



# Institute of Biostatistics and analyses: history and activities



Jiří Jarkovský, Adam Svobodník



# Institute of Biostatistics and Analyses

History  
Structure  
Collaboration

# Institute of Biostatistic and Analyses: History

## ☑ Since 2001

- group of scientists interested in analysis of biological and clinical data
- Centre of Biostatistics and analyses

## ☑ 2006

- Institute of Biostatistics and Analyses
- 45 employees and PhD students
- 40 students in pregraduate study of Computational Biology

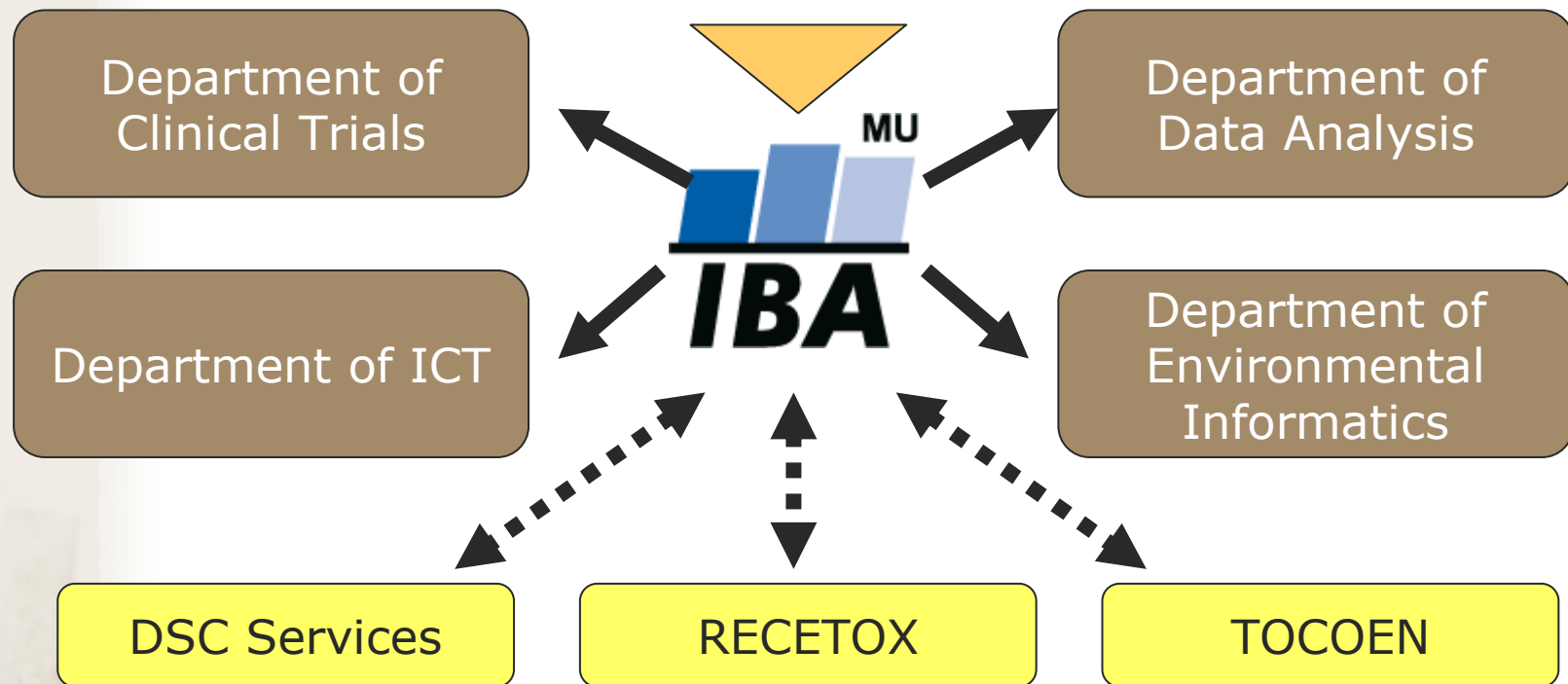


# Institute of Biostatistic and Analyses: Internal Structure

✓ Nowadays IBA has the following structure:



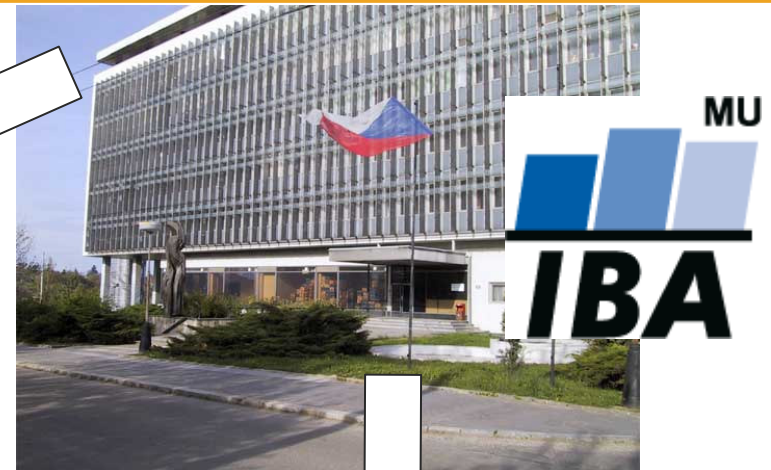
Masaryk university  
• Faculty of Medicine  
• Faculty of Science



# Institute of Biostatistic and Analyses: Collaboration

## Close mutual collaboration

- Faculty of Medicine (MU in Brno)
- Faculty of Informatics (MU in Brno)
- Faculty of Science (MU in Brno)
- Research Centre for Environmental Chemistry and Ecotoxicology (MU in Brno)
- Department of Mathematics (MU in Brno)
  
- Czech Environmental Information Agency
- National and Regional Centre for Persistent Organic Pollutants in Middle and East Europe
- Masaryk Memorial Hospital
- Czech Statistical Office
- Czech National Cancer Registry
- Committee for Breast Cancer Screening (Ministry of Health)
- Czech Oncology Society

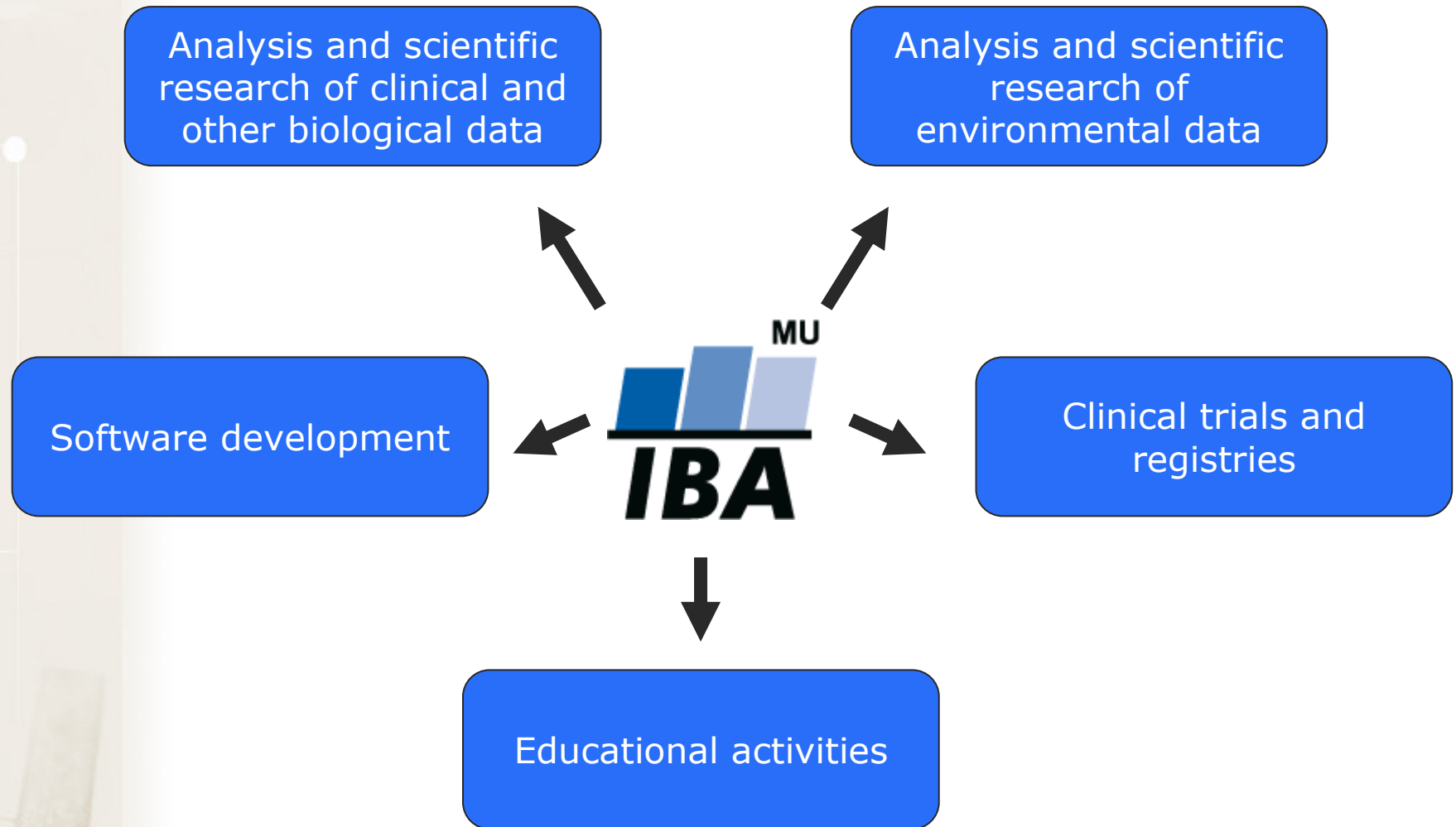


## Standard professional collaboration

More than 100 independent health care providers and hospitals, research institutions, environmental agencies, ministries etc.

# Institute of Biostatistic and Analyses: Areas of interests

## ☑ There are five main areas of interest





# Educational activities

Basic and advanced courses of data analysis  
Computational Biology  
Conferences  
Background of educational and scientific activities

# Basic and advanced courses of data analysis

- ✓ **18 courses for Faculty of Medicine and Faculty of Science**
- ✓ **Both undergraduate and postgraduate courses with 450 students per year**

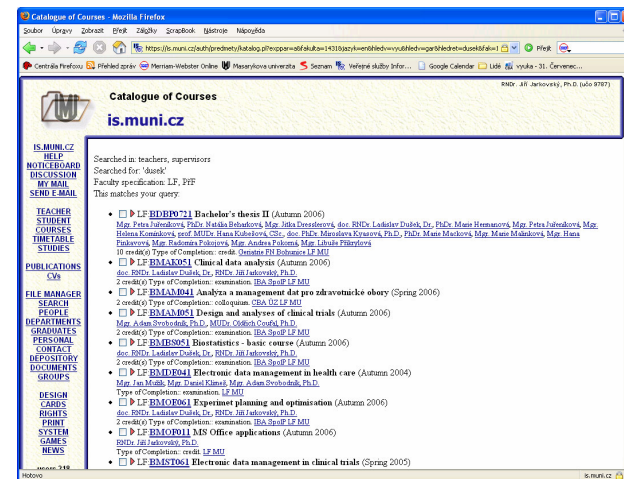
## → Analysis of biological and clinical data

- **Biostatistics**
- **Stochastic modelling**
- **Multivariate data analysis**
- **Experimental design**
- **Data mining**
- **Etc.**

## → Clinical trials

- **Design and analyses of clinical trials**
- **Data management in clinical trials**

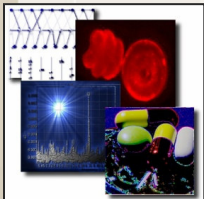
## → Courses of visiting professors: International collaboration in teaching



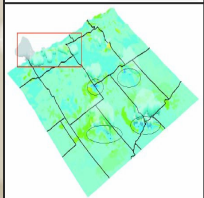


# Computational biology

- ☑ **Computational biology is newly (since 2001) established study programme at Masaryk University with high potential in almost all biological and clinical disciplines**
- ☑ **The concept of the computational biology is:**
  - ➔ **intrinsically inter-disciplinary**
  - ➔ **closely related to standard biological and clinical disciplines and their modern trends (bioinformatics, processing of large scale data from genomics to ecological biomonitoring, risk assessment, predictive modeling...).**
- ☑ **We define computational biology in a broad sense as umbrella branch for all standard biological and clinical disciplines with three main branches:**



- ☐ **Analysis of biological and clinical data**
  - **Applied analysis of data or development of new methods**



- ☐ **Environmental informatics**
  - **Environmental information systems, biomonitoring etc.**



- ☐ **Informatics in medical sciences**
  - **Informatics in clinical trials, information systems for clinical data, clinical registry**

# Conferences

## ☑ IBA organizes several international scientific meeting and conferences

### → 2005

- ☐ Summer School of Environmental Chemistry and Ecotoxicology 2005: tutorial of environmental risk assessment
- ☐ Enviroinfo 2005 – Informatics for environmental protection
- ☐ Summer School on Computational Biology

### → 2006

- ☐ Summer School of Environmental Chemistry and Ecotoxicology 2006: tutorial of environmental risk assessment
- ☐ Summer School on Computational Biology: Predictive modelling and ICT in Environmental Research – September 11-13

### → 2007

- ☐ TIES: Conference of International Environmetrics Society

# Background of educational and scientific activities

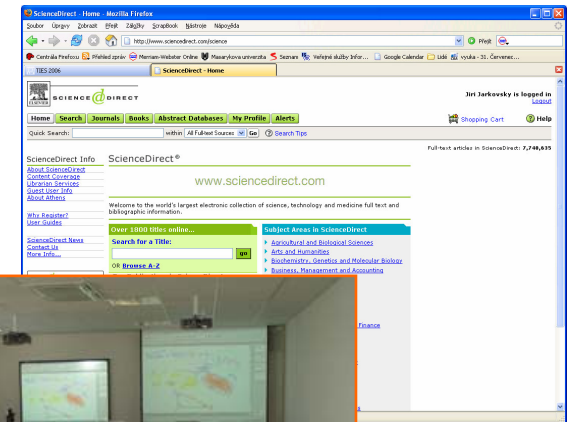
## ✓ Library

- Books and journals covering almost all areas of data analysis in biology and medicine
- Free access to many databases

## ✓ Software equipment

- Statistica
- SPSS + Clementine
- SAS
- R
- Matlab
- Maple
- ArcGIS

## ✓ Modern classroom



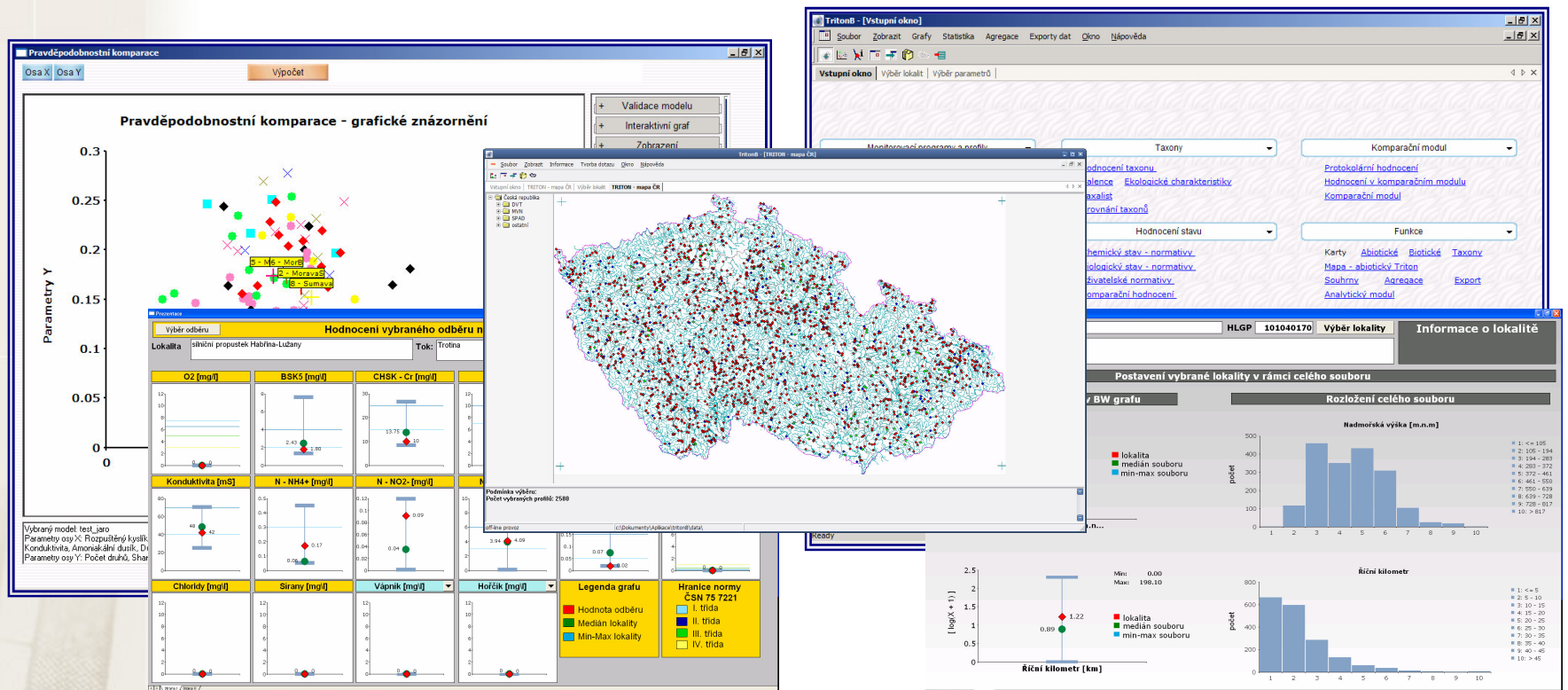


## **Important projects in environmental, biological and epidemiology research**

Environmental research  
Epidemiology  
Various biological research

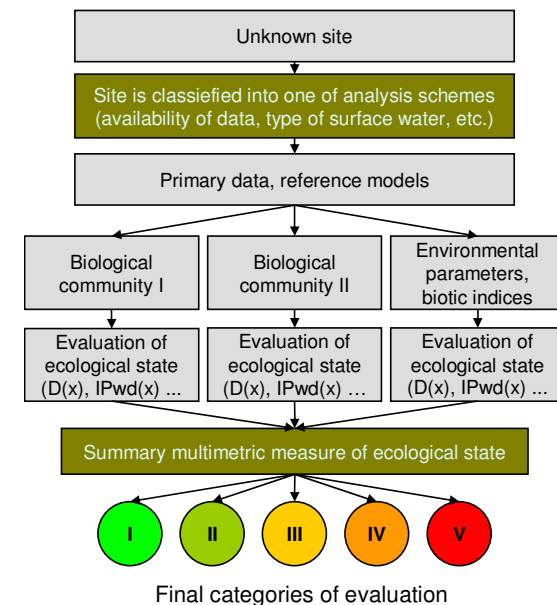
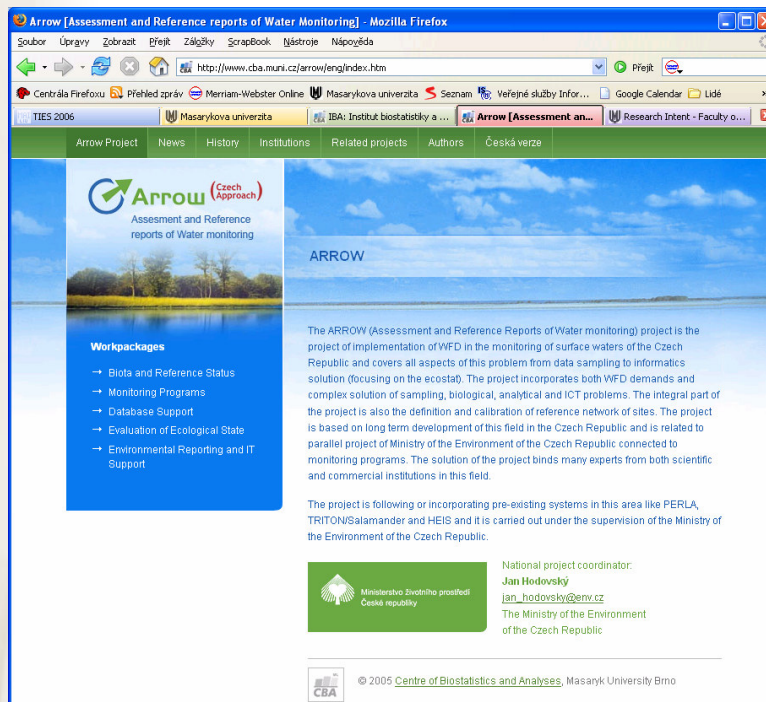
# TRITON project

- ✓ Project with Agricultural Water Management Authority
- ✓ Software and statistical methodology of data analysis in biomonitoring networks
  - ➔ Visualisation of data
  - ➔ Statistical analysis of time trends, biodiversity etc.



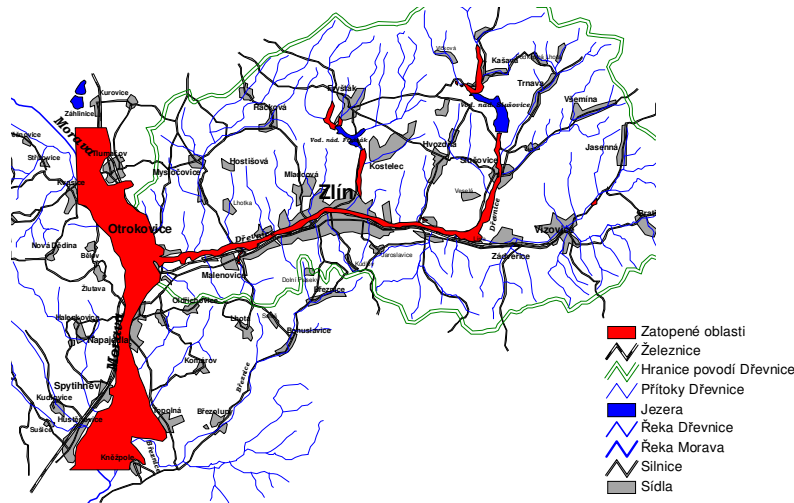
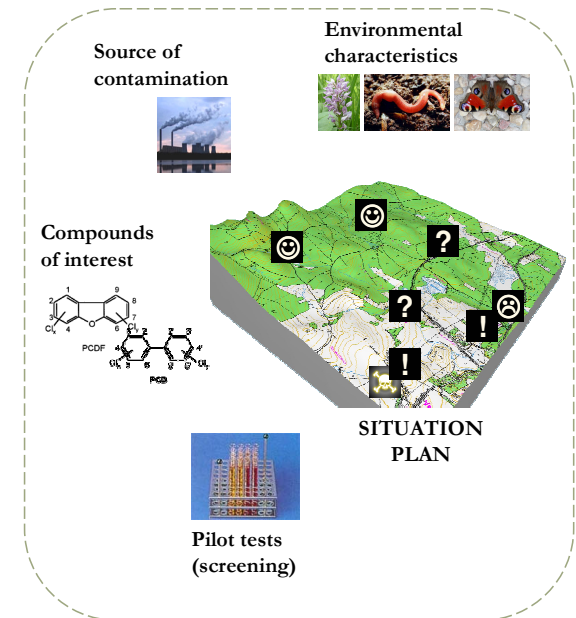
# ARROW project

- ✓ **Part of implementation of Water Framework Directive EU in the Czech Republic**
- ✓ **IBA tasks**
  - ➔ **Development of methodology and scientific research in analysis, typology and prediction of ecological state of localities**
  - ➔ **Software implementation of analytical methodology**



# INCHEMBIOL project

- ☑ **INTERACTIONS AMONG THE CHEMICALS, ENVIRONMENT AND BIOLOGICAL SYSTEMS AND THEIR CONSEQUENCES ON THE GLOBAL, REGIONAL AND LOCAL SCALES**
- ☑ **Long term projects (7 years) including**
  - ➔ **Experimental work**
  - ➔ **Data analysis**
    - ☐ **Modelling of POPs distribution and transport**
    - ☐ **Dose-response analysis**
    - ☐ **Biodiversity analysis**
    - ☐ **Analysis of data of biomonitoring networks**
    - ☐ **Environmental risk assessment**
    - ☐ **GIS**



# INCHEMBIOL: Environmental risk assessment



Uncertainty analysis

RISK CHARACTERIZATION



Benchmarking  
Probability estimation

Expert judgement  
EXPOSURE ASSESSMENT  
Sampling Modelling  
Biota

Bio-tests Bioindicators Biomonitoring

BIOLOGICAL EFFECTS

Data processing Modelling  
Multivariate analyses

CONCEPTUAL MODEL & SCENARIO



Model of area of interest  
Experimental design



Problem definition  
Hazard identification

Retrospective problem

Actual / urgent situation

Prospective evaluation

Data gathering  
Data aggregation  
Information services

Optimization & Processing in information systems & Communication

Autor: Název

© Institute of Biostatistics and analyses





# SVOD project: description

- ✓ **SVOD = EXPERT SYSTEM FOR ANALYSIS AND PRESENTATION OF EPIDEMIOLOGICAL DATA IN ONCOLOGY**
- ✓ **Data of National Cancer Institute: population database of cancer epidemiology since 1970**
- ✓ **Data includes incidence and mortality rates, time profiles, distribution of clinical stages, regional heterogeneity and more...**

http://www.svod.cz - SVOD Analyse - Mozilla Firefox

REGIONAL OVERVIEWS

Select a diagnose ?

II. TUMOURS OF DIGESTIVE ORGANS

C09,C10,C12-C14 - Malignant neoplasm of other sites in the pharynx

C15 - Malignant neoplasm of esophagus

C16 - Malignant neoplasm of stomach

C17 - Malignant neoplasm of small intestine

C18 - Malignant neoplasm of colon

C19 - Malignant neoplasm of rectosigmoid junction

C20 - Malignant neoplasm of rectum

C21 - Malignant neoplasm of anus and anal canal

C18,C21 - Malignant neoplasm of colon and rectum

C22 - Malignant neoplasm of liver and intrahepatic bile ducts

C23 - Malignant neoplasm of gallbladder

C24 - Malignant neoplasm of other and unspecified parts of biliary tract

C23,C24 - Malignant neoplasm of gallbladder and biliary tract

C25 - Malignant neoplasm of pancreas

C26 - Malignant neoplasm of other and ill-defined digestive organs

Display all diagnostic groups

http://www.svod.cz/analyse.php?modul=regionprehled#



# SVOD project: aims

## ✓ Project aims

### → Scientific research of epidemiology data

- ❑ General epidemiological and socio-economical analyses.
- ❑ Load of population by malignant tumours, population risks.
- ❑ Analysis of health care heterogeneity
- ❑ Approximate evaluation of the health care effectiveness according to therapeutic results.

### → Software development

- ❑ Desktop software
- ❑ Web visualization of epidemiology data – public access

<http://www.svod.cz/>

The screenshot shows a web browser window displaying the SVOD project website. The browser is Mozilla Firefox. The address bar shows the URL <http://www.svod.cz/?sec=aktuality&lang=en>. The website content includes a header with the title 'EPIDEMIOLOGY OF MALIGNANT TUMORS IN THE CZECH REPUBLIC' and a navigation menu. A 'NEWS' section is prominent, featuring several articles with dates and titles. The browser's taskbar shows several open windows, including 'SVOD' and 'CENIA'.

# INFORMATION AND ANALYTIC SERVICES OF THE SYSTEM SVOD

SVOD kompletní verze - C64 Zhoubný novotvar ledviny mimo pánevku

Soubor Nastavení Zobrazit Expertní služby Nápověda

SVOD – System for Viewing of Oncological Data

C60 Malignant neoplasm of breast

**Analytic tools – expert services**

SVOD – Main menu

- Diagnose selection
- COBRA - Expert services
- User preferred analyses
- Web portal SVOD
- Discussion club SVOD
- SVOD manual
- SVOD help
- Exit

COBRA – Expert services

Presentations Data browser Expert service

Epidemiology: incidence and mortality Comparative analysis Epidemiology

Age structure of patient population Survival analysis

Clinical stages Demographic data

**Communication and information tools**

**Predefined user controlled analytic tools**

Expert services

Expert services software SVOD

Epidemiology

Comparative analysis

Health care management

Základní epidemiologie

Základní epidemiologie  
Diagnóza C00

Počet případů

1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002

Epopy Tabulka hodnot Zavřít

Interaktivní analýza přežití

Analýza přežití  
Diagnóza C00

Kumulativní pravděpodobnost

Čas[měsíce]

Podíl případů s daným stádiem [%]

Diagnóza C00

St. I. (n=1948)  
St. II. (n=522)  
St. III. (n=208)  
St. IV. (n=91)  
St. neu. (n=1272)

Výběrý region: Jihomoravský kraj  
Srovnávací standard: všechny registry v ČR.  
Referenční hodnoty: minimum, medián, maximum

Věkové specifická incidence

Všechny složité pacienty  
v ČR srovnávací standardem  
Riziková incidence

Věkové skupiny

Výběrý region: Jihomoravský kraj  
Srovnávací standard: všechny registry v ČR.  
Referenční hodnoty: minimum, medián, maximum

**Comprehensive presentations with comments and access to appropriate analytic tools**

Vývoj věkové struktury v čase - všichni pacienti

Počet případů na 100 000 osob ve věku, let.

Věk do 45 let

Počet případů na 100 000 osob ve věku, let.

Věk 45 - 59 let

Počet případů na 100 000 osob ve věku, let.

Věk 60 a více let

Počet případů na 100 000 osob ve věku, let.

Další údaje o věkové struktuře pacientů a demografické údaje a údaje o věkové složce populace můžete získat v podobě grafických a tabulkových výstupů pomocí těchto uživatelských funkcí:

Analýza věkové složky Věkové specifické analýzy v čase Demografické údaje

**Data browser - analysis of individual parameters**

Prohlížeč dat - tvorba výběru

Negativně vyberte analyzovaný parametr ze seznamu. Po výběru analyzovaného parametru můžete v ostatních záložkách blíže určit skupiny pozorů ze kterých chcete data zpracovat.

Analýzovaný parametr	Region	Období	Věková skupina	Pohlaví	Diagnóza	Klinická stádia	Další výběr
1 pohlaví	katégorie	2470					
2 rok narození	datum	2470					
3 věk při stanovení dg.	spojitý	2470					
4 věk při dg. - 5-leté kateg.	katégorie	2470					
5 kraj trvalého bydliště	katégorie	2470					
6 sociální postavení	katégorie	374					
7 hlavní životní zaměstnání	katégorie	374					
8 ZN v rodinné anamnéze	katégorie	374					
9 kouření	katégorie	374					
10 diagnóza onemocnění	katégorie	2470					
11 rok stanovení diagnózy	datum	2470					
12 laterality	katégorie	2470					

Uživatelské parametry diagnostické parametry klinické parametry Umítl

Výsledný graf: katégorie

Analýzovaný parametr: klinické stádium

Stadion II 38,6%  
Stadion III 20,59%  
Stadion I 17,91%  
Stadion neznámé 12,76%  
Stadion IV 9,05%  
Celkem 100,104

Statistické výstupy - frekvenční tabulka

Diagnóza: C00 - Zhoubný novotvar prsu, Vybraný parametr: "klinické stádium".

Hodnota	Počet	%
1 stadium II	38761	38,69
2 stadium III	20626	20,59
3 stadium I	17947	17,91
4 stadium neznámé	12709	12,76
5 stadium IV	9071	9,05
6 celkem	100104	

1 2 3 4 5 6

klínické stádium

# Microarrays and analysis of their data

- ✓ Promising technique of molecular biology, nevertheless with lots of problems on the field of data analysis
- ✓ Analysis of expression and CGH microarrays data
- ✓ Project in cooperation with Masaryk Memorial Hospital
  - ➔ Genomic profiling in prediction of response of radiochemotherapeutic treatment in colorectal carcinoma

