling up several nouns in a row and try to rephrase the sentence, for example by replacing the noun with a verb. This will not only improve the text's stylistics but also the reader's clarity.

9. Try to avoid the term "person(s)". Your texts will be read by the general public, not by a small group of officials. For example, use the terms „people", "patients" etc. (depending on the context).

10. Limit the use of question marks in titles and subtitles. Use them at most 1 or 2 times for the whole text, preferably not at all. Make subtitles factual, and do not use foreign words in them. Specific example: ~~*How are children abused?*~~ Suggestion for a better subtitle: *Types of abuse*

11. Bullet points enrich the text, but do not use them throughout the text. This applies especially if bullet points are followed by complete sentences or clauses that normally follow. Bullet points are intended to catch the reader's attention. However, when used throughout the text, they are counterproductive.

12. If you provide links, ensure they actually work and direct the reader to the online resource.

Together with these rules, the NZIP team wants to ensure that the content is understandable to all interested visitor groups. Therefore, the NZIP team aims to gradually prepare alternative versions of selected articles for easy reading [2], which will be provided in a form that can be understood by people with intellectual disabilities, cognitive impairments etc.

TECHNICAL BACKGROUND

The design, development and implementation of NZIP follow good, proven and recommended practices in the field of web applications. It is based on PostgreSQL[3] – an open object-relational database system, which is contributed by a large global community of developers and IT companies. For the development itself, the Symfony web application framework[4] was chosen, designed primarily for PHP web applications. Modern technologies make the platform as user-friendly and efficient as possible. An example is the JavaScript framework React for creating responsive and interactive user interfaces. Attention was also paid to optimising the app's performance, including UX/UI design to make it accessible and easy to use across target audiences. As part of agile development, new functionality or entire modules incorporating a more extensive set of new features are added incrementally, depending on the priorities and specification of requirements. The development life cycle can be described in the following stages:

1. Analysis of the current state (needs survey and prioritisation)

2. Specification of requirements (functional and non-functional requirements, resource allocation)

3. Modelling (use case diagrams, entity relationship diagram)

[3] https://www.postgresql.org/

[4] https://symfony.com/

4. Static prototype (wireframes and graphical processing of individual components)

5. Functional prototype (interactive behaviour of individual components within a full-featured website)

6. Implementation (connection to real data and portal platform on development infrastructure)

7. Testing

8. Deployment (release of the new version on the production server)

9. Management and sustainability

10. Evaluation of user behaviour (and possible optimisation)

RESULTS

BUSINESS UNDERSTANDING

NZIP is built on a proprietary content management system that fully supports the proposed process of content publishing (Figure 1). The cornerstone is comprehensive and individual methodological support towards authors, clearly defined rules for editorial changes and a multi-step internal review mechanism before publication. Original articles must always be comprehensible and structured according to the needs of the reader, who can be from either the general or informed public.

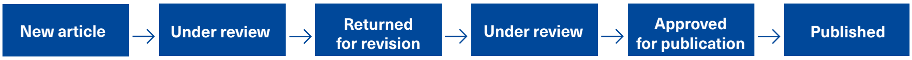


Figure 1: Scheme of article publication on NZIP

For the purposes of subsequent analytical processing of the available NZIP content, the format of individual datasets was designed to ensure that as much information as possible was made publicly available as part of the data opening strategy. The division of content into sections and the subsequent two-level tree structure allows articles and external sources to be clearly categorised with respect to the supporting topic.

DATA UNDERSTANDING

The data structure itself is not too complicated. The content is divided into three basic types: (i) article, (ii) external source, (iii) index term. There is always a set of metadata and the textual content itself. To ensure maximum accessibility, comprehension, and readability of the contribution according to the type of NZIP content published, the following open datasets have been proposed:

- Overview of categories
- Overview of original articles
- Overview of recommended external sources
- Overview of index terms

A simple mapping of specific records from the above sets makes it possible to link available content. For the purpose of textual analysis, descriptive attributes of individual records (such as title or description) are useful.

DATA PREPARATION

From the primary database, several SQL statements were used to retrieve files that describe the type of contribution. Complex transformations and mappings were not needed in this particular case. This is due to the relatively simple and complex database schema. These datasets do not include sensitive information such as the author's name; all available descriptive attributes are completely unproblematic from the perspective of personal data protection.

MODELLING

The following open datasets are designed to carry as much information value as possible. In addition to the title and text content itself, categorisations are available for the creation of global summaries or deeper analytical reports over individual modules, categories or subcategories. The individual datasets, a brief description and a summary of their descriptive characteristics are listed below.

Overview of categories: The dataset contains a list of all categories of articles from the National Health Information Portal (nzip.cz). It also contains the number of articles and sources in each category.

- The unique category identifier
- Category name

- Unique identifier of the parent category
- Parent category name
- Total number of articles in the category
- Total number of sources in the category

Overview of original articles: The dataset contains an overview of all articles published on the National Health Information Portal (nzip.cz). Each article is listed with its author (umbrella institution), title, category, date of publication, and date of last modification.

- Unique identifier of the article
- Title of the article
- Complete content of the article
- Guarantor (institution)
- Name of the parent category of the article (parent categories link categories into broader units)
- Unique identifier of the parent category of the article
- Name of the article category
- Unique identifier of the article category
- Flag indicating whether the article includes video
- Flag indicating whether the article is in an easy-to-read version
- Unique identifier to the original version of the article
- Length of the article (the total number of characters that the article contains)
- Date on which the article was published
- Date on which the article was last edited

Overview of recommended external sources: The dataset contains an overview of all external sources linked to the National Health Information Portal (nzip.cz). Each external source is listed with its author (umbrella institution), title, category and date of publication and last modification.

- Unique identifier of the external source
- Title of the external source
- Brief summary of the content of the external source

- Guarantor (institution)
- Name of the parent category of the external source (parent categories link categories into broader units)
- Unique identifier of the parent category of the external source
- Name of the external source category
- Unique identifier of the external source category
- Length of the external source's brief summary (the total number of characters contained in the brief summary of the content of the external source)
- Date on which the external source was published
- Date on which the external source was last modified

Overview of index terms: The dataset contains an overview of all index terms listed on the National Health Information Portal (nzip.cz). Each of the index terms is explained.

- Unique identifier of the index term
- Title of the index term
- Explanation of the index term
- Length of the index term (the total number of characters that the index term contains)

EVALUATION

The control of datasets designed and filled in the above-described manner was performed on several levels. First, a summary check was made on the total numbers of each type of contribution. As a second step, specific records (individual rows) and available content on NZIP were checked. Categories that had only a navigational or orientation role, but had no real content in the form of an article or an external source, proved problematic. However, the overall concept and logic of the NZIP information architecture (menu structure, individual items and related content) also justifies this anomaly in the data, although it is not entirely understandable at first glance.

DEPLOYMENT

Similar to the systematic approach to the design and development of NZIP, the NZIP open data were also prepared systematically. The process of dataset generation was first run on the development server; after successful completion and internal validation, the same was done on the production server. These outputs are available within the local open data catalogue and are updated annually.

DISCUSSION

The National Health Information Portal is the primary source of guaranteed and comprehensible information across the Czech healthcare system. It is unquestionably one of the strategic priorities for communication from professionals to the general and informed public. Together with open data from the National Health Information System, it is a source of valid and correctly interpreted information. The meaningful opening of data for machine-readable use within NZIP is also a further step towards openness. As new modules are implemented (and thus logically new content, such as games), new open datasets will also be added. The main motivation and goal of NZIP is to provide guaranteed content in accordance with current legislation for further descriptive or analytical tasks.

EVALUATION OF THE AIMS OF THE CHAPTER

- To understand the role and goals of the National Health Information Portal in continuous health literacy.
    - The main objective is to promote the development of health literacy through comprehensible content. The basic division into main sections (life situations, prevention and healthy lifestyle, information on diseases, recommended websites and data reporting) covers basic information in the form of articles, explanatory index terms and additional external sources.
- To distinguish different types of published content concerning the design of open datasets.
    - Distinguishing individual datasets according to the type and main categories of NZIP allows not only basic descriptive characterisation, but also preparation of complex analytical reports, where it is possible to work, for example, with the frequency of occurrence of selected keywords.

- To design the structure and content of open datasets to maximise information value.
  - The format and content of the above mentioned datasets were primarily inspired by several requirements for subsequent machine processing of published content on NZIP. The proposed outputs contain all possible descriptive attributes, with the exception of information that would interfere with personal data protection.

LESSONS LEARNED

The information published on the National Health Information Portal in the form of static articles, external sources and index terms has an exceptional added value. All texts are original and provide scope for further descriptive and analytical insights. The open data sets, which will be updated on an annual basis, provide structured textual data with the ambition of extracting keywords and terms and then linking them to other related domains such as education. Data on the last modification will allow retrospective tracking of changes and updates made by authors in view of the possibility of the published content becoming obsolete.

The collaboration of the umbrella institutions and technical teams has not only made it possible to build a comprehensive and unified information base for health literacy development but also contributed to the publication of text data in an open format for further processing. Contrary to expectations, the evaluation was a time-consuming phase, during which several iterations occurred in structuring the datasets.

REFERENCES

[1] Klímek J. DCAT-AP representation of Czech National Open Data Catalog and its impact. J Web Semant. 2019;55:69–85.

[2] Nomura M, Skat Nielsen G, Tronbacke B. Guidelines for easy-to-read materials. International Federation of Library Associations and Institutions (IFLA); 2010 ISBN 9789077897423.

SECTION C

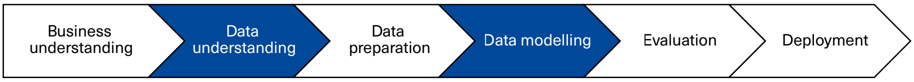
| 15 |

INNOVATIONS IN PATIENT-CENTRED CANCER CARE: ONLINE BENCHMARKING TOOL

Martin Komenda, Tit Albreht, Marek Svoboda, Ondřej Májek, Jakub Gregor, Ladislav Dušek

This case study was supported by the Interreg CENTRAL EUROPE Programme funded under the European Regional Development Fund, project no. CE1047 "INTENT – Using Guidelines to Trigger Social Entrepreneur Solutions Towards Better Patient-Centred Cancer Care in Central Europe".

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2017–2020
Keywords	Cancer care, patient-centred approach, data collection, data analysis, benchmarking, interactive visualisation
Research question	How to create a new benchmarking tool supporting patient-centred cancer care and health innovation?
Type of result	Complex web application
Level of data processing	Descriptive statistics, Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

Cancer is the second leading cause of death in the European Union, accounting for approximately 26% of all deaths [1]. As cancer patients' treatment outcomes improve and survival is longer, the nature of cancer care (and patients' expectations) change from acute to chronic disease settings. A patient's chances of long-term survival depend not only on their age and type of cancer, but unfortunately also on the country or region. The shift in "what cancer is" requires a more patient-centred approach to cancer care. This approach should deliver more equitable access to patients, harmonised and coherent care pathways, policy incentives for performance improvement and openness to more relevant and affordable innovations. This reflects evidence that the model of cancer care based solely on acute inpatient care is costly and economically unsustainable in its present form and that current and emerging innovations will accelerate the capacity to transfer more services to outpatient care and local communities closer to where patient lives. In this changing context, several European cancer research and health institutes joined the INTENT project[1] to find solutions for innovative patient-centred cancer care. It targeted and involved various stakeholders: patients, cancer care providers, clinicians and other healthcare professionals, and policymakers. However, knowledge sharing and dialogue need a practical focus on making changes if a new patient-centred and financially sustainable model for continuity of care is to be realised and implemented. One of the main objectives of INTENT was to understand what patient-centred care means and explore/adopt new tools and methods supporting this approach in the context of relevant policy recommendations and local stakeholders' needs. To help achieve this, an online benchmarking tool based on valid data from a broad survey was designed, developed, and implemented to show what needs to be done better to deliver patient-centred care.

AIMS

- To define a patient-centred care model and multidimensional interregional descriptive indicators.
- To effectively collect data using online surveys, which describe the quality of cancer care from different points of view (cancer care providers and their staff, patients, and policymakers).
- To benchmark comprehensive cancer care among selected European institutions using a proper interactive visualisation tool.

[1] https://www.interreg-central.eu/Content.Node/INTENT.html

METHODS

The process of designing and implementing the benchmarking tool was based on the need to comprehensively map the patient-centred cancer care area in the selected regions and to define a suitable innovative model and guidelines based on the active involvement of pilot regions in Central Europe. Below is a brief description of the complete process of all addressed critical topics.

- Sharing understanding of a patient-centred care model.
- Performance indicators identification by INTENT partners and integration into local information systems.
- Survey, review, and integration policy recommendations produced by INTENT at the local/regional/national level.
- Implement the innovative model and benchmarking tool in all selected pilot regions with cancer care providers.
- A critical mass of professionals trained in participating regions to ensure proper application of this model, associated benchmarking, and innovative solutions.
- Sharing good practices, benchmarking results, and innovative ideas using the INTENT Virtual Know-How Centre (VKHC) platform.

PATIENT-CENTRED CANCER CARE MODEL

Before the design of the online benchmarking tool started, a new patient-centred cancer care model (PCCM), which guides the effective integration of health services offered by the five pilot regions (Budapest – Hungary, Slovenia, South Moravia – Czech Republic, and the Veneto and Friuli Venezia Giulia regions – Italy) taking part in cancer policy mapping, was developed. The entire model proposal went through the following three stages of consensus building: (i) a situational analysis of existing care models in pilot regions, (ii) conducting an online survey and follow-up interviews with local stakeholders to identify their expectations for a new model, (iii) development and adoption of a patient-centred cancer care model. Based on mentioned model with implementation guidelines, sets of indicators were defined, either as new ones, or adopted from previous projects, such as CANCON[2], BenchCan[3],DanuBalt[4], or iPAAC[5]. The

[2] https://cancercontrol.eu/

[3] https://www.oeci.eu/Attachments/BenchCan_A5_2014_5.pdf

[4] https://cordis.europa.eu/project/id/643738

[5] https://www.ipaac.eu/

next step was the tool's design, development, testing and implementation and subsequent integration of these indicators, which are compatible with the model of patient-centred cancer care.

SURVEY ON EXPECTATIONS

The benchmarking tool was planned from the beginning as an authenticated (identifying someone's identity by assuring that the person is the same as what he is claiming) and authorised (granting someone to do something – a way to check if the user has permission to use a resource or not) section of the INTENT Virtual Know-How Centre platform. A survey of expectations was carried out to obtain and adequately evaluate the tool's relevant ideas, specifications, and functions in the form of user stories. The collection of short statements from participants of local stakeholder panels, according to the given instructions and examples, was done in the following general format:

- As a [*please insert your role*], I want to [*please insert the requirement or feature of the virtual know-how centre*] so that [*please insert the desired objective*].

Each partner site provided at least five user stories showing basic expectations from the platform and its primary functions, features, and modules. Below are some examples of correctly defined user stories from various perspectives.

- As a hospital manager, I want to benchmark the performance of my hospital according to the different PCCM dimensions so that I can quickly identify in which aspects the centre possibly deviates from PCCM standard or practice achievable best performers.
- As a patient representative, I want to find out easily where the role of patient representatives is within the PCCM so that I can effectively participate in implementing the PCCM at the particular cancer centre.
- As a potential user, I want to access the VKHC centre on mobile devices smartphones or tablets easily) so that I can use the knowledge without a PC immediately available.

QUESTIONNAIRE MANAGEMENT

A final multidimensional set of indicators has been defined among all project partners. Questions for individual target groups were formulated to be included in a survey and analysed using the online benchmarking tool. Respondents from five target groups (management, medical doctors, nurses, expert patients, and patients) answered questions about the six following dimensions:

1. Patient-centred culture
2. Communication, information, and education
3. Accessibility and continuity of care
4. Shared decision-making and multidisciplinary approach
5. Enhancing the quality of life
6. Research and innovation

A total of 110 questions were prepared with a parametric "yes/partially/no" response structure. This allowed for subsequent standardised assessment of individual dimensions and institutions. All questionnaires were designed in the online form and, upon a request from some partners, as a structured MS Excel sheet. All questionnaires were completed anonymously, and the respondents' identity was not disclosed to hospital staff, data managers or analysts.

TECHNICAL BACKGROUND

Integrating a new set of indicators required robust software support that enabled subsequent storage, validation, and data preparation for further processing by related services. In general, the PostgreSQL database system stores individual indicators and records, including metadata. It allows easy manipulation of available data using SQL language and ensures data integrity by defining complex constraints. Data from a structured questionnaire were imported into the central database via a standardised export file. It was then possible to transform the data into a suitable format for subsequent visualisation on the online benchmarking tool. Visualisations were developed and implemented using JavaScript libraries (d3.js, NVD3), which are directly designed for this purpose.

RESULTS

BUSINESS UNDERSTANDING

The INTENT project team asked representatives of all involved project partners and their stakeholder panels for opinions and requirements for the design of the Virtual Know-How Centre, notably the online benchmarking tool. Short statements from participants of local stakeholder panels were collected according to the description and examples specified in Methods. Each partner provided user stories that showed critical expectations from the virtual know-how centre with the online benchmarking tool and its functionalities in different target groups: In total, 25 user stories from different target group levels were specified: manager – 7, patient – 10, medical doctor – 6, nurse – 2. This input was the basis for designing and developing all information and analytical tools under the umbrella of VKHC.

DATA UNDERSTANDING

Data collection on patient-centred cancer care was performed in the following target groups (with further specification on their involvement and experience):

- Managers (only a few managers per hospital, authentication required)
  - One person from your institute working on the Intent project would circulate the questionnaire among the various management board members and collect their feedback.
- Hospital staff (tens of employees, no authentication required)
  - Medical doctors: Both junior and senior physicians who have been at the institute for at least two years. The following departments could be covered: surgery, medical oncology, radiotherapy, etc.
  - Nurse: Both junior and senior nurses have been at the institute for at least two years. The following departments could be covered: surgery, medical oncology, radiotherapy etc.
- Patients (hundreds of patients, no authentication required)
  - Patients: Preferably inpatients or outpatients who already have a diagnosis and have started their treatment plan.
  - Expert patients: Any expert patients from patient associations who are also familiar with the operation of your Institute.

The survey structure in terms of target groups and dimensions is given in Table 1.

Table 1: Structure of the survey. Numbers indicate the number of questions answered in each target group and dimension

Dimension / Target group	Manager	Doctor	Nurse	Patient	Expert patient
Patient-centred culture	7	4	4	1	5
Information, communication, education	7	2	2	5	0
Accessibility and continuity of care	9	4	4	6	0
Shared decision-making and multidisciplinary approach	6	6	5	3	0
Enhancing quality of life	5	5	4	2	0
Research and innovation	7	5	0	2	0

DATA PREPARATION

The questionnaires were prepared in the languages of all partners involved in the INTENT project (i.e. Czech, Hungarian, Italian and Slovenian), especially given the expected more limited language skills in some target groups and also for the convenience of the respondents. In total, the following numbers of questionnaires were thus obtained from all project partners: management – 5, doctor – 53, nurse – 53, patient – 510, expert patient – 26. The questionnaires were collected online or office using MS Excel structured sheets. Data cleaning, transformation, and mapping were performed for further analytical processing to prepare valid data for the interactive visualisation tool.

MODELLING

Responses to questions of the *Single Choice* type were in the Y-P-N format, where 1 point was assigned for the response Y (yes), 0.5 points for the response P (partly) and 0 points for the response N (no). Answers to multiple choice questions were in the format of an empty field or a Y response, corresponding to 1 point. The average value for a multiple choice question was calculated by summing the Y responses and dividing by the number of options, except for "None of the above". If one of the options was "None of the above", and a respondent selected it, then the question value was 0.

QUESTION TOTAL SCORE OVER TARGET GROUPS

Since the number of respondents in each target group differed (e.g., management N = 1 and patients N = 98 for one of the centres), the total question score was calculated as the average of the question values of all target groups. For example, if the resulting values were 0.7 (patients), 0.8 (doctors) and 1.0

(management), the sum of these values was 2.5; dividing the number of target groups gave a score of 0.8.

DIMENSION VALUE

The values of all responses of a dimension were summed, and the sum was divided by the number of responses. In the case of multiple choice questions, the question value (the normalised 0-1 value) was not considered, but each option was considered a separate response.

TARGET GROUP TOTAL SCORE

The target group total score was calculated similarly to the question total score over target groups (see above) because there was a different number of questions in each dimension, and we aimed to use the same weight for all dimensions.

DIMENSION TOTAL SCORE

Since the number of respondents in each target group was different and we aimed to use the same weight for all target groups, the total dimension score was calculated as the average of the dimension values of all target groups (see example in Figure 1).

BENCHMARKING OF THE CENTRES

Each involved centre was compared with the centres with the highest or lowest scores in benchmarking graphs and tables. The values of benchmarking charts were composed of dimension values, and benchmarking tables were written by question values. In both cases, only two values were displayed if the compared centre had the highest or the lowest score (see Figure 1, Figure 2, Figure 3).

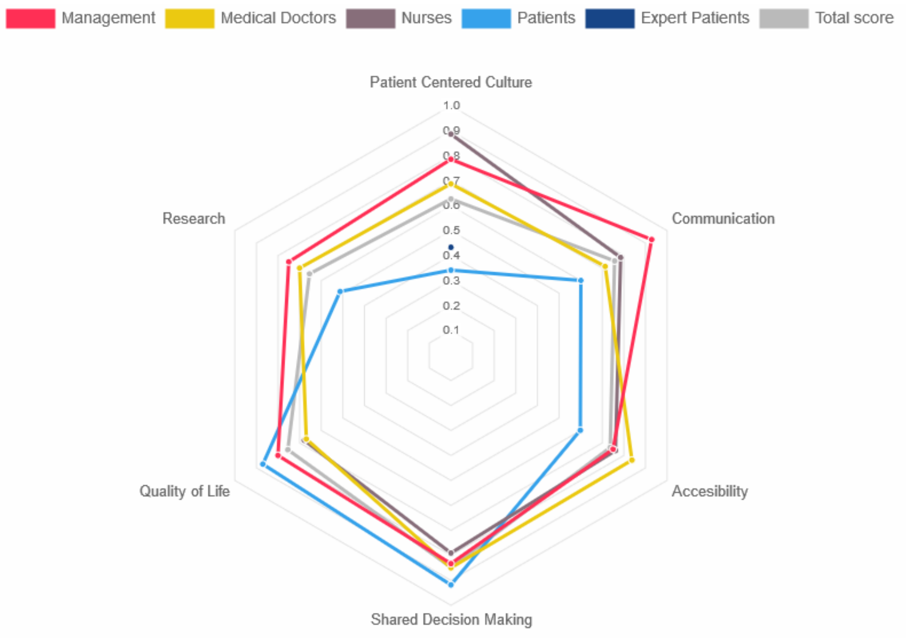
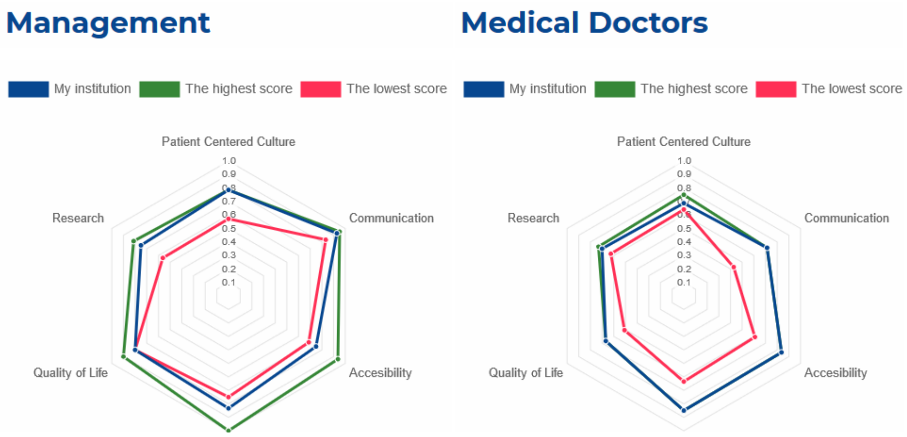


Figure 1: Data visualisation – overall performance (total score) of a particular centre in six evaluated dimensions over the target groups



Management

Medical Doctors

Figure 2: Data visualisation in a benchmarking graph – benchmarking of a particular centre compared to the best and worst-performing centres

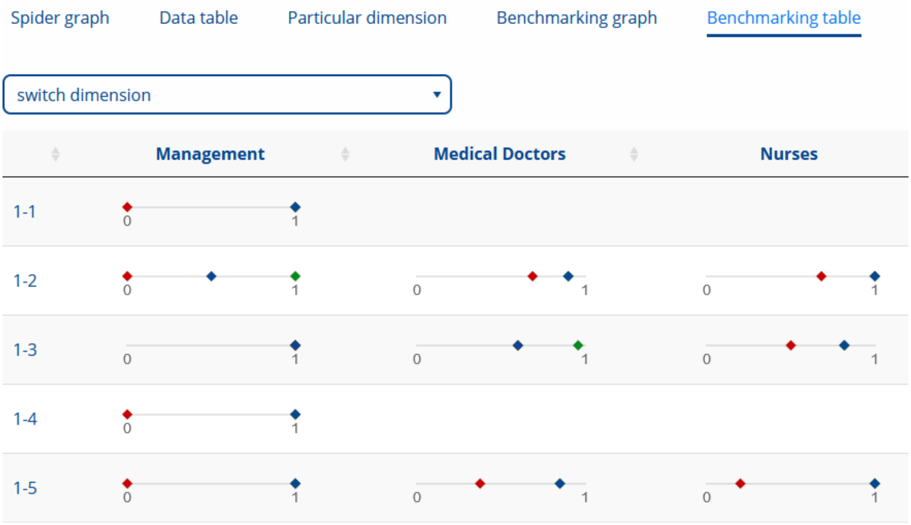


Figure 3: Data visualisation in a benchmarking table – benchmarking of a particular centre compared to the best and worst-performing centres

The interactive data presentation module contains five analytical tools designed to analyse the particular centre's results in different views, target groups and dimensions (spider graph, data table, particular dimension) and to analyse benchmarking and comparison with other centres (in de-identified mode).

EVALUATION

Assessment and analysis of partners' and pilot sites' expectations served as a basis for developing the Virtual Know-How Centre. The international cooperation allowed the INTENT project team to gain insight into the needs of different stakeholders in different contexts. The development of the tool and its content included the experience of centres from other countries. It should therefore help the international usability of the outputs within the Central European region and beyond. The evaluation and review process included several successive phases to ensure maximum correctness and user clarity of all presented outputs: (i) testing during programming and deployment to the development server, (ii) internal user testing of all functionalities within the development team, (iii) verification of presented outputs against input data, (iv) final user testing and commenting by the INTENT team.

DEPLOYMENT

The final version of the benchmarking tool and its development was incorporated as a part of the Virtual Know-How Centre. The VKHC and the benchmarking tool were developed to serve and collect data from European cancer care providers (centres). Cancer centres interested in benchmarking exercises can contact the INTENT partners – directly or via the Organisation of European Cancer Institutes (OECI) – and organise a new round of the benchmarking exercise together. The developed software tools are flexible and can be used in other new benchmarking activities within the new centres, countries, territories, etc.

DISCUSSION

The Virtual Know-How Centre[6] is an online repository of good practices for patient-centred care. It includes outputs prepared by the INTENT project: the patient-centred care model, its implementation guidelines, indicator sets, etc. This platform links them to examples of good practices and innovative solutions for implementing the patient-centred cancer care model principles. One of the VKHC integral parts is an online benchmarking tool, which allows the authorised personnel from participating centres to collect data on indicators of a particular centre or to collect individual responses (from patients or staff) through online surveys and to perform benchmarking exercises between participating centres (e.g. comparison of results with standards or best performers). This tool is based on new performance indicators on quality of care, patient benefits, cost-effectiveness, and innovation capacity. It enables comparison, organisational change, and social entrepreneurship solutions. The main impact of the benchmarking tool on the cancer care centres both in concerned territories and outside is as follows:

— A complex online platform for data collection among all involved/participating healthcare providers

— Benchmarking in all comprehensive cancer care centres, particularly in European regions

— Leverage improvements in patient-centred cancer care and social solutions to support the centres' performance and associated supportive services for cancer patients

[6] https://intent-benchmark.uzis.cz/

- Improving the performance at each participating centre by comparative analysis of provided services and the quality of these services
- Providing the participating centres with best practice examples in a way that contributes to improving the quality of interdisciplinary patient treatment

The INTENT project provided a unique opportunity to develop and deploy tools based on comprehensive multi-stakeholder international situational analysis and experiences and skills of crucial cancer centres of the Central European region. This should improve patient outcomes and experience in Central Europe and beyond.

EVALUATION OF THE AIMS OF THE CHAPTER

- To define a patient-centred care model and multidimensional interregional descriptive indicators.
  - In cooperation with several European cancer centres involved in the INTENT project, based on the mutual sharing of experience in the field of patient care, it was possible to design multidimensional indicators according to defined target groups, which took into account the attitudes of all stakeholders.
- To effectively collect data using online surveys, which describe the quality of cancer care from different points of view (cancer care providers, patients, and policymakers).
  - Based on the defined user stories and given indicators, a comprehensive questionnaire system was created, available online, where it was possible to provide feedback from a given point of view in a user-friendly environment. All data were comprehensively stored directly in the designed database.
- To benchmark comprehensive cancer care among selected European institutions using a proper interactive visualisation tool.
  - The results of the questionnaire survey analysis were presented in the form of an online visualisation (dedicated authorised module of the Virtual Know-How Centre), where comparisons between the participating centres are available via suitable graphs. Data export for further processing purposes is also an integral part of this.

LESSONS LEARNED

A consortium of European partners, namely several cancer care facilities, built the outputs of the INTENT project based on previous cooperation, experience and achieved results. For the understanding phase, this meant that a significant amount of time was saved. The data understanding phase included the design of a model for collecting structured feedback with subsequent evaluation in the form of easy-to-understand graphical outputs. Due to the different domains in which individual partners had different levels of expertise, it was possible to appropriately combine healthcare, oncology specialty, IT and communication with a result that was highly rated internationally.

The appropriately chosen structure and technologies for data collection made the subsequent phases easier to implement. An interactive web application including a module for comparing the participating centres can be considered as one of the project's main outputs. The results achieved in the INTENT project, together with other international experience gained, have helped in the preparation of a strategic document called the Czech National Cancer Control Plan 2030[7], where an entire chapter is devoted to the issue of patient-oriented care and efforts to ensure the maximum possible quality of life during the disease, after cure and in the terminal stage.

REFERENCES

[1] OECD European Union. Health at a Glance: Europe 2020: State of Health in the EU Cycle [Internet]. Paris: OECD Publishing; 2020. Available from: https://doi.org/10.1787/82129230-en.

[7] https://www.mzcr.cz/wp-content/uploads/2022/06/Narodni-onkologicky-plan-Ceske-republiky-2030.pdf

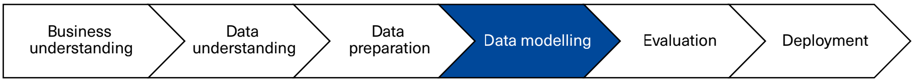
SECTION C

16

INTERACTIVE TOOL FOR CRISIS MANAGEMENT IN TIMES OF COVID-19 EPIDEMIC IN THE CZECH REPUBLIC

Martin Komenda, Michal Vičar, Jaroslav Číhal, Daniel Klimeš, Petr Šnajdárek, Jan Ryšavý, Lenka Šnajdrová, Ladislav Dušek

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2020–now
Keywords	COVID-19, dashboard, crisis management, open data, epidemic
Research question	How to design complex advanced analytical reporting services using business intelligence to support data-oriented decision-making during an epidemic.
Type of result	Complex web application
Level of data processing	Advanced analyses, Open datasets

DATA TO DOWNLOAD

INTRODUCTION

The subsequent waves of the COVID-19 epidemic, caused by a new coronavirus known as SARS-CoV-2, were contained in the Czech Republic, mainly thanks to the timely and vigorous introduction of blanket measures. Although the Smart Quarantine project was only introduced as the number of new cases began to decline, it is still one of the tools to support data-driven decision-making when discussing proposed blanket measures. In general, the Smart Quarantine project allows the responsible authorities to substantially increase their ability to react quickly to the emergence of new local outbreaks, and thus maintain the downward trend in the number of cases. This activity comprises a set of measures, procedures and tools that, once fully implemented and with the continuous development of digitisation, will create a sufficiently robust system that can respond very quickly even to sudden surges in outbreaks, including large-scale epidemics and pandemics of global significance. The main objectives of this project were to build an operational, sufficiently flexible and standardised tool for dealing with public health threats (environmental, biological, chemical, radiation and combined threats impacting public health). The three main areas covered by the project were: (i) tracing and isolation of at-risk contacts, (ii) ensuring rapid testing and laboratory investigation, (iii) multi-source data integration and central monitoring to provide information service and system communication.

Especially at the beginning of the epidemic in the Czech Republic, it was complicated and time-consuming to prepare ad hoc reports in cooperation with the Czech Armed Forces, the National Agency for Communication and Information Technologies and the Institute of Health Information and Statistics of the Czech Republic as a basis for the Integrated Central Management Team. Gradually, new information and communication tools began to be developed and subsequently used for crisis management purposes, such as (i) ArcGIS situational maps with the ability to interactively change the data view based on users' requests, (ii) a tracking system for the movement of mobile collection vehicles that reported the number of samples and their performance, (iii) the COVIDForms application designed to centralise up-to-date data from the regional public health authorities, laboratories and sampling points, (iv) the Dashboard, which provided advanced analytical reporting as an interactive online visualisation tool, covering the last of the three areas mentioned above. The main objective was early threat identification and risk analysis based on orchestrating all available data processed in real time from various information systems.

AIMS

— To understand, systematically store, and consolidate data from various information systems focusing on the epidemic caused by COVID-19 in the Czech Republic.

— To design proper interactive visualisations presenting required reports daily.

METHODS

In the early days of the epidemic, it was essential to build a consolidated reporting capability as soon as possible to serve as an important source of information for the public, health authorities and government. However, data collection has not been electronic or fully automated since the outbreak. Active assistance was provided by the army and hospital dispatch, which called individual persons and manually entered the results into dedicated information systems. A comprehensive and machine-based way of collecting and processing relevant data is required, but human factors will always be needed, given the requirement to provide crisis management. Therefore, an online Dashboard tool was created to visualise and monitor epidemiological data on COVID-19 in the Czech Republic. This platform provided the professional community with up-to-date information on the number of confirmed cases, hospital admissions, deaths and other relevant statistics related to the pandemic. Vaccination information, including the number of vaccinated persons with coverage in different parts of the country, was an integral part of the platform.

This section describes the heterogeneity of solutions (cloud-based, On-Premise) for processing input data sources across platforms, which were systematically used during the COVID-19 epidemic. It also describes the basic technological framework for the effective use of business intelligence tools with respect to the On-Premise Dashboard solution as a necessary requirement from the umbrella institution, namely the Institute of Health Information and Statistics of the Czech Republic (IHIS).

DATA SOURCES

This case study works with data in different formats, structures or representations. This often poses a challenge in processing, systematic storage and subsequent analysis. In particular, this involves processing data in file formats such as TXT, CSV, XML, XLSX or JSON. Just as the input data and formats are diverse, there are several different ways of retrieving data for further processing (API, web scraping, open data, database queries, and shared storage). Below is a list of some of the most commonly used methods.

- **Application Programming Interface (API):** Many web services provide APIs that allow automated communication with their data sources and enable data retrieval in a structured and standardised way. Examples include data on booking and vaccination registration or data from the Covid Forms App.
- **Web scraping:** This technique extracts data from web pages through an automated process. Scripts or programs can crawl web pages, analyse their content and extract the required data. One example is the global overview of COVID-19[1].
- **Open data:** An integral part of valuable resources are open data provided by government institutions – in this case, by the Institute of Health Information and Statistics of the Czech Republic through the local catalogue of open data in healthcare[2].
- **Database queries:** In cases where it is possible to access database systems, data are retrieved directly using SQL query scripts.
- **Shared storage:** Data are retrieved from secure data stores where individual files are stored.

[1] https://www.worldometers.info/coronavirus/

[2] http://data.mzcr.cz

TECHNOLOGICAL BACKGROUND

The first version of Dashboard was created as a cloud-based solution that allowed the entire solution to be hosted in the cloud. Data could be managed using a web browser while making full use of the Microsoft Azure cloud.

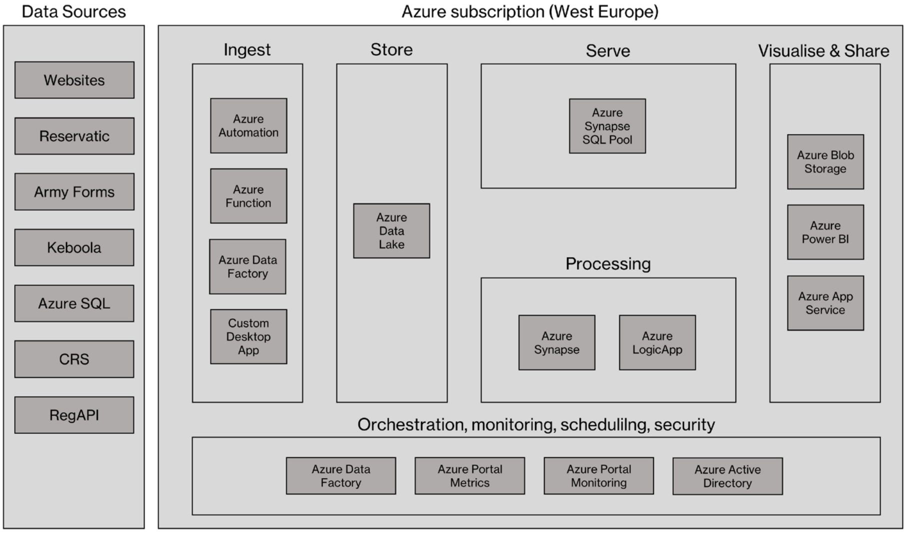


Figure 1: Scheme of solution in the cloud platform

Below is a brief description of the most important technologies and approaches that were part of the technical design of the first version of Dashboard.

— **Data sources:** This system uses data from different data sources, employing various techniques such as Python, C#, PowerShell, database processing.

— **Ingest:** In the context of Dashboard, it is the process of collecting and ingesting data from various sources and incorporating them into the system. The process includes data extraction, transformation, and loading into the database.

— **Azure function:** This feature makes it possible to provide and run serverless functions in the cloud, run a code, and integrate with other Azure services.

— **Azure Data Factory:** This mainly includes capabilities to manage pipelines that define and control the movement of data. Triggers must be set, making it possible to define when and how often pipelines should be run. Last but not least, they allow for monitoring the running of individual pipelines.

- **Azure Blob Storage:** Cloud storage within the Microsoft Azure platform. It is designed to store large volumes of unstructured data while providing high availability, resistance to faults, as well as additional scalability.
- **Synapse SQL (formerly SQL Data Warehouse):** This distributed relational database supports scalable data processing. It is used for operations over large volumes of data using T-SQL.
- **Serve (Azure App Service):** This service is of crucial importance, providing an environment for running a variety of application types, including web applications written in different types of environments (.NET, Python, Java). The service enables high scalability and relatively easy deployment and integration with other Microsoft Azure services.
- **Orchestration, monitoring, security.** The Synapse Pipelines are an important part of Azure Synapse. This service is used to orchestrate and schedule data flows or transformations, making it possible to create and manage ETL processes (Extract, Transform, Load), which are processes for integrating data from different data sources.
- **Azure Active Directory:** Access security management is an integral part of the Dashboard. It is a fundamental building block for managing identities and access across the Microsoft Azure ecosystem – not only for the cloud, but also for on-premise environments.
- **Azure Web Application:** This feature is used to deploy and run web applications in the cloud. It involves hosting web applications that can integrate with other cloud services.
- **Power BI:** This comprehensive business intelligence solution provides tools to analyse, visualise and share data. It includes a wide range of features for extracting insights and information from their data, visualising and analysing it. In particular, the Dashboard concept uses Power BI Embedded Services, which makes it possible to integrate visualisations into users' applications and websites. Users can thus interactively work with the data.
- **DAX language:** The DAX language is designed to support interactivity in visualisations. DAX (Data Analysis expression) provides a rich set of functions that make it possible to perform complex calculations, aggregations, filters and other manipulations with data. Calculations can be performed at rows, columns, or even whole tables.

As already mentioned above, the original technology environment for the epidemiological dashboard was Microsoft Azure Cloud, which was later migrated to an On-Premise solution. This strategy was chosen based on evaluating operational efficiency, overall management costs, and other factors. Integral parts of the migration from Microsoft Azure Cloud to the On-Premise solution were:

- analysis of used data and calculations,
- rewriting the solutions for data retrieval,
- creation of database structures and rewriting of procedures providing calculations,
- scheduling tasks to manage the process of data processing.

RESULTS

BUSINESS UNDERSTANDING

For a detailed understanding of this domain, it was necessary to master not only the basic epidemiological characteristics and their interrelationships but, above all, to thoroughly and in great detail understand all the input data sources that were produced daily through the Smart Quarantine information systems used in the fight against the epidemic in the Czech Republic. With the ambition of designing, developing and operating a long-term tool that would exceed the needs of the unexpected COVID-19 epidemic in the Czech Republic, it was crucial at the very beginning to document the data flows between the systems, to be in constant contact with the operators of these systems despite possible changes in subcontractors, and to communicate any changes, outages or inaccuracies to the target users in a coordinated manner. Below is a comprehensive diagram that captures all the key players involved in the process of data entry and processing. Figure 2 shows how the different stakeholders (Ministry of Health, regions, regional public health authorities, health care providers, general practitioners, laboratories) were involved in the complex process of sampling, evaluation, reporting to the central system, online reporting and data-driven management.

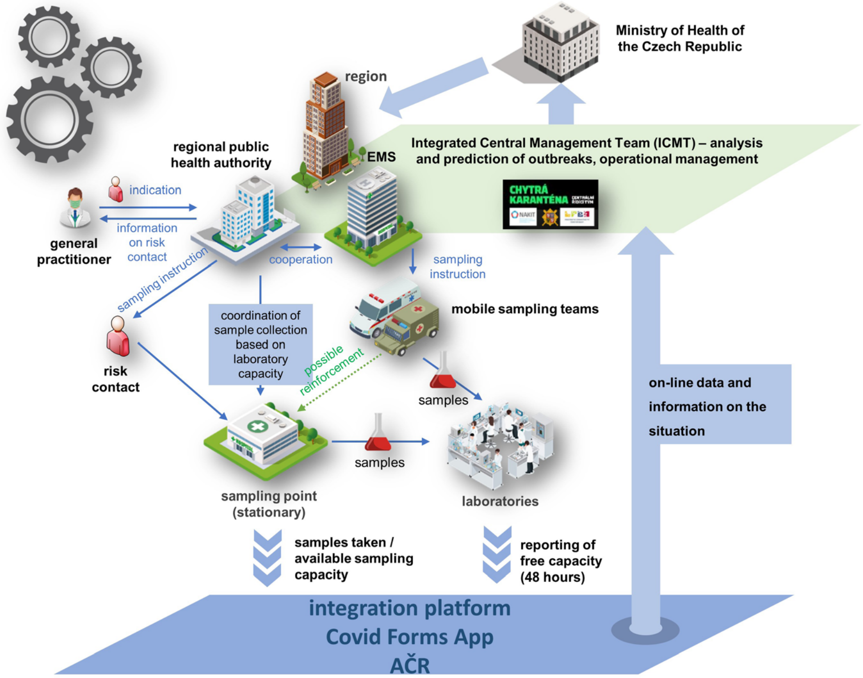

Ministry of Health of the Czech Republic
region
regional public health authority
EMS
Integrated Central Management Team (ICMT) – analysis and prediction of outbreaks, operational management
CHYTRÁ KARANTÉNA
indication
information on risk contact
general practitioner
sampling instruction
cooperation
sampling instruction
mobile sampling teams
coordination of sample collection based on laboratory capacity
possible reinforcement
risk contact
samples
samples
sampling point (stationary)
laboratories
on-line data and information on the situation
samples taken / available sampling capacity
reporting of free capacity (48 hours)
integration platform
Covid Forms App
AČR

Figure 2: Data distribution between sub-entities in the process of COVID-19 reporting[3]

DATA UNDERSTANDING

The data provide the basis for the production of reports with a summary of the daily trends in incidence, the numbers of hospital admissions, people who were cured, those who died, as well as prevalence; all these numbers are provided in detail by region and district, with a breakdown by age group and sex and a range of other inputs.

- Positively tested: This dataset provides line-by-line de-identified data on incidence (i.e. numbers of infection cases). Each dataset row describes the case by the person's residence, sex and age group.
- Cured: The dataset contains records of people cured of COVID-19 as reported by regional public health authorities.
- Hospitalised: Information on hospitalised patients describing the course of hospitalisation – current and total number of hospital admissions, breakdown by symptoms.

[3] Designed by the team of the Smart Quarantine project.

- Deaths: The dataset contains records of deaths related to COVID-19 as reported by regional public health authorities. It includes deaths of persons who tested positive for COVID-19.
- Overview of cases worldwide: Data from the Worldometer website, which provide information on COVID-19 trends worldwide.
- Testing and tracing: Data are obtained from the Covid Forms Application, where reports from regional public health authorities, laboratories, sampling sites and vaccination sites are available.
- Booking: The source data are retrieved using several APIs. Of these, two provide information on the list of places where it is possible to book for antigen and PCR tests, or vaccinations. The other APIs provide data with an overview of the capacity to book for antigen and PCR tests or vaccination (for either primary dose or booster dose).
- Registration: The data source is information from the registration system, providing information on the status of the registration request. Each registration line falls into one particular group.
- Vaccination: The data source is open datasets provided by the analytical department of the IHIS. The documents are also published in the catalogue of open data.
- Intensive care capacities: These data provide information on daily summaries of changes in free capacities in acute care facilities.

The complete data flow through the entire system for COVID-19 epidemic monitoring are documented in Figure 3.

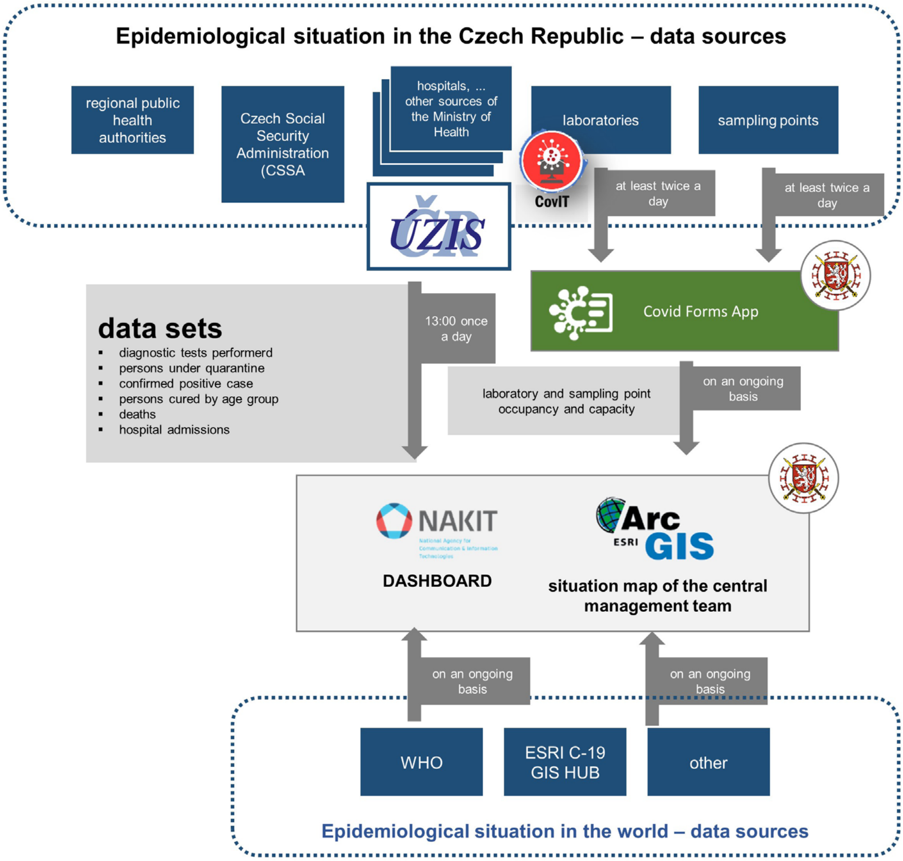

Epidemiological situation in the Czech Republic – data sources
regional public health authorities
Czech Social Security Administration (CSSA
hospitals, ... other sources of the Ministry of Health
laboratories
sampling points
CovIT
ÚZIS ČR
at least twice a day
at least twice a day
Covid Forms App
data sets
diagnostic tests performerd
persons under quarantine
confirmed positive case
persons cured by age group
deaths
hospital admissions
13:00 once a day
laboratory and sampling point occupancy and capacity
on an ongoing basis
NAKIT
National Agency for Communication & Information Technologies
DASHBOARD
ESRI ArcGIS
situation map of the central management team
on an ongoing basis
on an ongoing basis
WHO
ESRI C-19 GIS HUB
other
Epidemiological situation in the world – data sources

Figure 3: Data sources and data flow in the process of COVID-19 reporting

DATA PREPARATION

Data preparation usually involves several steps. This section focuses on the ETL (Extract, Transform, Load) transformation layer.

— Extraction: Data are obtained from several different sources (state institutions such as the Ministry of Health of the Czech Republic, Institute of Health Information and Statistics of the Czech Republic, Czech Statistical Office, laboratories, systems for vaccination registration, booking and registration systems).

— Transformation: Extracted data may be in a form that requires further modification and transformation to achieve compatibility with the data model.

This includes activities such as modifying the format of the data type, column naming conventions, normalisation of values, or enrichment with additional attributes.

- Validation and cleaning: Data are checked for inconsistencies and further cleaned, e.g. each record must have information about the region where it was made.
- Loading into the data repository: Data that have passed validation and transformation are then loaded into the data warehouse.

MODELLING

The resulting reports in the Dashboard are organised into several sections by topic – infections, testing, vaccinations, bed capacity, etc. In addition to the standard statistics and visualisations, each report also includes the date and time when the report was updated, which is important for the credibility of the report and its possible control. Thanks to the implementation of several filtering components, it is possible to quickly display a detailed view by (for example):

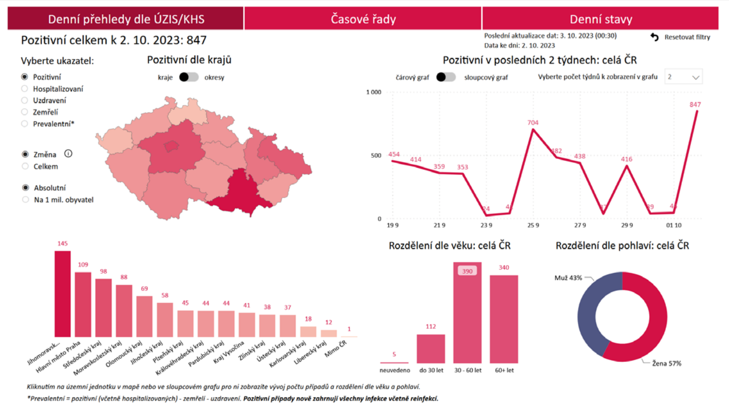
- region, district and quarter of the Capital of Prague – both as a bar chart and as a map;
- data in the basic overview up to 8 weeks back, either as a line or bar chart (in more detailed reports as a timeline with the possibility of advanced statistics from the beginning of pandemic monitoring);
- resolution into daily increments or as a cumulative total;
- absolute values and per population;
- sex and age categories.

One of the many interactive reports is a case overview of the country, which captures the main attributes of COVID-19. This overview shows the following epidemiological characteristics:

1. Positive cases: how many cases of the disease are captured. In the first phase of the pandemic, the number of infections was equal to the number of people. In the next phase, the figure included the number of repeated infections (reinfections) in the same person.

2. Hospitalisations: the number of cases with a more severe disease course, which required hospital admission. This number was used to assess the severity of disease course in the population and to provide an overview of the

hospital system utilisation rate. Unlike positive cases, which only showed the number with a positive test result (while far from every real positive case was actually tested positive), hospitalisations were a figure captured in hospitals, i.e. numbers of people with a more severe disease course. (iii) Cured: The number of people who contracted COVID-19 and recovered. The condition was usually a negative test for COVID-19 after several days of quarantine. (iv) Deaths: The number of people who died and had previously tested positive for COVID-19. (v) Prevalence: The number of patients at a given time. It was calculated as the number of positive cases minus the number of cured people and those who died.



Pozitivní celkem k 2. 10. 2023: 847

Poslední aktualizace dat: 3. 10. 2023 (00:30)
Data ke dni: 2. 10. 2023

Vyberte ukazatel:
- ● Pozitivní
- ○ Hospitalizovaní
- ○ Uzdravení
- ○ Zemřelí
- ○ Prevalentní*

- ● Změna
- ○ Celkem

- ● Absolutní
- ○ Na 1 mil. obyvatel

Pozitivní dle krajů

Pozitivní v posledních 2 týdnech: celá ČR

Vyberte počet týdnů k zobrazení v grafu: 2

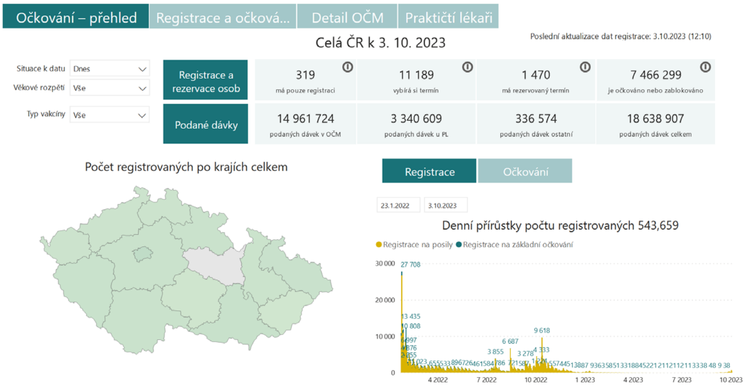
Rozdělení dle věku: celá ČR

Rozdělení dle pohlaví: celá ČR

Kliknutím na územní jednotku v mapě nebo ve sloupcovém grafu pro ni zobrazíte vývoj počtu případů a rozdělení dle věku a pohlaví.

Prevalentní = pozitivní (včetně hospitalizovaných) - zemřelí - uzdravení. ***Pozitivní případy nově zahrnují všechny infekce včetně reinfekcí.*

Figure 4: Screenshot from the Dashboard platform: Overview of cases in the Czech Republic (available only in the Czech)



Celá ČR k 3. 10. 2023

Poslední aktualizace dat registrace: 3.10.2023 (12:10)

Situace k datu: Dnes
Věkové rozpětí: Vše
Typ vakcíny: Vše

Registrace a rezervace osob	319 má pouze registraci	11 189 vybírá si termín	1 470 má rezervovaný termín	7 466 299 je očkováno nebo zablokováno
Podané dávky	14 961 724 podaných dávek v OČM	3 340 609 podaných dávek u PL	336 574 podaných dávek ostatní	18 638 907 podaných dávek celkem

Počet registrovaných po krajích celkem

23.1.2022 – 3.10.2023

Denní přírůstky počtu registrovaných 543,659

● Registrace na posily ● Registrace na základní očkování

Figure 5: Screenshot from the Dashboard platform: Overview of vaccinations (available only in the Czech)

Another example is the vaccination report, which mapped the main statistics on vaccination registrations and bookings, and the vaccination itself. This report allowed for capacity planning and establishing vaccination delivery sites, both at vaccination sites (used mainly in the first phase of the pandemic) and at GPs (to whom vaccination has gradually moved as the disease has begun to recede).

The report details the entire vaccination lifecycle: from the initial registration process, through booking a specific appointment, to the vaccination itself. It was possible to filter by age range, priority groups and vaccine type (manufacturer). An interactive map by region showed the regional distribution of registrations and vaccinations; the timeline showed daily increments and trends in vaccination in the Czech Republic.

EVALUATION

Data checking and validation are carried out in several ways and involve different techniques and approaches. As part of the checking and validation process, it is important to define appropriate rules and standards that determine what data are considered valid. When implementing external sources, individual attributes are validated to ensure they contain only the expected values. Input parameters such as date must not allow the insertion of another data type; the "date from" parameter cannot be later than the "date to" parameter. The actual data sources prepared by the analytical department of the IHIS are validated at several levels after uploading to the data warehouse, as listed below:

- validation of input data format,
- checking of value ranges,
- checking of reference integrity,
- checking of duplication,
- checking of logical rules.

Visual checks are performed by eye to identify any inconsistencies or other problems associated with the data. During a visual data check, the data are analysed by a human who carefully goes through it to assess its accuracy and quality, possibly by back-checking it in the source system. It is mainly used in the phase of finalising the outputs for the users before handing them over for use. The outputs in the reports are regularly checked daily to prevent escalation by users.

Process data control is another checking procedure that ensures that data are processed correctly in the required order and time, and are ready for further processing. The goal is to verify the following areas thoroughly:

— Error-free process of data preparation and processing; in case of a process failure, the support team is notified.

— Checking of the calculation process: each process is independently monitored with the possibility of attempting an automatic recovery, the support team is notified in the event of a process failure. In this situation, if data retrieval from an external source fails, for example, the attempt to retrieve the data is made again within the given time interval. If completion is still unsuccessful, the process is evaluated as an error.

DEPLOYMENT

The process of design, development, implementation and deployment took place in the Microsoft Azure Cloud environment, where changes were deployed manually without integrating the now quite traditional methods of automation (continuous integration and continuous delivery). This was mainly due to the lack of an intermediate step for User-Acceptance-Testing steps and the inability to efficiently update data sources. In the test mode, changes were developed and approved, and once approved, they were deployed into the production environment.

DISCUSSION

The Dashboard web application for visualising data on the course of the COVID-19 epidemic in the Czech Republic and in selected countries of the world offers the user the possibility to easily change the parameters of the plotted graphs, to display individual views based on extension filters (regions, districts, quarters of the Capital of Prague, age groups, etc.) and to display the required data clearly. For international comparisons, predefined country groups are available (with the option to change to individual selection), together with various indicators, including normalisation to population size and recalculated ratios. Daily case summaries are available, along with selected indicators to refine the required view of available data, graphical representation of utilisation and capacity of sampling sites and laboratories, as well as an overview of utilisation by region and detailed reports for individual laboratories and sampling sites. There is also monitoring of bed occupancy and, of course, a report on the current vaccination status. Microsoft Azure Cloud, the technology for the first version of the application, is undoubtedly one of the leading cloud service providers on the market, providing a wide range of options for developing, operating and managing applications across the entire infrastructure. However, when selecting a technology, it is always important to carefully consider how it will be used in the long term, including the benefits and pitfalls of the overall solution. At the time of the peak of the COVID-19 pandemic, no comprehensive reporting platform with such consolidated data was available. Development was brisk, with new data sources being added several times a week. After a longer operation and the necessary time interval, the described technology was evaluated as rather inefficient and cost-ineffective. The key factors were global infrastructure, management, cost, scalability, availability, data protection and sustainability. Therefore, a migration to an On-Premise solution was made in mid-2023, along with a significant optimisation of the source data processing and individual visualisations.

EVALUATION OF THE AIMS OF THE CHAPTER

- To understand, systematically store, and consolidate data from various information systems focusing on the epidemic caused by COVID-19 in the Czech Republic.
    - MS Azure was used to store and process many different data sources describing the epidemic in the Czech Republic, which was robust enough to provide variability of data processing and the easy possibility of parallel development of required functionalities of various Dashboard modules.
- To design proper interactive visualisations presenting required reports daily.
    - Using the MS Power BI Embedded environment and expert skills, it was possible to process and display all relevant data in individual reports, which were designed primarily for crisis management over current daily data. The interactive reports allowed a quick view of diverse data and personalised views of charts and tables.

LESSONS LEARNED

This long-standing project has provided experience from several perspectives: (i) design and development of web applications (registries) for data collection in accelerated timeframes due to the rapidly spreading epidemic; (ii) communication with target groups (government, interested ministries, officials, regional leadership: mayors, etc.; (iii) design and publication of open data sets including unexpected modifications and implementation of changes in line with the current epidemiological situation. The Dashboard as a tool was based on modelling new views and scenarios as requested by the crisis teams. It proved necessary to have the tools used in practice constantly ready for further implementation - i.e., to perform their technical update, maintenance and development according to current requirements. The database and the individual heterogeneous sources must be up-to-date and always contain relevant information to help deal with a given situation (convening a crisis headquarters, a staff exercise, activation of dispatching centres, a decision-making process supported by user data, albeit from a calm situation).

SECTION C

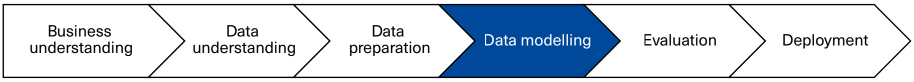
| 17 |

MANAGEMENT TOOL FOR EPIDEMIC MONITORING, ANALYSIS, AND VISUALISATION

Tomáš Pavlík, Lenka Přibylová, Veronika Eclerová, Ondřej Májek, Andrea Kraus, David Kraus, Ladislav Dušek, Martin Komenda

The case study was supported by the following grant project: Online platform for real-time monitoring, analysis and management of epidemic situations — reg. no.: MUNI/11/02202001/2020, funded by the specific internal research programme of Masaryk University.

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2020–2021
Keywords	Predictive models, COVID-19, capacity mapping, data analysis, interactive visualisation
Research question	How to develop a complex epidemic monitoring, analysis, and visualisation tool based on the proven methodological background?
Type of result	Complex web application
Level of data processing	Advanced analyses, Open datasets

DATA TO DOWNLOAD

INTRODUCTION

The global pandemic of COVID-19 has brought several new challenges for a wide range of scientific disciplines. Researchers from Masaryk University helped to address many of these issues. In general, quick and straightforward access to relevant and valid information is crucial for providing objective decision-making support when dealing with any possible epidemic situation in the Czech Republic in future, which might put pressure on the health system again. It can also contribute to evaluating better the effectiveness of the measures taken in the long term. This case study aims to present a solution for an online analytical tool for monitoring epidemic situations, taking into account the available options for managing bed, equipment and staff capacities and prospectively modelling the burden of the population under study, including its structure determined by the different stages of the disease and other factors.

A robust web-based application has been designed to meet the requirements of integrating different data sources, including real-time analysis, and providing transparent interactive reporting of current status over available data in a standardised format. Its architecture allows for a quick and easy selection of suitable predictive models in several steps, then populating them with sample or custom data inputs according to the defined structure, setting the necessary parameters together with model calibration, and then displaying the visualisation of the mentioned models. An integral part of this is reporting the results of the predictions about the data describing the occupancy of inpatient care capacities.

AIMS

- To understand and integrate various data sources describing the epidemic caused by COVID-19 in the Czech Republic.
- To propose models based on methodologies allowing short- and medium-term epidemic predictions and long-term scenarios.
- To develop an interactive web application for modelling the future evolution of the COVID-19 epidemic in the Czech Republic about the expected need for the hospitalisation of patients.

METHODS

This section aims to present the solution of the online analytical tool for prospective modelling of the population burden of the monitored disease, including the structure determined by individual stages of the disease and other modelling factors and supporting applications for monitoring epidemic situations about the possibility of processing information on the bed capacity of the health system in the Czech Republic.

DATA SOURCES

Our tool is based on publicly available data sources, published in the format of open datasets and datasets available after approval by the Ministry of Health of the Czech Republic. The data source on hospital capacity is a dataset containing daily summaries of changes in functional ability in acute inpatient care hospitals related to the spread of COVID-19. These data have been regularly updated and published in the catalogue of health open data[1]. Taking into account the evolution of the COVID-19 epidemic in the Czech Republic, the data were divided by period, namely Wave 1 (from 11 April 2020 to 4 May 2020) and Wave 2 (from 1 September 2020 to 1 April 2022); at that time, the data were updated daily. The Infectious Disease Information System (IDIS) is this project's primary source of epidemiological data. The main objective of this nationwide registry is to obtain information on the incidence of infectious diseases to assess the development of the epidemiological situation in various regions of the Czech Republic, monitor the population's health status and manage the provision of healthcare. This registry is not limited to COVID-19; however, data on patients with COVID-19 are its major component, due to the recent epidemic.

MATHEMATICAL MODELS

The developed web application uses selected epidemiological models for short- to medium-term predictions of the development of the monitored disease, regarding not only the number of newly diagnosed cases but also the number of newly hospitalised patients or deaths.

Epidemiological structured models are extensions of epidemiological SIR models that are standard in modelling the evolution of acute infectious diseases: (i) the adaptive SIR [1] for short-term predictions, and (ii) the ZSEIAR [2–4] In addition, **statistical time series models** (SARIMA [5]) of hospital admissions, severe hospital admissions and deaths are implemented. These models provide short-term point and interval predictions. The **segmented regression** [6] (SRM-R) of the reproduction number offers a flexible method for statistical analysis of time series trends, representing a compromise between standard stochastic ARIMA-type time series models and structured epidemiological models.

[1] https://opendata.mzcr.cz/

TECHNOLOGICAL BACKGROUND

The tool for modelling and reporting epidemic situations was designed as a web application. It is based on implementing a set of scripts in the R language that allow the calculation of four predictive models. The Shiny framework was used to standardise the format of the input datasets and subsequent configuration of the actual analysis of these models (setting of the parametric values of the models and their calibration). Thus, this platform provides a user-friendly interface for loading input datasets in a defined format, running the proposed algorithms for calculating predictive models, and visualising the results in graphs, tables and downloadable data. An integral part of this is a visualisation that combines the monitoring of inpatient care capacity with data defining the potential load on the system. This graphical tool summarises the use of these capacities in changes in disease epidemiology over selected time units (Figure 1).

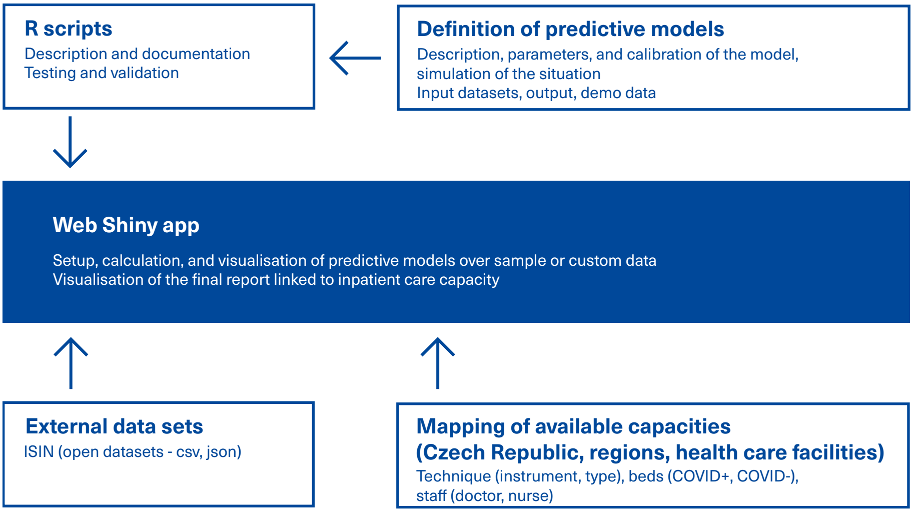

R scripts
Description and documentation
Testing and validation
Definition of predictive models
Description, parameters, and calibration of the model, simulation of the situation
Input datasets, output, demo data
Web Shiny app
Setup, calculation, and visualisation of predictive models over sample or custom data
Visualisation of the final report linked to inpatient care capacity
External data sets
ISIN (open datasets - csv, json)
Mapping of available capacities (Czech Republic, regions, health care facilities)
Technique (instrument, type), beds (COVID+, COVID-), staff (doctor, nurse)

Figure 1: Web platform communication schema

For the design, development, testing, validation, implementation and maintenance of the web application, the following tools enabling the effective collaboration of the expert team were used:

— Balsamiq Wireframes[2] – a desktop application for designing screen layouts for web applications;

[2] https://balsamiq.com/wireframes/

- R studio[3] – a desktop scripting application for the R programming language designed for statistical data analysis and graphical display;
- shinyapps.io by R Studio[4] – a platform for creating interactive web applications within the R environment;
- analytical applications (MS Excel, SPSS) for validating results.

RESULTS

BUSINESS UNDERSTANDING

COVID-19 has a basic reproduction number many times higher than influenza, which inherently means a much more dynamic and rapid increase in infected people. This also directly results in many times higher absolute numbers of people infected at one time with the potential need for hospitalisation. Therefore, there was a need for a unique online tool that combines modelling and real-time monitoring of healthcare system management in the Czech Republic. This case study put the models and their predictions into context in scenarios with accurate inpatient and ICU capacity. The following factors were the primary motivation for designing and developing a unique web-based application for real-time monitoring, analysis, and management of epidemic situations: (i) information on health system capacities and care needs in an evolving infectious disease epidemic, (ii) mapping and managing these capacities of the ICU system, (iii) appropriate use of analytical tools using deterministic and stochastic models with innovative features that go beyond current standards in infectious disease modelling.

There was no online analytical tool integrating acute or intensive care capacities with real-time epidemiological models of the evolution of the disease of interest. The new web-based application enabled rational and coordinated decision-making, e.g. at the level of individual regions or healthcare facilities within an area, including the integrated emergency system, through detailed reports on the dynamics of the use of relevant capacities. This tool has thus enabled the coordination of information related to critical national infrastructure in epidemic situations such as the spread of COVID-19. The mathematical innovation of the project included the enrichment of mathematical models of the reach of infectious viral diseases (deterministic and stochastic) with new elements; in particular, the inclusion of asymptomatic or otherwise undetected cohorts and the consideration of the infectiousness of infected individuals

[3] https://www.r-project.org/

[4] https://www.shinyapps.io/

before the manifestation of symptoms. The main outputs involved predictions of the evolution of the epidemic and available capacity, both short-term and long-term, including ongoing evaluation of restrictive measures, which could support the selection of appropriate measures at the moment and in the long term and thus contribute to the control of the epidemic.

DATA UNDERSTANDING

The understanding of the input data in this report is based on the four open datasets that are published within the Ministry of Health catalogue.

COVID-19: Online control centre for intensive care – free capacity contained reports on instrumentation (ECMO – extracorporeal membrane oxygenation, mechanical ventilation, CRRT, IHD, ventilators - portable and those in operating theatres)) and bed occupancy (A&E+ICU adult beds only and standard oxygen beds for the entire hospital). The numbers of beds that allowed the provision of the type of care, i.e. had the necessary staff and adequate material and technical support, were regularly updated. The staffing capacity was not separately recorded; it was always linked to the available bed capacity. The so-called reprofiled capacity was monitored as well, i.e. beds that had been originally (in the ordinary course of the healthcare facility) intended to provide a different type of care or care of a diverse speciality/expertise.

- Date
- Region in CZ-NUTS format
- Device: ECMO - Number of ECMO devices in the A&E+ICU - Available capacity / total capacity
- Device: Mechanical ventilation - Number of devices at A&E+ICU - Available capacity / total capacity
- Device: CRRT (continuous dialysis) - Number of continuous dialysis machines - Available capacity / total capacity
- Device: IHD (intermittent dialysis) - Number of intermittent dialysis devices - Available capacity / total capacity
- Device: Ventilators (portable) - Number of portable/transport ventilators - Available capacity / total capacity
- Fans (operating room) - Number of fans in operating rooms - Available capacity / total capacity

- Beds (A&E+ICU adults) - Number of beds in the anaesthesia and resuscitation department - Available capacity divided into COVID-19 positive and COVID-19 negative/total capacity
- Beds (standard with oxygen) - Of the total number of standard acute beds, the number of beds to which oxygen is supplied - Available capacity with breakdown into COVID-19 positive and COVID-19 negative/total capacity
- Reprofiled capacity: IP beds (Mechanical ventilation +/-) - beds with and without UPV option - Available power / actual capacity / planned capacity
- Reprofiled capacity: Standard beds with oxygen - Available capacity / actual capacity / planned capacity

COVID-19: positives, cures and deaths at district level provided aggregated data on the number of positive, cured deaths on each day stratified by the district of residence, sex and age categories based on reports from district health stations, laboratories and hospitals.

- Date (positivity, cure or death)
- District in CZ-NUTS format
- Sex
- Age categories 0–19, 20–64 and 65+ years
- Number of new positive cases
- Number of cured
- Number of deaths
- Number of deaths per hospital admission
- Number of deaths with recorded complications

COVID-19 dataset: time efficiency of testing – a dataset at the level of anonymised records of positive patients already processed by the KHS and containing time data on the interval between symptoms of the disease, collection, test result, KHS reporting and isolation of the positive patient based on reports from regional health stations and laboratories.

- Date of the first symptom
- Date of collection
- Date of first positive PCR test result

- Date of KHS report
- Date of isolation

COVID-19: hospitalised patients dataset provided aggregated data on the number of hospitalised patients with aggregation by county of stay, gender, age category, presence of the complicating condition, and patient outcome according to hospital data supplemented with KHS data.

- District in CZ-NUTS format
- Sex
- Age category 0–19, 20–64 and 65+ years
- Complicated condition during hospitalisation
- Outcome status at the date of export: Discharged, still hospitalised, death during hospitalisation
- Number of patients

DATA PREPARATION

In this particular case study, the data preparation phase was trivial. All input data did not need to be modified significantly because the application processed only open datasets directly in a uniform and standardised format. Therefore, the data structure was kept the same, and if the user wanted to work with his data, he always had to follow the prescribed format strictly.

MODELLING

WEB APPLICATION DESIGN AND DEVELOPMENT

All basic web application features were described through wireframes (Figure 2). The schematic layout of the functional elements on the individual pages provided a precise specification for the subsequent development. The user was guided step by step through the complete modelling process in an attempt to simplify the usage and overall navigation of the application as much as possible:

1. Selecting one or more available predictive models.

2. Loading sample data or custom datasets to simulate epidemics in the appropriate structure.

3. Setting the parameters and calibrating the selected models.

4. Visualise the results of the selected predictive models (graph, table, source data in .csv format).

5. Loading sample data or custom datasets to the bed capacities of the system in the appropriate structure.

6. Reporting the results provides insight into the use of inpatient care capacity about changes in the epidemiology of the disease of interest according to selected predictive models.

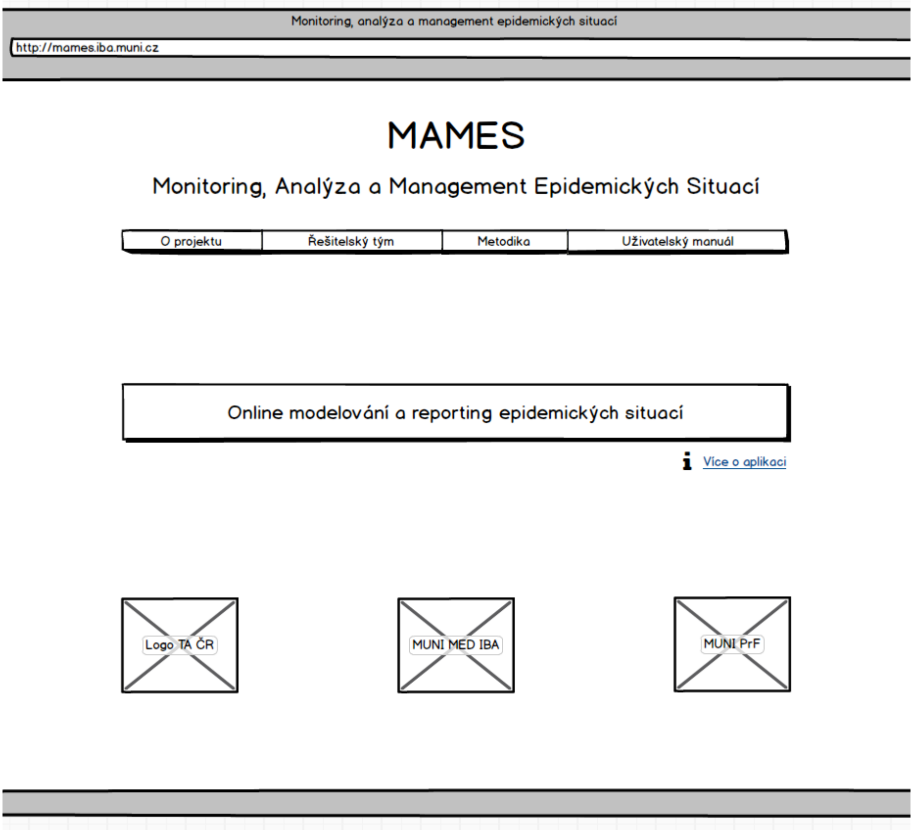

Monitoring, analýza a management epidemických situací
http://mames.iba.muni.cz
MAMES
Monitoring, Analýza a Management Epidemických Situací
O projektu
Řešitelský tým
Metodika
Uživatelský manuál
Online modelování a reporting epidemických situací
Více o aplikaci
Logo TA ČR
MUNI MED IBA
MUNI PrF

Figure 2: Homepage wireframe (available only in Czech)

The output of the predictive models in step 4 and the subsequent input of the reporting tool were the following characteristics:

- Number of newly diagnosed patients with the disease under study (COVID-19).
- Number of currently hospitalised patients.
- Number of currently hospitalised patients requiring intensive care (ICU).
- Number of deceased patients.

The output that could be visualised together with the capacity data, i.e. data on available standard and intensive care beds, were the following characteristics:

- Number of currently admitted patients.
- Number of currently admitted patients requiring intensive care (ICU).

Using the input data described above, this simple system highlighted the inadequacy of existing capacity in the identified trend in developing the disease under study. The application has a folder structure, which follows widely accepted recommendations, and generally uses good practices in designing, developing and implementing web applications using the R language and its Shiny library (the main script app.R consists of three essential parts).

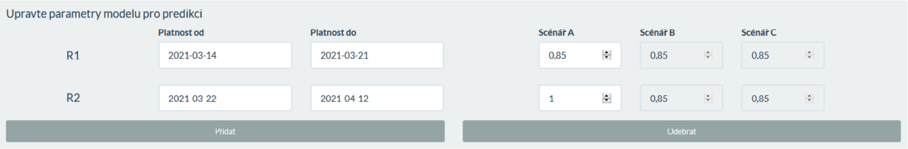
- Loading the global R script includes loading all the libraries needed to run the application, loading global variables, and loading all the predefined functions used on the server side. This section also defines the application's core functions containing calculating the individual predictive models.
- User interface definition. This part of the script defines all elements related to the appearance and user elements of the application. The user interface is divided into several tabs, each represented by a sub-script (scripts with UI_ prefix). These scripts are then called within the main application script. The R language does not allow defining all the necessary user elements, so some parts of the code are written in HTML. The final appearance of the application elements is modified using the theme.css cascading stylesheet.
- Definition of the server part, which contains the so-called logical part of the application. All data operations are defined here, from data retrieval to the definition of the resulting graphs with predictions. Furthermore, the server part contains all the code ensuring the functionality of the user interface. This part is called sub-scripts with the Server_ prefix, where longer code sections are moved for clarity of the central part of the script.

PREDICTIVE MODELLING

It was necessary to set its parameters individually for each of the four implemented models. Each model had a predefined default setting for predictions, but it was possible to change this setting. An example is the setup of the adapted SIR model for short-term forecasting, which allows working with one to three different forecasting scenarios. This model does not primarily work with a population of cured patients, so the end state is diagnostic confirmation of the presence of the disease of interest (state R corresponds to the number of newly confirmed positive individuals we assume are in quarantine), not cure. The first step was to retrieve accurate open data on the COVID-19 epidemic in the Czech Republic. The following set of characteristics was shown in the form of a data table: date, number of positive cases, number of newly diagnosed patients with the disease under study (COVID-19), number of currently hospitalised patients, number of presently hospitalised patients requiring intensive care (ICU), number of deceased patients. The next step was to select one of the available scenarios set by default according to the primary SIR model optimisation result. However, the model also allowed the user to work with other scenarios of the possible evolution of the COVID-19 epidemic:

- The adequate reproductive number in the SIR model reflected the average number of additional persons directly infected by a single infected individual. This parameterised the disease's infectivity and considered the measures taken over time, and the immunity acquired.
- The risk reflected the likelihood that a patient who developed symptoms of COVID-19 would subsequently be hospitalised due to their medical condition.
- The risk of the need for intensive care reflected the likelihood that a hospitalised patient would deteriorate enough on a given day to require intensive care, i.e. would require a transfer from a standard ward bed to an intensive care unit (ICU) bed.
- The risk of death was the probability that a patient who had developed symptoms of COVID-19 would subsequently die.

For example, we wanted to set up a scenario for the model where the reproduction number R was unchanged for the following week, and R increased to 1 after one week. R was set to 0.85 for the entire 30-day prediction. This default setpoint was automatically filled in from the last known value of R and therefore did not need to be adjusted in our assumed scenario. However, additional discounts could be added. For example, the truth was be from 22 March 2021 to 12 April 2021, and the R-value would be 1 (Figure 3). The end of validity of the previous reproduction number did not need to be changed; the value would adjust itself so that the days follow each other. The other parameters of the model, i.e. the risk of hospitalisation, the risk of needing intensive care and the risk of death, could be adjusted similarly.

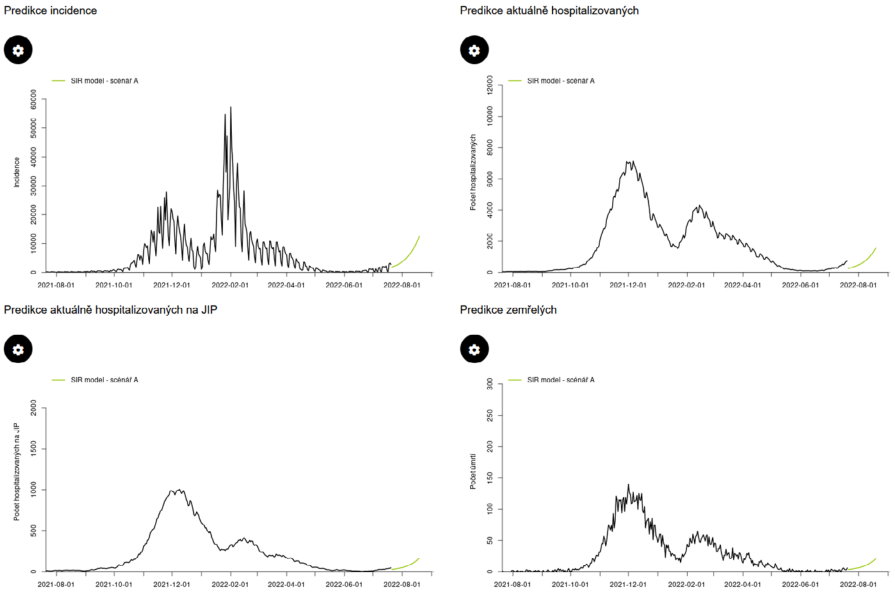


Upravte parametry modelu pro predikci

	Platnost od	Platnost do	Scénář A	Scénář B	Scénář C
R1	2021-03-14	2021-03-21	0,85	0,85	0,85
R2	2021 03 22	2021 04 12	1	0,85	0,85

Figure 3: Adjusting the reproduction number in the SIR model (available only in the Czech)

The results of modelling the COVID-19 epidemiological data using the selected SIR model are presented below (Figure 4). This tab allowed a visual check of the correctness of the settings of individual parameters of the models chosen under all defined scenarios. If the user was unsatisfied with the model results, they could modify the input model parameters.



Predikce incidence

Predikce aktuálně hospitalizovaných

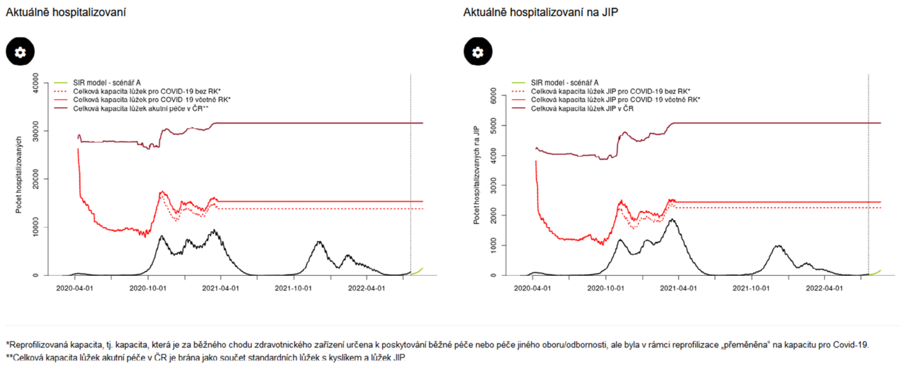
Predikce aktuálně hospitalizovaných na JIP

Predikce zemřelých

Figure 4: Visualization of various predictions based on the selected SIR model (available only in the Czech)

One of the other main objectives of the application was to relate the results of the predictive models to data on inpatient care capacity, namely total bed capacity, bed capacity for COVID-19 patients (including reprofiled beds), both for standard beds and intensive care unit (ICU) beds. The application enabled the retrieval of accurate data containing daily summaries of changes in the available capacity in acute inpatient care hospitals about the spread of COVID-19 in the Czech Republic. Alternatively, users could use their data, which had to be in a defined structure corresponding to the open dataset's format.

The graphs below (Figure 5) represent a common visualisation of the available capacities and outputs of the selected mathematical models. While in the case of the number of patients admitted to the hospital, real and modelled values were shown, the capacity data were not modelled. We assumed a constant evolution for the prediction horizon. Data were also available in the form of the data table and summary files in CSV format to download.


Aktuálně hospitalizovaní
SIR model - scénář A
Celková kapacita lůžek pro COVID-19 bez RK*
Celková kapacita lůžek pro COVID-19 včetně RK*
Celková kapacita lůžek akutní péče v ČR**
Počet hospitalizovaných
0
10000
20000
30000
40000
2020-04-01
2020-10-01
2021-04-01
2021-10-01
2022-04-01
Aktuálně hospitalizovaní na JIP
SIR model - scénář A
Celková kapacita lůžek JIP pro COVID-19 bez RK*
Celková kapacita lůžek JIP pro COVID-19 včetně RK*
Celková kapacita lůžek JIP v ČR
Počet hospitalizovaných na JIP
0
1000
2000
3000
4000
5000
6000
2020-04-01
2020-10-01
2021-04-01
2021-10-01
2022-04-01

*Reprofilizovaná kapacita, tj. kapacita, která je za běžného chodu zdravotnického zařízení určena k poskytování běžné péče nebo péče jiného oboru/odbornosti, ale byla v rámci reprofilizace „přeměněna" na kapacitu pro Covid-19.
**Celková kapacita lůžek akutní péče v ČR je brána jako součet standardních lůžek s kyslíkem a lůžek JIP.

Figure 5: Reporting on the available capacities and outputs of the selected mathematical models (available only in the Czech)

EVALUATION

All the achieved results from individual predictive models were thoroughly validated in collaboration with the mathematical and analytical teams involved in this case study. Given the complexity of the calculations of each model, it was necessary to verify the correctness of each step in the modelling and visualisation process. User testing of the web application was also an integral part of the process and was conducted across the development team. Emphasis was placed on completing the specification according to the designed wireframes, simplicity and functionality of all navigation elements, accessibility of the application and overall user-friendliness.

DEPLOYMENT

The web application for Monitoring, Analysis and Management of Epidemic Situations (MAMES)[5] allowed the user to prospectively model the COVID-19 epidemic in the Czech Republic about the expected need for hospitalisation of patients, including estimating critically ill patients in direct relation to intensive care capacities. In addition to static content, where theoretical and descriptive information was published, the application implemented a module for modelling epidemic situations. It offered four robust predictive models (adapted SIR model for short-term predictions, epidemiological model ZSEIAR, time series model and segmented regression model of the reproduction number) that worked with the total number of patients, newly diagnosed, newly hospitalised and the structure of the population concerning the different stages of the disease, including the estimation of critically ill patients in direct relation to intensive care capacities. In general, this application targets users with expert knowledge of infectious disease behaviour and epidemiology and can navigate the parametric description of individual models. It is easy to select suitable predictive models in a few steps, populate them with sample or custom data inputs according to a defined structure, set the necessary parameters together with model calibration, and then visualise the results. The application also includes linking the prediction results with another web-based application that monitors the potential burden on the healthcare system by monitoring inpatient care capacities. As mentioned earlier, the target audience is primarily interested professionals involved in predictive modelling of COVID-19 epidemic data. In-depth knowledge of data analysis, modelling, and basic epidemiology is required to use the MAMES project's proposed outputs properly.

DISCUSSION

The output of the case study in the form of a web application allowed the general public to model the evolution of the COVID-19 epidemic in real-time and thus responded to the need at the time for rapid availability of adequate information related to the spread of COVID-19. Due to the implementation of flexible epidemiological models, the application is not strictly linked to COVID-19 but can be adapted to other epidemic periods and situations. A supporting application for monitoring the potential burden on the health system is available, including source codes, and can be further developed to meet the current user needs. From the perspective of the society-wide application, the target group of the

[5] https://mames.iba.muni.cz/

application is a wide range of users from the state administration authorities (Regional Public Health Authorities, National Institute of Public Health, Ministry of Health), health service providers (hospitals and other healthcare facilities) and their founders (Ministry of Health and other central authorities, regions and others), as well as other academic and non-academic institutions.

EVALUATION OF THE AIMS OF THE CHAPTER

- To understand and integrate various data sources describing the epidemic caused by COVID-19 in the Czech Republic.
  - The prerequisite for the subsequent modelling phase was the detailed adoption of all relevant input data. All open datasets are described through a metadata schema in the open data catalogue. Subsequent integration in real-life modelling applications was possible because the team became familiar with the data.

- To propose models based on methodologies allowing short- and medium-term epidemic predictions and long-term scenarios.
  - The existence of methodological background allowed the proposal of detailed documentation for prospective modelling of the population burden of COVID-19, including the structure determined by disease stage and other modelling factors. The relevant open datasets have been appropriately pre-processed to be correctly used as input to the models.

- To develop an interactive web application for modelling the future evolution of the COVID-19 epidemic in the Czech Republic about the expected need for the hospitalisation of patients.
  - The proposed algorithms for the predictive models were suitably optimised and implemented into a web application[6] for monitoring epidemic situations using the R package Shiny application concerning the possibility of processing information on bed capacities of the health system. The primary added value is the possibility of setting input values and parameters for individual predictive scenarios on the user side and visualising the results in tables and graphs.

[6] https://mames.iba.muni.cz/

LESSONS LEARNED

The academic collaboration at Masaryk University focused on the design and development of an innovative platform for monitoring and prediction in crisis situations, specifically focusing on the COVID-19 epidemic. Thanks to the experience of the teams involved, the understanding phases were relatively quick; all the more attention was paid to working with the selected models, which were tested in practice. The result was an interactive application with the possibility of custom data import and display, which was very positively evaluated during the project evaluation. The interactivity of the application was ensured by the R Shiny technology, with which it was also possible to set many variables concerning the current or predicted state of the epidemic.

In this case, open data helped as a data basis for a stand-alone and free-standing application, which, however, was highly innovative.

REFERENCES

[1] Májek O, Pavlík T, Dušek L. Adaptive SIR model for short-term predictions [Internet]. Brno: Masaryk University; 2021 [cited 6 Jun 2023]. Available from: https://is.muni.cz/www/98951/47857356/mames/MAMES_SIR_metodika.pdf.

[2] Přibylová L, Hajnová V. Model ZSEIAR [Internet]. Brno: Masaryk University; 2021 [cited 6 Aug 2023]. Available from: https://is.muni.cz/www/98951/47857356/mames/MAMES_ZSEIAR_metodika.pdf.

[3] PAQ research IA. Life in pandemic: How many people were the respondents in closer contact with? [Internet]. 2020 [cited 20 Dec 2021]. Available from: https://zivotbehempandemie.cz/kontakty.

[4] Google. Google Mobility reports [Internet]. 2020 [cited 24 Jun 2023]. Available from: https://github.com/ActiveConclusion/COVID19_mobility/blob/master/google_reports/mobility_report_europe.xlsx.

[5] Kraus A, Kraus D. Statistical time series model [Internet]. Brno: Masaryk University; 2021 [cited 12 May 2022]. Available from: https://is.muni.cz/www/98951/47857356/mames/MAMES_SARIMA_metodika.pdf.

[6] Uher M. Segmented regression model of the reproduction number [Internet]. Brno: Masaryk University; 2021 [cited 9 Jun 2023]. Available from: https://is.muni.cz/www/98951/47857356/mames/MAMES_SRM_metodika.pdf.

SECTION C

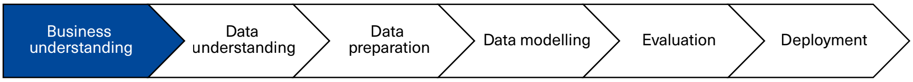
| 18 |

A PILOT TOOLBOX AGAINST HEALTH MISINFORMATION THE CZECH REPUBLIC IN THE COVID-19 AGE

Václav Moravec, Jakub Gregor, Radek Mařík, Ivan Vodochodský, Ladislav Dušek, Martin Komenda

The case study was supported by the Technology Agency of the Czech Republic, project no. TL04000176 "COVID-19 infodemic: an AI communication platform suppresses infodemic relation to journalistic and media ethics".

CRISP-DM CRUCIAL PHASES



GENERAL INFORMATION

Year	2020–2022
Keywords	Infodemic, misinformation, COVID-19, pandemic, media
Research question	How can disinformation be systematically countered using correctly interpreted data?
Type of result	Static analytical report
Level of data processing	Advanced analyses

DATA TO DOWNLOAD

INTRODUCTION

One month after the Chinese authorities released the first information about a hitherto unknown respiratory disease that affected dozens of people in Wuhan at the end of 2019, the World Health Organization (WHO) issued Situation Report No. 13, which contained a short chapter "Managing the 2019-nCoV 'infodemic'" [1]. The Introduction of the document of 2 February 2020 stated: "The 2019-nCoV and response have been accompanied by a massive 'infodemic' – an over-abundance of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance when they need it." Two weeks later, WHO Director-General Tedros Adhanom Ghebreyesus said at the Munich Security Conference that the world is fighting not only an epidemic but also an infodemic, because "fake news spreads faster and more easily than this virus, and is just as dangerous" [2]. Since then, the relatively unknown term "infodemic" became part of the COVID-19 pandemic news coverage, as well as of many scientific studies across disciplines, from medicine to media and journalism studies. The two years of the COVID-19 pandemic have made it possible to study and evaluate the phenomenon of infodemic in the countries that are among the most affected in Europe. One of them is the Czech Republic, which has been reported as one of the top five positions in COVID-19 – related incidence, hospitalisation, and mortality rates [3]. Last but not least, the vaccination rate should be mentioned as an indicator highly sensitive to the infodemic and as a measure of population sensitivity to various types of information disorders (e.g. disinformation and misinformation). And again, the Czech population is worth to be studied as it stands with its reached standardised vaccination rate nearly in the middle of EU countries.

It would be wrong to derive the notion of infodemic from the onset of the COVID-19 pandemic. Its roots are older and date back to the beginning of the new millennium and the advent of the SARS epidemic. The authorship of the concept is attributed to Gunther Eysenbach [4], who became the promoter of the new scientific discipline of infodemiology, and to David Rothkopf [5], who used it as a journalistic metaphor for the phenomenon permeating the world of networked digital media. There is no doubt that its deeper conceptualisation in many disciplines can contribute to solving interdisciplinary problems. With the cooperation of individual professional fields (and with the preservation of deep scientific analysis and proposed solutions), it is possible to face such complex phenomena, which is contained in the metaphor of infodemic.

An effective combination of good journalistic practice, medical expertise and modern information and communication technologies makes it possible to identify selected forms of information distortion in various media. These include, for example, inadequate or incorrect work of journalists with data, wrong choice of reference examples and comparisons, or the need for more knowledge of medical terminology. The course of individual waves of the COVID-19 epidemic in the Czech Republic offers a unique input suitable for further analytical treatment. The methods and techniques of the social sciences, humanities and arts presented here allow us to understand the leading causes of an infodemic. The result is a knowledge base for innovations in journalism education and the redefinition of journalistic and media ethics in the Czech Republic about the medical field, which is one of the tools to mitigate or minimise the impact of current and future infodemics.

AIMS

- To identify the essential characteristics influencing a digital infodemic in the Czech Republic about detailed knowledge of media communication and the performance of key stakeholders on national and regional levels.
- To find approaches to systematically eliminating false or misleading information during a pandemic? If so, how?

METHODS

A systematic approach in the form of a pilot toolbox facing health misinformation in the Czech Republic during the COVID-19 epidemic is based on a proven methodological background. The information "cake" model (Figure 1) provides the very first broad roadmap on how to fight an infodemic [6], which supplements the World Health Organization (WHO) framework [1]. This model introduces four pillars of infodemic management: (i) information monitoring (infoveillance); (ii) building eHealth literacy and science literacy capacity; (iii) encouraging knowledge refinement and quality improvement processes such as fact-checking and peer-review; and (iv) accurate and timely knowledge translation, minimising distorting factors such as political or commercial influences.

Especially fact checking, peer review and the circulation and exchange of information between four levels (social media, news media, policy and health care practice, and science), including a correct interpretation, are crucial for significant improvement of information delivery to the professionals and the general public.

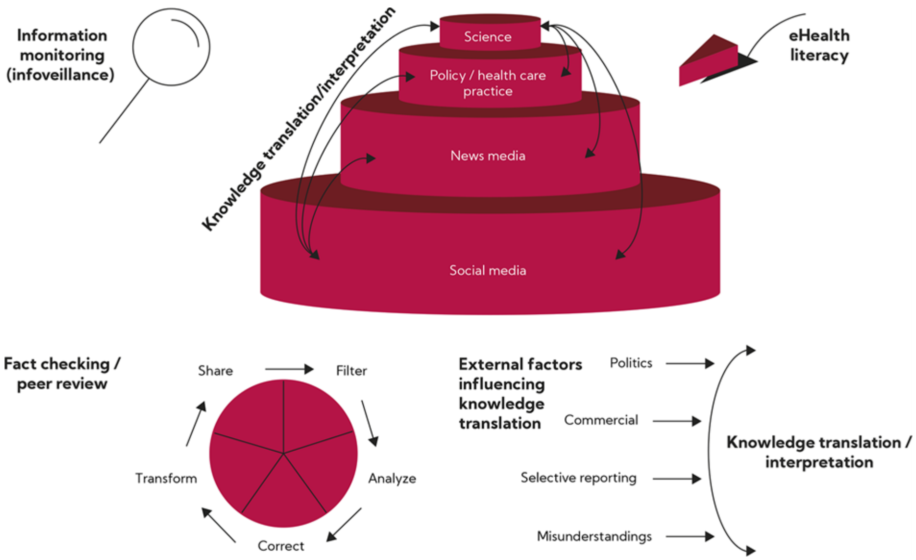

Information monitoring (infoveillance)
Knowledge translation/interpretation
Science
Policy / health care practice
News media
Social media
eHealth literacy
Fact checking / peer review
Share
Filter
Analyze
Correct
Transform
External factors influencing knowledge translation
Politics
Commercial
Selective reporting
Misunderstandings
Knowledge translation / interpretation

Figure 1: The information "cake" model describing the basics of infodemic management according to Eysenbach

To design and implement a pilot toolbox against health misinformation in the Czech Republic in the COVID-19 age, we combined the basic idea of infodemic management in the form of the information "cake" model. Our interdisciplinary team focused on a normative framework for news media and journalism, with analytical data processing and analysing media coverage of selected topics in medical terminology. Moreover, the National Health Information Portal improving health literacy, supported by health open data driven by state authorities, is also emphasised as a needed component.

RESULTS

BUSINESS AND DATA UNDERSTANDING

With a thorough understanding and knowledge of the government, media, journalistic and scientific environment related to COVID-19 communication in the Czech Republic, we proposed an information matrix (Table 1) that contains three crucial groups (expert associations and entities, official government tools, and media) about the particular pillar of the information "cake" model. The primary objective of this matrix is to identify how the comprehensive model of infodemic management is covered in the Czech Republic and where are its main issues and gaps. At first glance, it is noticeable that some domains are only partially covered (as a pilot project activity) or not covered at all. Within the case study below, the activities and results related to each pillar are localised to the Czech environment and described in detail. It is important to note that an institutionalised platform under government supervision is needed to support the systematic fight against misinformation and misinterpretation.

The COVID-19 pandemic has been unprecedented in terms of the requirement to transfer scientific knowledge as quickly as possible to policymakers, stakeholders, the media, and the general public worldwide. In the case of the Czech Republic, it can be stated that all the groups involved needed to prepare for this situation. This was particularly evident in the second and subsequent waves of the epidemic (i.e. from autumn 2020 onwards) when the latest scientific knowledge and information reached the target groups with a high degree of inaccuracy and "noise" without proper context and with inaccurate interpretation. The individual pillars in the Czech environment are described below.

Table 1: The matrix combines four pillars of infodemic management and involved stakeholders and tools driving the publishing of information about the COVID-19 epidemic in the Czech Republic

Pillar	**Experts, researchers**	**Government, Ministry of Health**	**Media**
P1: Facilitate Accurate Knowledge Translation	Snow Initiative A multidisciplinary group of experts (civic activity) https://www.iniciativa-snih.cz/ Interdisciplinary Group for Epidemic Situations A multidisciplinary group of experts (ministerial advisory board) https://www.meses.cz/ National Institute for Pandemic Management A multidisciplinary group of experts (ministerial advisory board)	Actual Information about COVID-19 + COVID Portal Official website of the Ministry of Health and the Government – actual situation, restrictions, professional guidelines, recommendations for life situations https://koronavirus.mzcr.cz, https://covid.gov.cz Disease at the Moment Official website of the Ministry of Health and the Government – epidemiology https://onemocneni-aktualne.mzcr.cz/covid-19 Czechia Vaccinates Official website of the Ministry of Health and the Government – vaccination https://www.ceskoockuje.cz Traffic Light System + Anti-epidemic System Guidelines for the implementation and lifting of anti-epidemic measures Analytical reporting services Reports and analyses for epidemic management, media, and public	Public broadcasting Czech Television, Czech Radio, Objective news, educational programs, interactive data analyses "Mobile Radio" radio A platform for communication between local authorities and citizens using official data

Pillar	Experts, researchers	Government, Ministry of Health	Media
P2: Knowledge Refinement, Filtering, and Fact-Checking	Epidemiologists, researchers, and experts in social media	Open data Official datasets published in open or semi-open mode (incidence, hospitalisations, deaths, vaccination) https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19 Smart Quarantine A set of tools and applications for data collection and visualisation at all levels of the epidemic management	Demagog.cz, Manipulatori.CZ, Cesti-elfove.cz Independent civic fact-checking platforms https://demagog.cz, https://manipulatori.cz, https://cesti-elfove.cz "Ověřovna!" ("Verification room!") A fact-checking platform of the Czech Radio
P3: Build eHealth Literacy	National Health Information Portal Source of validated and credible information on health https://www.nzip.cz		--------------
P4: Monitoring, Infodemiology, Infoveillance, and Social Listening	Infomore.cz interdisciplinary project for coronavirus-related media analyses https://www.infomore.cz	Semantic Vision analyses Detection of disinformation and adversarial propaganda	Infomore.cz interdisciplinary project for coronavirus-related media analyses https://www.infomore.cz

FIRST PILLAR: FACILITATE ACCURATE KNOWLEDGE TRANSLATION

Experts and researchers

Dozens of experts and researchers in epidemiology, immunology, clinical medicine, and many other disciplines have commented on coronavirus during the epidemic in the Czech Republic. Discussions from scientific conferences, including several conflicting views and claims, thus spilt over into the public domain, which of course, did not positively affect how the public perceived the latest scientific findings.

Government, Ministry of Health

The website koronavirus.mzcr.cz became an official communication platform of the Ministry of Health, which provided official and up-to-date information on current measures, testing, vaccination, information for health professionals etc. It was then followed in a broader and more parameterised form by the COVID Portal[1], which aimed to describe specific life situations during the epidemic and how to deal with them in more detail. Current daily epidemiological data and datasets have been presented on the Disease at the Moment platform [7]. Throughout the epidemic, there has been an effort to link the measures and restrictions taken to the actual trends of the epidemic situation and to communicate clearly under what circumstances they would be lifted or tightened (e. g. in terms of restrictions on the movement of people, school closures, wearing of masks, mandatory testing, holding of group events, etc.). It should be noted, however, that neither of them was of long duration, and epidemic management on the governmental level has instead been guided by ad hoc adopted sets of measures.

Media

During the COVID-19 epidemic, the transfer of the most accurate knowledge from scientific discourse to the field of journalism was ensured by specialised scientific editors of public service media from Czech Television and Czech Radio. These public service media provided context and educated the public with information on significant news programs and specialised programs (e.g., "Earth in Need" on Czech Television or "Science Plus" on Czech Radio). The public appreciated reputable continuous news and public service media programs during the pandemic. After the first half of the COVID-19 epidemic, people identified the public service media as the second most trusted source of information about this severe health crisis after health experts. After half a year of the COVID-19 epidemic in the Czech Republic, 76% of the population trusted health experts as sources of information, 69% trusted the public service media, 52% trusted the Minister of Health, 45% trusted the online media, and 36% trusted the Prime Minister [8]. Moreover, thanks to the initiative of "Mobile Radio" in cooperation with the technology start-up MAMA AI and the Institute of Health Information and Statistics of the Czech Republic, a tool for citizens of individual municipalities with precise statistics on the ongoing epidemic in the Czech Republic was designed and deployed. Data in this reporting system are validated and, where needed, accompanied by explanatory notes to ensure a correct interpretation.

SECOND PILLAR: KNOWLEDGE REFINEMENT, FILTERING, AND FACT-CHECKING

Experts and researchers

Epidemiologists and other experts became the new stars during the COVID-19 pandemic. They received much coverage in the news media, and many used

[1] https://covid.gov.cz/

the modern phenomenon of social networks. However, this proved to be a double-edged weapon. On the one hand, they accelerated the transfer of verified and valid knowledge, and many experts helped to debunk the proliferating misinformation and myths on social media. On the other hand, it also led to sharing of preliminary and unverified information and findings that were later proven false or exaggerated.

Government, Ministry of Health

The so-called Smart Quarantine tools are software solutions used as a crisis management system during the COVID-19 epidemic in the Czech Republic. Specifically, these include a central registry for infectious diseases; an application for collecting data and information from laboratories, sampling points and health stations; an online Dashboard for up-to-date visualisation of the development of the epidemic; a situational map; a mobile application for monitoring persons tested positive for COVID-19; and linking isolated tools to get a comprehensive picture of the current situation in the Czech Republic. These centrally coordinated platforms provide large volumes of data. However, not all of them are intended to be shared with the general public, as their primary purpose is to support crisis management at national and regional levels. The publication of only valid information from these systems under the Ministry of Health's umbrella towards the public concerning the COVID-19 epidemic in the Czech Republic has proved crucial from the beginning. This information in open data has been published continuously according to the evolution of the epidemic and various requests of interested stakeholders in the National Catalogue of Open Data. From 1 March 2020 to 1 April 2022, a total of 57 open datasets related to COVID-19 have been published and widely used by the general public, scientists, public authorities and decision-makers [9]. The global aim is to provide complete factual information, including correct interpretation, for further communication fully guaranteed by the national authority.

Media

In journalism, an essential role in verifying facts is played not only by traditional news media, whose routine procedures ensure that published information has been confirmed from at least two relevant sources, but also by a relatively new area of journalism called fact-checking journalism. According to Luengo and García-Marín [10], fact-checking journalism directly responds to the growing spread of disinformation and misinformation in networked digital media. The traditional news media (e.g., news agencies such as AFP and Reuters or public service media such as BBC and Deutsche Welle) systematically focus on refuting disinformation and verifying the factual correctness of selected informational errors in public space. Fact-checking journalism platforms take various forms – from specialised websites (e.g. AFP, BBC) to regular radio or television broadcasting sections. Before the COVID-19 epidemic, three media based on civic engagement (civic journalism) focused on developing fact-checking journalism in the Czech Republic, namely Demagog.cz, Manipulatori.cz and Cesti-elfove.cz. Their impact and influence on the audience are negligible compared to traditional news media, including public service media. During the COVID-19 epidemic, none of the public service media in the Czech Republic, not even Czech Television, the most trusted media brand in the Czech Republic, established a fact-checking journalism platform that would systematically verify and correct disinformation disseminated in digital media. In September 2021, Czech Radio set up a section "Ověřovna!" ("Verification room!") within their online portal irozhlas.cz. However, from September 2021 to December 2021, only five fact-checks related to the COVID-19 epidemic were published.

THIRD PILLAR: BUILD EHEALTH LITERACY

The Czech population is below average among European countries in terms of the overall level of health literacy and the proportion of people with low health literacy [11,12]. Health literacy development and continuous improvement concerns all three of these bodies (experts, government, media). For this reason, this chapter is conceived more generally. A systematic approach to educating the general population and providing validated and credible health information has yet to be established nationally. The National Health Information Portal (NHIP) has been developed since 2019 under the umbrella of the national authorities and professional medical societies. The COVID-19 epidemic delayed the launch of the NHIP; the portal was officially launched in July 2020. However, its operation is not a reaction to the COVID-19 epidemic/infodemic but rather an effort to provide a more comprehensive approach to health education for the Czech population. Somewhat surprisingly, this intention was also reflected by users. The coronavirus-related terms ranked third (behind "homecare" and "spa") from July 2020 to January 2022. The platform is an understandable, clear and user-friendly online source for a wide range of information on health, diseases and other medical conditions, proper health and disease care and services at the health-social interface. The published content is always carefully reviewed and validated by experts, stakeholders, and lay readers.

FOURTH PILLAR: MONITORING, INFODEMIOLOGY, INFOVEILLANCE AND SOCIAL LISTENING

The first Czech platform that tried to fulfil Eysenbach's process of continuous monitoring of the media landscape and the associated infoveillance during the COVID-19 epidemic was the portal Infomore.cz. It was an interdisciplinary project of experts in journalism and media studies from Charles University, health statisticians and computer scientists from Masaryk University, and media analysts from the NEWTON Media agency, which has the largest archive of media content in the Czech Republic. Since December 2020, the Informore.cz website has regularly published various quantitative and qualitative analyses. The quantitative analyses covered the most frequent words of media coverage in the form of monthly word clouds, individual elements of media panic based on the concept of the linguist Roger Fowler (e.g., inappropriate use of war metaphors, frequent occurrence of quantification rhetoric, or inappropriate use of medical terms), and the occurrence of individual disinformation narratives in social and traditional media. The qualitative analyses concerned the basic parameters of the functioning of disinformation resources on social networks or alternative framing of vaccination against COVID-19. Elements of machine learning were used for the analyses. Without them, infoveillance would not be possible.

DATA PREPARATION

During the two years of the COVID-19 epidemic (from 1 December 2019 to 31 January 2022), an extensive corpus of 2,485,724 media articles in the Czech language dedicated to this health crisis was created from the data of the Newton Media company (co-investigator of the Infomore.cz project[2]). The articles were collected from more than 4,000 media sources (traditional and online media). Based on this corpus, it is possible to reconstruct the media coverage of individual stages of the epidemic in the Czech Republic.

Data on the number of disinformation articles related to the COVID-19 epidemic was obtained from Semantic Vision's corpus. The company operates its own Open Source Intelligence (OSINT) system, which builds on elements of artificial intelligence, especially advanced semantic analysis and big data semantics. Semantic Visions collects and analyses 90% of the world's online news content. The corpus of disinformation narratives of the COVID-19 epidemic, on which our analyses are based, is the result of monitoring 4,313 websites in the Czech language. An article is included in the corpus if it is more than 75% based on manipulation and false claims.

The official source of information on COVID-19 epidemiology in the Czech Republic is the Information System of Infectious Diseases, operated by the Institute of Health Information and Statistics. Data on individual cases are entered by laboratories and Regional Public Health Authorities. Outputs from the system and open datasets are available on Disease at the Moment[3], the official website of the Ministry of Health for publishing data on the COVID-19 epidemic.

[2] https://www.infomore.cz/

[3] https://onemocneni-aktualne.mzcr.cz/covid-19

MODELLING

Through different types of media analysis, it is possible to point out good and bad media practices. The outputs in the form of word clouds and qualitative and quantitative content analyses aim to hold a mirror up to not only journalistic practice but also all who express themselves in the media space. Below, selected examples are briefly described.

ARTICLES IN THE EPIDEMIC'S FIRST PHASE

The number of articles devoted to the epidemic peaked in the epidemic's first phase (see Figure 2). Semantic Visions is a significant player in infoveillance in the Czech Republic. It is a software-based actionable analytics company based in Prague and London. It operates a military-grade Open-Source Intelligence (OSINT) system that collects and analyses 90% of the world's news content. Semantic Visions focuses on the detection of disinformation and negative propaganda. The company examines the occurrence of individual disinformation narratives on 4,313 websites in the Czech language. Semantic Vision regularly provided the results of its findings to the Ministry of Health of the Czech Republic. The ministry's work with these data was not systematic.

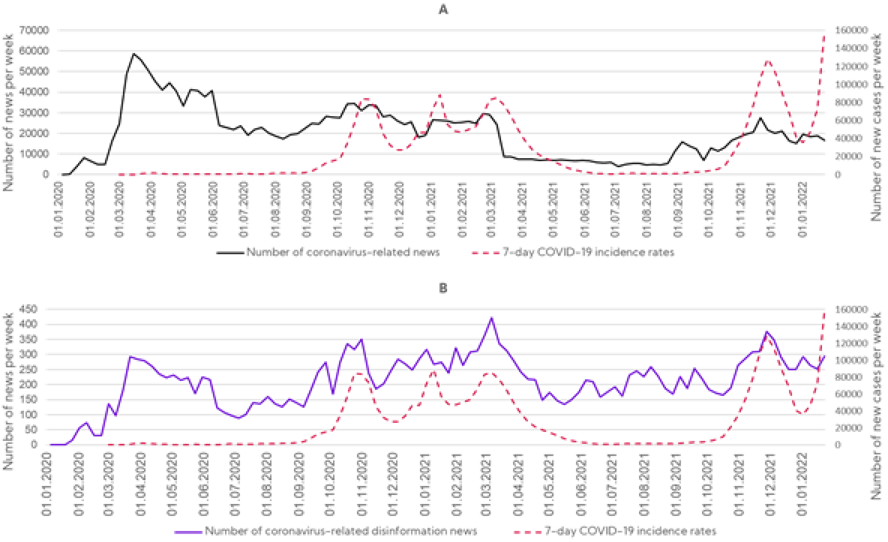


Figure 2: COVID-19 weekly incidence rates and weekly numbers of coronavirus-related news and disinformation news in Czech media. Data source: Newton Media, Semantic Visions, Institute of Health Information and Statistics of the Czech Republic

The highest media coverage occurred at the beginning of the COVID-19 epidemic, in March 2020; at times of the highest incidence and death rates, the number of contributions was less than half of the original coverage. On the other hand, the number of misinformation posts followed the course of the epidemic much more closely. The most mentioned conspiracy theories about the causes and origins of COVID-19 were the installation of 5G networks and Bill Gates. Other approaches received much less publicity – mainly about origins in Chinese or American laboratories or uncontrolled spread through migration [13].

WORD CLOUDS ABOUT COVID-19

Newton Media's media monitoring corpus on COVID-19 contained 1,153,999 messages as of 31 October 2020. By querying the identification of media sources, it can be determined that the news came from 3,392 different media sources. In addition to analysing all data, 358 references were selected from the total, and 7 clusters were created above them with specific media type characteristics:

- national and regional daily newspapers, including their websites (excluding tabloids),
- primary tabloid print media and leading tabloid websites,
- commercial television and their websites,
- the leading economic media and websites,
- commercial radios and their websites,
- news magazines and their websites,
- public service media and their websites.

All forms of the word coronavirus, which naturally dominated the clouds due to the topic of the news, were removed from the word clouds, taking space away from words related to the news, which were of particular interest to us.

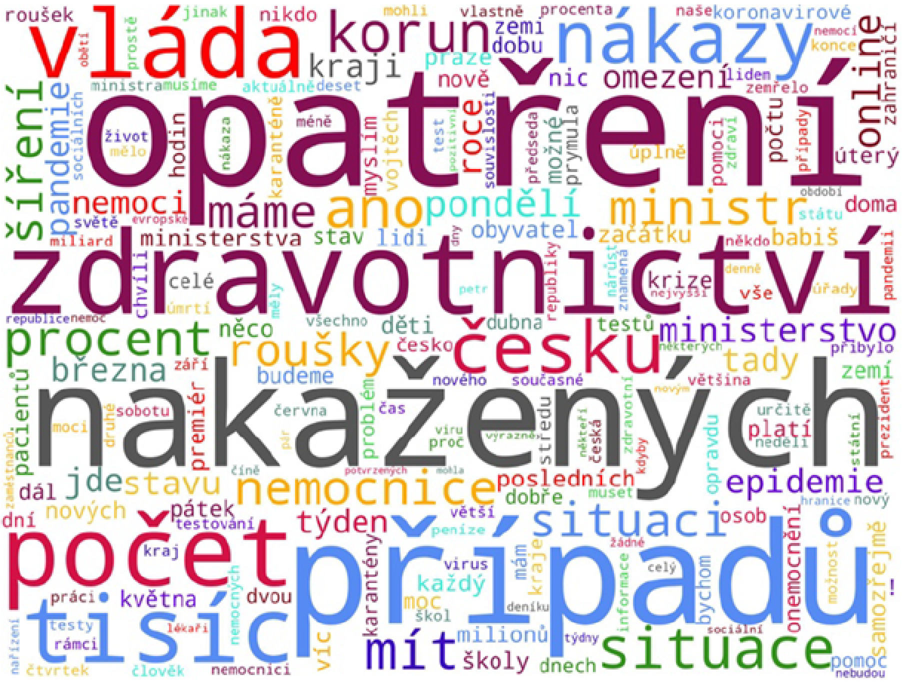


Figure 3: Word cloud of the words used in all monitored media coverage of COVID-19 between December 1, 2019, and October 31, 2020. Data source: the research team of the Infomore.cz project (TL04000176), available only in Czech

As seen in Figure 3, news coverage of COVID-19 disease in the overall news sample analysed was dominated by the words measures (opatření), infected (nakažených), cases (případů), healthcare (zdravotnictví), number (počet), government (vláda), infections (nákaza), and thousand (tisíc), between January and October 2020. The comments varied over time according to the evolution of the pandemic and the type of media. Since the most relevant in terms of the daily life of the population was the media coverage of different kinds of measures (which were brought by the government, especially by the ministers of health, while health and hospital capacity were another frequently mentioned topic), they were constantly being prepared and dynamically changing, so that they were the focus of attention for a long time, as were the numbers of infected. Especially in spring, when they were a symbol of solidarity, the word masks (roušky) occupied a prominent place. Discussions of the impact on the economy are illustrated by the numerals and the word crowns (korun). The complete analytical report is available at www.infomore.cz.

EVALUATION AND DEPLOYMENT

The individual analytical outputs, which can significantly help in the correct understanding of the presentation of facts, including the necessary interpretation, were always thoroughly validated in terms of content and visuals by members of the research team. Internal reviewers have involved health experts, computer scientists, and communication and media experts. The primary objective was the clarity and comprehensibility of the outputs subsequently published online for the general public without access restrictions. The results in the form of media analyses divided into word clouds, qualitative and quantitative content analyses, static educational articles, open datasets, and a unique dictionary of terms related to the COVID-19 epidemic are intended to support overall health literacy across the Czech population. Following the theoretical principles mentioned above, these results have been published on three portals interlinked to cover selected domains that present only objective and valid information.

DISCUSSION

Two years of the COVID-19 epidemic in the Czech Republic – one of the most affected countries in Europe – have shown that the metaphor of infodemic has not become an essential topic in the public sphere. It is even though an infodemic can be a symbol of a comprehensive approach to spreading excessive information, including disinformation and misinformation, about a severe health crisis, which may make it harder to find a solution to the crisis. In the corpus of media coverage of the COVID-19 epidemic in the Czech Republic, which amounts to a total of 2,485,724 articles published in the Czech language between 1 December 2019 and 31 January 2022 in more than four thousand media titles and brands (traditional and online media), the term "infodemic" appeared only in 487 of them. As a result, the Czech Republic lacks an institutionalised platform to support the fight against disinformation and erroneous interpretations. Systematic fact-checking is developed only by non-profit organisations with limited impact (Demagog.cz and Manipulatori.cz). The development of fact-checking journalism remains outside the attention of public service media, including Czech Television as the most trusted media brand in the Czech Republic. At the same time, the infodemic issue concerns not only health crises but crises in general (for instance, political or war-related), as demonstrated by the Russian occupation of Ukraine at the beginning of 2022. The Czech Elves, a civic initiative monitoring the Czech disinformation scene, pointed out the following: "During February, disinformation about COVID-19, which had completely dominated in recent months, was replaced by disinformation supporting Russia's claims to Ukraine and its subsequent brutal military attack" [14].

EVALUATION OF THE AIMS OF THE CHAPTER

- To identify the essential characteristics influencing a digital infodemic in the Czech Republic about detailed knowledge of media communication and the performance of key stakeholders on national and regional levels.
  - These characteristics are based on the information "cake" model [6], which illustrates the very first broad roadmap to fight an infodemic. A newly designed information matrix of COVID-19 infodemic management in the Czech Republic extended this approach and properly combines four fundamental pillars (facilitate accurate knowledge translation; knowledge refinement, filtering, and fact-checking; build eHealth literacy; monitoring, infodemiology, infoveillance, and social listening), involved stakeholders and tools (experts and researchers; government and Ministry of Health; media) driving publishing information about the COVID-19 epidemic.
- To find approaches to systematically eliminating false or misleading information during a pandemic? If so, how?
  - The selected projects, under the auspices of government organisations collaborating with health experts, computer scientists, and communication and media experts in the Czech Republic, provide or produce the necessary online tools and communication platforms with fully guaranteed and proven information. In this way, they significantly support broader education and health literacy improvement of the general public and professionals. Three examples of good practice are the Infomore.cz[4] project, the National Health Information Portal[5], and the National Catalogue of Open Data[6], which contains various healthcare datasets. Separately, these modern platforms have no chance of influencing the opinions and attitudes of the general public. However, with systematic and institutionally covered and managed support from governmental organisations, they can form, together with other elements, a solid background for the fight against the infodemic in the Czech Republic. The combination of proven attitudes of professional societies under an umbrella of the Ministry of Health and the Czech Medical Association of

[4] https://www.infomore.cz/

[5] https://www.nzip.cz/

[6] https://data.mzcr.cz/

J. E. Purkyně (associating experts from various medical disciplines) and the correct handling of data, including accurate and non-misleading interpretation, is the key to improve the currently not entirely satisfactory situation in the field of infodemic management, not only in connection with the COVID-19 epidemic.

LESSONS LEARNED

Valid and correctly interpreted data proved to be crucial for decision-making and communication during the COVID-19 epidemic. The information matrix and the described pillars fall under the domain of domain and data understanding, which is one of the main outputs of the Informore.cz project, documents the collection of experience, communication channels and the ability to disseminate information among the general public. Analytical perspectives combining media space and open data on the course of the COVID-19 epidemic have provided unique outputs that highlight the phenomenon of an infodemic.

A systematic approach to combating disinformation requires not only existing and validly processed data but also nationwide education of the general and informed public with the support of interested public authorities and the media space.

ACKNOWLEDGEMENTS

This chapter is based on the results of the Infomore.cz interdisciplinary project, in which Charles University, Masaryk University, and NEWTON Media cooperated. We want to thank our colleagues from the team of project investigators, especially: Alice Němcová Tejkalová, Veronika Macková, Alžběta Krausová, Victoria Nainová, Jiří Jarkovský, Ondřej Májek, Petr Herian, Tomáš Kůst, Alena Zachová, Vít Kadlec, and Věra Čarná.

REFERENCES

[1] World Health Organization. Novel Coronavirus (2019-nCoV) Situation Report – 13, 2020 [Internet]. 2 Feb 2020 [cited 16 Jan 2022]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf.

[2] United Nations Department of Global Communications. UN tackles 'infodemic' of misinformation and cybercrime in COVID-19 crisis [Internet]. 31 March 2020 [cited 15 May 2023]. Available from: https://www.un.org/en/un-coronavirus-communications-team/un-tackling-'infodemic'-misinformation-and-cybercrime-covid-19.

[3] European Centre for Disease Prevention and Control. COVID-19 Coronavirus data – weekly (from 17 December 2020) [Internet]. 2020 [cited 15 May 2023]. Available from: https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data-weekly-from-17-december-2020.

[4] Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med. 2002;113(9):763–5.

[5] Rothkopf D J. When the Buzz Bites Back. The Washington Post; 2023. Available from: https://www.washingtonpost.com/archive/opinions/2003/05/11/when-the-buzz-bites-back/bc8cd84f-cab6-4648-bf58-0277261af6cd/.

[6] Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res. 2020;22(6):e21820.

[7] Komenda M, Bulhart V, Karolyi M, Jarkovský J, Mužík M, Májek O, et al. Complex reporting of coronavirus disease (COVID-19) epidemic in the Czech Republic: use of interactive web-based application in practice. J Med Internet Res. 2020;22(5):e19367.

[8] Czech Television. In coronavirus-related information, people trust experts the most, prime minister the least, Kantar survey shows [Internet].2020 [cited 16 Jun 2023]. Available from: https://ct24.ceskatelevize.cz/domaci/3209500-lide-u-covidu-nejvice-duveruji-expertum-nejmene-premierovi-ukazal-pruzkum-kantaru.

[9] Komenda M, Jarkovský J, Klimeš D, et al. Sharing datasets of the COVID-19 epidemic in the Czech Republic. PLoS One. 2022;17(4):e0267397.

[10] Luengo M, García-Marín D. The performance of truth: politicians, fact--checking journalism, and the struggle to tackle COVID-19 misinformation. Am J Cult Sociol. 2020;8:405–27.

[11] Baccolini V, Rosso A, Di Paolo C, et al. What is the prevalence of low health literacy in European Union member states? A systematic review and meta--analysis. J Gen Intern Med. 2021;36(3):753–61.

[12] The HLS19 Consortium of the WHO Action Network M-POHL. International Report on the Methodology, Results, and Recommendations of the European Health Literacy Population Survey 2019–2021 (HLS19) of M-POHL [Internet]. Vienna: Austrian National Public Health Institute; 2021 [cited 2 Nov 2022]. Available from: https://m-pohl.net/Int_Report_methdology_results_recommendations.

[13] Vodochodský I. Myths connected with COVID-19 [Internet]. 2021 [cited 18 Oct 2022]. Available from: https://www.infomore.cz/res/file/analyzy/20210308-myty-klasicka-media/20210308-myty-klasicka-media.pdf.

[14] Czech Elves. Regular overview of the Czech disinformation scene – February 2022 [Internet]. 2022 [cited 7 Apr 2023]. Available from: https://cesti-elfove.cz/wp-content/uploads/report_unor.pdf.

Data-driven decision-making in medical education and healthcare

Martin Komenda et al.

Designed and typeset by Ing. Martin Janíček, illustrated by Martin Pešta
Proofreading by Lenka Šnajdrová, PhD
Printed by powerprint s.r.o., Brandejsovo nám. 1219/1, 165 00 Prague Suchdol
Published by Masaryk University, Žerotínovo nám. 617/9, 601 77 Brno
and Institute of Health Information and Statistics of the Czech Republic, Palackého náměstí 4,
P.O. BOX 60, 128 01 Prague 2
First edition, 2023

ISBN 978-80-280-0392-0 (Masaryk University)
ISBN 978-80-7472-198-4 (Institute of Health Information and Statistics of the Czech Republic)

MEFANET

MEDICAL
FACULTIES
NETWORK

nzip.cz

https://iba.med.muni.cz/en/data-rulezzz