

Centre CERIT-SC

scientific computations,

collaborative research &

support services



Tomáš Rebok

CERIT-SC, Institute of Computer Science MU

MetaCentrum, CESNET z.s.p.o.

(rebok@ics.muni.cz)

Overview

- Centre CERIT-SC – brief introduction
- National Grid Infrastructure (NGI) for research computations
- CERIT-SC & NGI
- Research support by CERIT-SC
- Selected research collaborations
- Additional services available to academic research community

A computing and research centre operating at Masaryk University in Brno, Czech Republic

- long-term history (→ long-term experience in ICT science)
 - CERIT-SC evolved from Supercomputing Center Brno (established in 1994), and
 - participates on the operation of National Grid Infrastructure

Our mission:

<http://www.cerit-sc.cz>

- production services for computational science
 - high-performance computing clusters
 - large data storage, back-ups and data archives
 - web portals & projects' back-office
- an application of top-level ICT in the science
 - own research in e-infrastructures (know-how)
 - novel forms of infrastructure utilization (experimental usage support)
 - research collaborations with other science areas

A long-term experience with:

- operation of large HW/SW & communication infrastructure → High Performance Computing
 - including internal research in e-infrastructures (identity management, security, scheduling algorithms, large data processing – parallel and distributed algorithms, etc.) and computing methods/algorithms
- cooperation in large EU projects and their support
- web portals and projects' back-office
- data back-ups and archiving
- research in collaboration with partners of different science-fields
- additional services for researchers

National Grid Infrastructure (NGI) for research computations

National Grid Infrastructure (NGI)

CERIT-SC resources integrated into the NGI

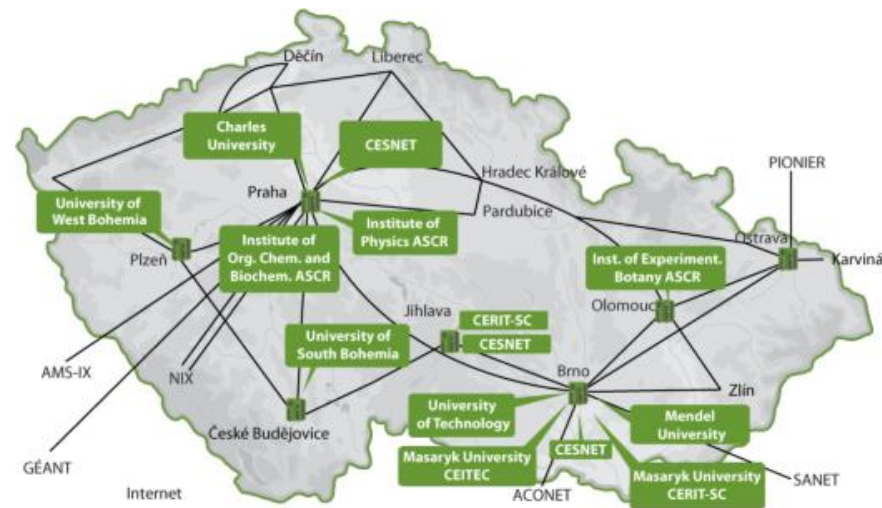
- operated by MetaCentrum NGI (CESNET) since 1996
- MetaCentrum was established by CERIT-SC (previously called SCB)

<http://www.metacentrum.cz>

National Grid Infrastructure

Integrates medium/large HW centers (clusters, powerful servers, storages) of several universities/institutions

- → environment for work/collaboration in the area of research computations and data handling
- NGI further integrated into the European Grid Infrastructure (EGI.eu)



Computing clusters

a group of “common” interconnected computers



(previously)

Computing clusters

a group of “common” interconnected computers



(now)

MetaCentrum Virtual Organization (Meta VO)

Available to all academic users from Czech universities, Academy of Science, research institutes, etc.

- commercial bodies just for public research

Offers:

<http://metavo.metacentrum.cz>

- **computing resources**
- **storage resources**
- **application programs**

After registration, all the resources/services are available free of charge

- users “pay” via publications with acknowledgements
→ results in user priorities in cases of high load



MetaVO – basic properties

After registration, **the resources are available without any administrative burden**

- → ~ immediately (based on the actual load)
- **no resource applications** have to be provided

User accounts periodically extended every year

- a proof of continuing user's academic affiliation
- **publications with acknowledgements simultaneously reported**
 - could help us when asking for funds from public authorities

Best-effort service

Meta VO – computing resources available

Computing resources: ca 10000 cores (x86_64)

- nodes with lower number of computing cores: 2x4-8 jader
- nodes with medium number of comp. cores (SMP nodes): 32-80 cores
- memory (RAM) up to 1 TB per node
- a node with high number of computing cores: 288 cores, 6 TB of RAM
- other „exotic“ hardware:
 - nodes with GPU cards, etc.

CERIT-SC: important resource provider (4512 cores)

<http://metavo.metacentrum.cz/cs/state/hardware.html>

Meta VO – storage resources available

ca 1 PB (1063 TB) for operational data

- centralized storage arrays distributed through various cities in the CR
- user quota 1-3 TB on each storage array

ca 19 PB (19000 TB) for archival data

- “unlimited” user quota

CERIT-SC: important resource provider (5 PB)

<http://metavo.metacentrum.cz/cs/state/nodes>

Meta VO – software available

~ 250 different applications (commercial & free/open s.)

– see <http://meta.cesnet.cz/wiki/Kategorie:Aplikace>

- **development tools**

– GNU, Intel, and PGI compilers, profiling and debugging tools (TotalView, Allinea), ...

- **mathematical software**

– Matlab, Maple, Mathematica, gridMathematica, ...

- **application chemistry**

– Gaussian 09, Gaussian-Linda, Gamess, Gromacs, ...

- **material simulations**

– Wien2k, ANSYS Fluent CFD, Ansys Mechanical, Ansys HPC...

- **structural biology, bioinformatics**

– CLC Genomics Workbench, Geneious, Turbomole, Molpro, ...

CERIT-SC: important commercial SW provider

Meta VO – grid environment

- *batch jobs*

- the computations described by script files

- *interactive jobs*

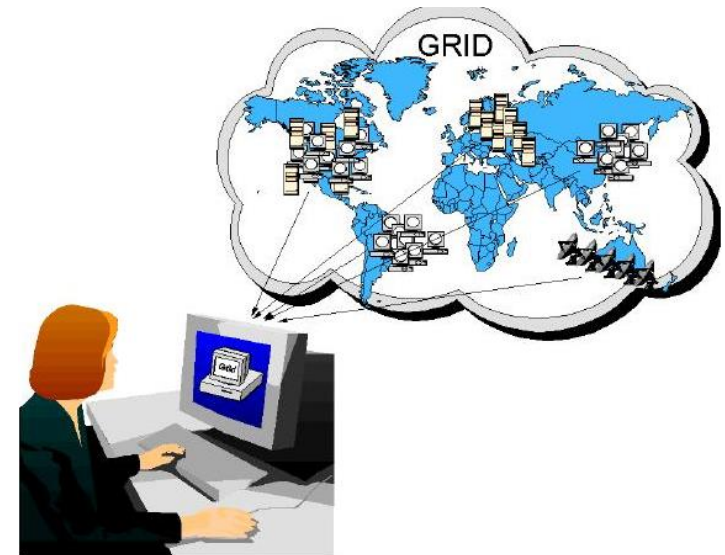
- text & graphical environment

- *cloud computing*

- instead of running jobs with computations, users run the whole virtual machines (the whole OS becomes under their control)

- focused on research computations again (not for webhosting)

- Windows & Linux images provided, user-uploaded images also supported



CERIT-SC & NGI

Centre CERIT-SC & NGI

CERIT-SC is an important NGI partner

- **HW & SW resources provider**

 - SMP nodes (1600 cores)

 - HD nodes (2624 cores)

 - SGI UV node (288 cores, 6 TB RAM)**

 - storage capacity (~ 5 PB)

- **significant personal overlaps with NGI exist**

remember, **CERIT-SC (SCB) established MetaCentrum NGI**

- → **much research/work is performed in collaboration**



CERIT-SC & NGI – production services

High-performance computing

- parallel/distributed computations

Data back-ups and archiving

- multiple storage systems in geographically distant locations
- advanced hierarchical storage systems

Web portals & projects' back-office

- *for general public & dissemination*
web pages, RSS feeds, blogs, social media, ...
- *for projects' internal needs*
data & document servers, request tracking, messaging, meeting planners, collaborative environments, ...

Authentication and Authorization Infrastructure, Identity Management, Data Security, ...

■ CERIT-SC & NGI – participation in large EU projects

Building European grid research infrastructure:

DataGrid, EGEE, EGEE II, EGEE III, EGI DS, EGI InSPIRE, EMI, EUAsiaGrid, CHAIN, CHAIN-REDS, Thalamos, ...

Basic research in grid infrastructures:

GridLab, CoreGrid, Moonshot, ...

Other projects' support:

ELIXIR (European life-science infrastructure for biological information)

BBMRI (Biobanking and Biomolecular Resources Research Infrastructure)

ELI (Extreme Light Infrastructure)

Pierre Auger Observatory

Thalassemia

...

■ CERIT-SC & NGI – services for selected projects being supported I.

EGI.eu (European Grid Infrastructure):

- **web pages:** <http://www.egi.eu/>
- **authentication & authorization infrastructure:** <http://www.egi.eu/sso/>
- **blogs:** <http://www.egi.eu/blog/>
- **event webs:** <http://tf2012.egi.eu> <http://tf2011.egi.eu> ...
- **wiki pages:** <http://wiki.egi.eu/>
- **mailinglists:** <http://mailman.egi.eu/>
- **document server:** <http://documents.egi.eu/>
- **request tracking:** <http://rt.egi.eu/>
- **discussion forum:** <http://forum.egi.eu/>
- **Indico (meeting planner):** <http://indico.egi.eu/>
- **Jabber (no web):** <jabber.egi.eu>

EGI DS:

- **web pages:** <http://web.eu-egi.eu/>

■ CERIT-SC & NGI – services for selected projects being supported II.

MetaCentrum NGI + VO:

- **web pages:** <http://www.metacentrum.cz> , <http://metavo.metacentrum.cz/>
- **authentication & authorization infrastr.:** <http://perun.metacentrum.cz/>
- **mailinglists:** <https://www.metacentrum.cz/mailman/admin/>

MediGrid:

- **web pages:** <http://www.medigrid.cz/cs/>
- **application for searching drug interactions:** <http://www.medigrid.cz/interakce/>

Pathological atlases:

- **web pages, data storage & archive:** <http://atlases.muni.cz/>

EEF - European E-infrastructure Forum

- **web pages:** <http://www.einfrastructure-forum.eu/>

Research support by CERIT-SC

■ Research support by CERIT-SC

Fact I. Common HW centers provide just a “dumb” power without any support how to effectively use it

Fact II. Common HW centers do not participate on the users’ research aiming to help them with ICT problems

CERIT-SC collaborates with its users:

- to help them effectively use the provided resources
- to help them to cope with their ICT research problems focusing on an application of top-level ICT in the science

What's the idea?

We focus on intelligent & novel usage forms of the provided infrastructure

- the provided HW/SW resources serve just as a tool for research and development
 - highly-flexible infrastructure (convenient to experiments)
- in comparison with NGI resources, the production computations are at the second-level of interest
- the centre aims to be equipped with cutting-edge technologies in order to allow top-level research (both internal & collaborative)
- real research collaboration with our partners
 - the collaborations generate new questions/problems for IT
 - the collaborations generate novel opportunities for the science (we DON'T want to be a common service organization)

■ How do we fulfill the idea?

How are the research collaborations performed?

- the work is carried via a diploma/doctoral thesis of a FI MU student
- the CERIT-SC staff supervises/consults the student and regularly meets with the research partners
- the partners provide the expert knowledge from the particular area
- in an ideal case, once the thesis become defended, the collaboration continues via an externally funded project

Strong ICT expert knowledge available:

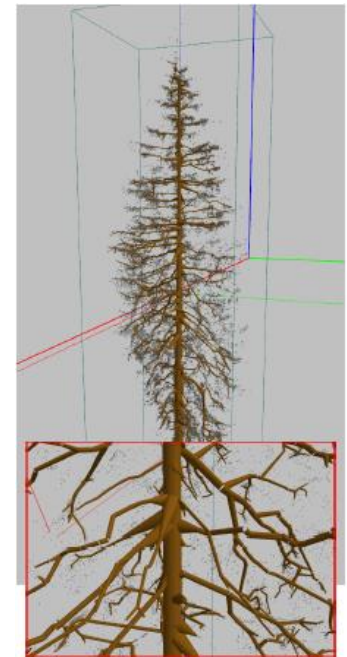
- long-term collaboration with Faculty of Informatics MU
- long-term collaboration with CESNET
- consultations with experts in the particular areas

Selected research collaborations

Selected (ongoing) collaborations I.

3D tree reconstructions from terrestrial LiDAR scans

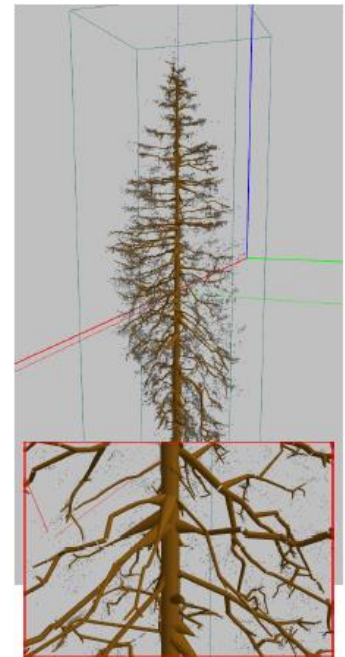
- partner: Global Change Research Centre - Academy of Sciences of the Czech Republic (*CzechGlobe*)
- **the goal: to propose an algorithm able to perform fully-automated reconstruction of tree skeletons (main focus on Norway spruce trees)**
 - from a 3D point cloud
 - scanned by a LiDAR scanner
 - the points provide information about XYZ coordinates + reflection intensity
 - *the expected output: 3D tree skeleton*
- **the main issue:** overlaps (→ gaps in the input data)



Selected (ongoing) collaborations I.

3D tree reconstructions from terrestrial LiDAR scans – cont'd

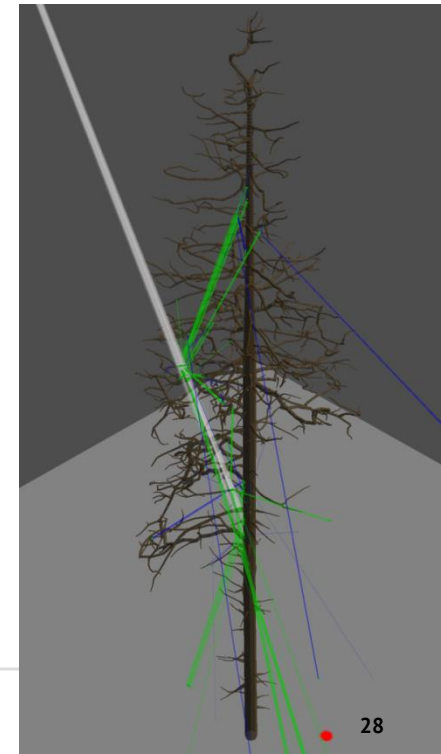
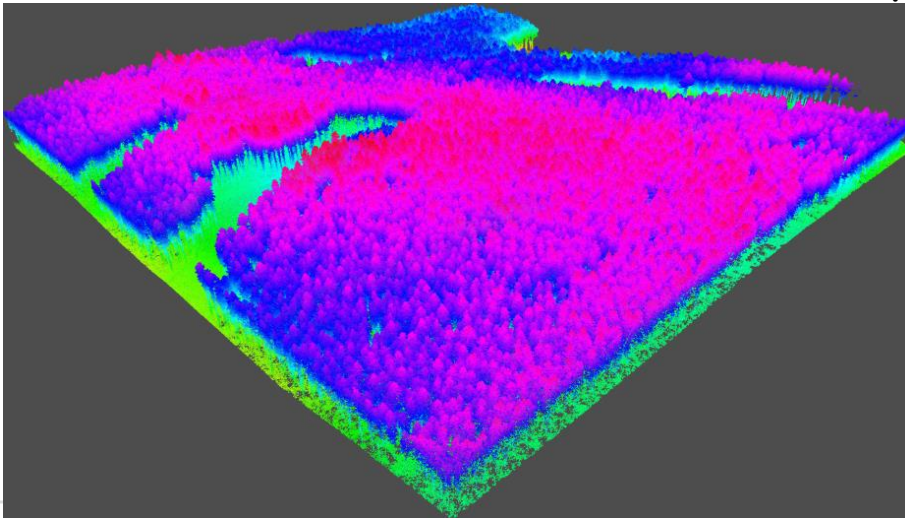
- the diploma thesis proposed a novel innovative approach to the reconstructions of 3D tree models
- the reconstructed models used in subsequent research
 - **determining a statistical information** about the amount of wood biomass and about basic tree structure
 - **parametric supplementation of green biomass** (young branches+ needles) – a part of the PhD work
 - **importing the 3D models into tools performing various analysis** (e.g., DART radiative transfer model)



Selected (ongoing) collaborations II.

3D reconstruction of tree forests from full-wave LiDAR scans

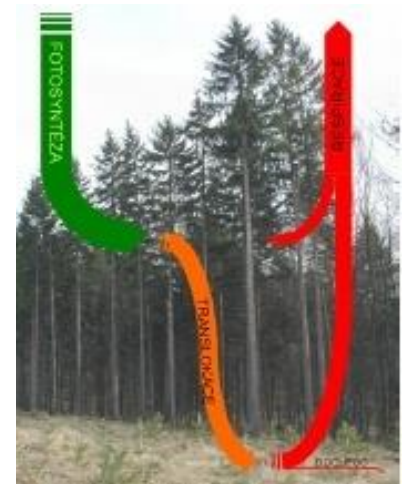
- subsequent PhD thesis, a preparation of joint project
- **the goal: an accurate 3D reconstruction of tree forests scanned by aerial full-waveform LiDAR scans**
 - possibly supplemented by hyperspectral or thermal scans, in-situ measurements, ...



Selected (ongoing) collaborations III.

An application of neural networks for filling in the gaps in eddy-covariance measurements

- partner: Global Change Research Centre - Academy of Sciences of the Czech Republic (*CzechGlobe*)
- **the goal: to propose a novel fully-automated method for gap-filling of eddy-covariance data**
 - based on historical measurements and self-learning
 - *accompanying characteristics* – temperature, pressure, humidity,...
- **main issues:**
 - historical data have to be taken into account
 - the forest evolves (grows)



Selected (ongoing) collaborations IV.

Identification of areas affected by geometric distortions in aerial landscape scans

- partner: Global Change Research Centre - Academy of Sciences of the Czech Republic (*CzechGlobe*)
- **the goal: to propose a novel, fully-automated method for an identification of regions within the scans, where the airplane suddenly deviated**
 - and thus introduce distortions in the scanned data
 - → *image processing*
 - current approaches are suitable for determining distortions in the scans of regular objects (like buildings in the city scans) rather than their determination in the diverse vegetable
- **main issue: diverse tree structure**

Selected (ongoing) collaborations V.

De-novo sequencing *Trifolium pratense*

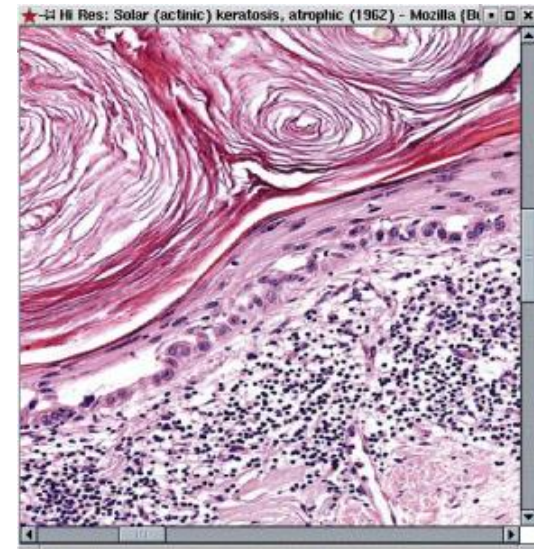
- partner: *Institute of Experimental Biology SCI MU*
- the goal: evaluation and optimization of available tools for DNA reads corrections and assembly
 - *Trifolium pratense* analysis results in large computations
 - ~ 500 GB of memory
 - computations take weeks/months
- main issue: computation demands



Selected (ongoing) collaborations VI.

Virtual microscope, pathologic atlases

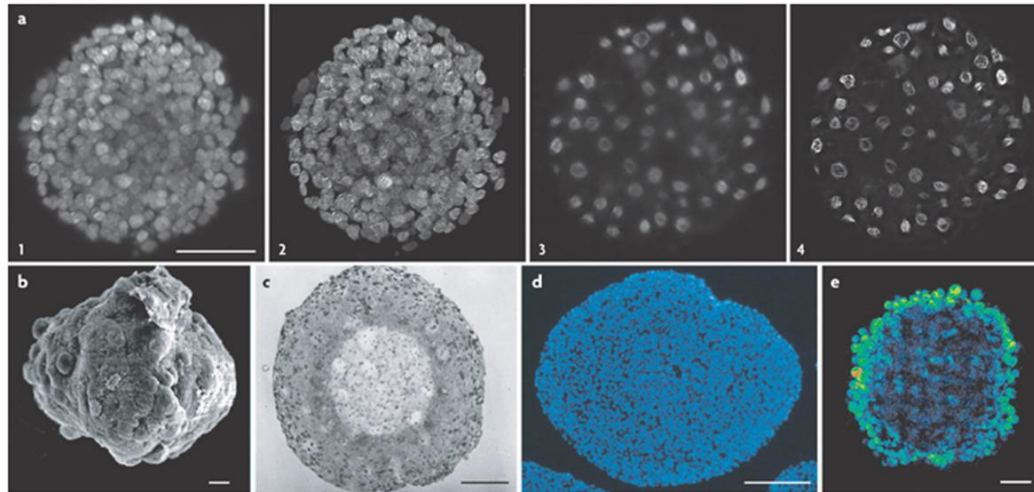
- partner: *Faculty of Medicine MU*
- **the goal: an implementation of virtual microscope for dermatology atlas (web application)**
 - shows the tissue scans
 - resolution up to 170000x140000 pixels
 - composed from tiles (up to 30000 of tiles)
 - allows to „focus“ like real microscope
- **main issues:**
 - optimization of scans processing (GPU)
- **the result is available at <http://atlases.muni.cz>**



Selected (ongoing) collaborations VII.

Segmentation of live cell cultures in microscope images

- partner: *University of South Bohemia*
- **the goal: to determine interesting/important objects in the images of live cell cultures, filtering the noise out of attention**
 - implemented in C and CUDA
 - achieved acceleration: 10x – 1000x



Selected (ongoing) collaborations VIII.

An algorithm for determination of problematic closures in a road network

- partner: *Transport Research Centre, Olomouc*
- **the goal: to find a robust algorithm able to identify all the road network break-ups and evaluate their impacts**
- **main issue: computation demands**
 - the brute-force algorithms fail because of large state space
 - 2 algorithms proposed able to cope with multiple road closures

Selected (ongoing) collaborations IX.

- **Biobanking research infrastructure (BBMRI_CZ)**
 - *partner: Masaryk Memorial Cancer Institute, Recamo*
- **Propagation models of epilepsy and other processes in the brain**
 - *partner: MED MU, ÚPT AV, CEITEC*
- **Photometric archive of astronomical images**
- **Extraction of photometric data on the objects of astronomical images**
 - *2x partner: partner: Institute of theoretical physics and astrophysics SCI MU*
- **Bioinformatic analysis of data from the mass spectrometer**
 - *partner: Institute of experimental biology SCI MU*
- **Synchronizing timestamps in aerial landscape scans**
 - *partner: CzechGlobe*
- **Optimization of Ansys computation for flow determination around a large two-shaft gas turbine**
 - *partner: SVS FEM*
- **3.5 Million smartmeters in the cloud**
 - *partner: CEZ group, MycroftMind*
- ...

**Additional services available
to academic research
community**

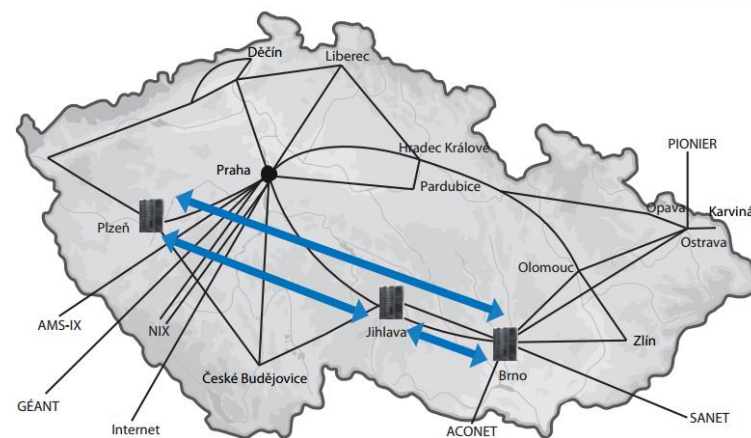
Storage and archival services

The need to archive long-term scientific data increases

- e.g., archival of data used in experiments
in order to allow further usage or results revision

Centralized storage infrastructure:

- 3 hierarchical storage systems available
located in Pilsen, Jihlava (CERIT-SC) and Brno
the total capacity available: ca 19 PB
- suitable for backups, archival, and
data sharing
- additional services:
FileSender
OwnCloud



Remote collaboration support

Support for interactive collaborative work in real-time

- **videoconferences**
HD videoconferencing support via H.323 HW/SW equipment
- **webconferences**
SD videoconferencing support via Adobe Connect (Adobe Flash)
- **special transmissions**
HD, UHD, 2K, 4K, 8K with compressed/uncompressed video transmission (UltraGrid tool)
- **IP telephony**

Support for offline content access

- **streaming**
- **video archive**



Security services

Security incidents handling

- detailed monitoring of possible security incidents
- the users/administrators are informed about security incidents, and
- helped to resolve the incident
- additional services:
seminars, workshops, etc.

Security teams CSIRT-MU and CESNET-CERTS

- several successes:
e.g., Chuck Norris botnet discovery



Federated identity management

Czech academic identity federation eduID.cz

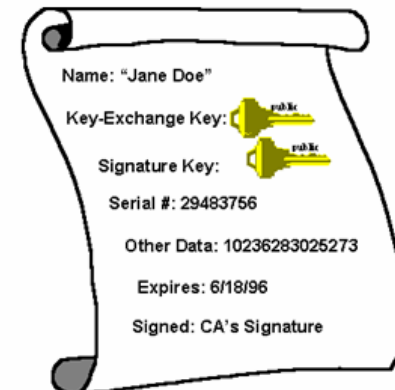
- provides means for inter-organizational identity management and access control to network services, while respecting the privacy of the users
- users may access multiple applications using just a single password
- service provider administrators do not have to preserve user's credentials and implement authentication
- user authentication is always performed at the home organization, user credentials are not revealed to the service providers



PKI – users and servers certificates

CESNET CA certification authority

- provides the users with TERENA (Trans-European Research and Education Networking Association) certificates
 - usable for electronic signatures as well as for encryption
- CESNET CA services:
 - issues personal certificates
 - issues certificates for servers and services
 - certificates registration offices
 - certificates certification offices





Mobility and roaming support

Eduroam.cz

- idea to enable transparent usage of (especially wireless) networks of partner (Czech as well as abroad) institutions



<http://www.eduroam.cz>

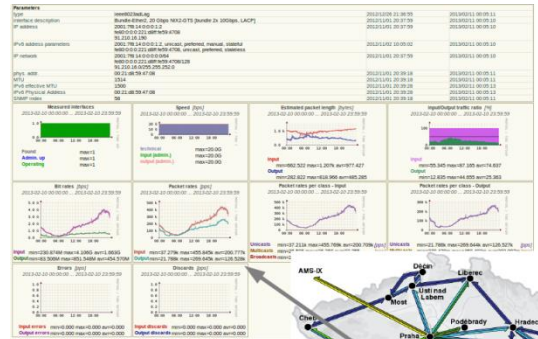
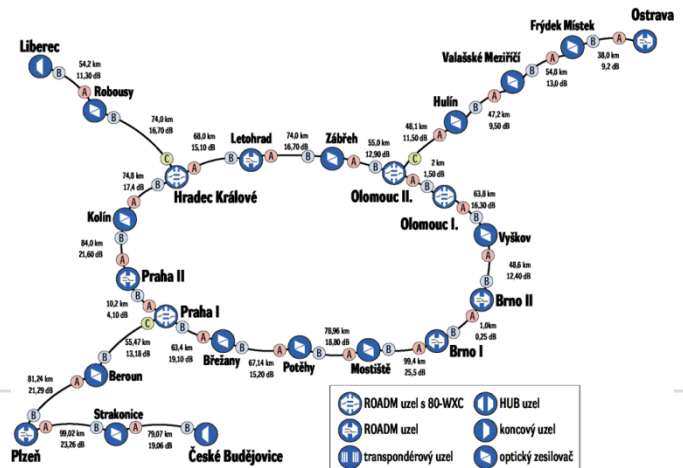
Communication infrastructure and its monitoring

The basis of all the services: high-speed computer network

- 100 Gbps, called CESNET2
- interconnected with pan-european network GÉANT

and its monitoring

- detailed network monitoring (quality issues as well as individual nodes behaviour) available
- automatic detection of various events, anomalies, etc.



Conclusions

Conclusions I.

There're three computing e-infrastructures being established in the Czech Republic

IT4Innovations (VŠB-Technical University of Ostrava)

- **currently ca 3300 cores (around 30000 cores planned)**
- **intended for large production academic/commercial computations** (more resources available thanks to integration into PRACE) **on more or less homogeneous infrastructure**
 - formal applications (research project proposals) required
 - financial participation required (highly welcomed)

National Grid Infrastructure + CERIT-SC

- **currently ca 10000 cores, available for public research only**
- **free of charge, heterogeneous resources (exotic HW available)**
- **intended for common small-to-medium scientific computations or IT4I projects preparation**

Conclusions II.

CERIT-SC aims to provide additional services beyond the scope of common HW centers

an environment for collaborative research

- not only HW/SW provider, but
- → a real collaboration of IT experts and users

we focus on novel and beneficial approaches to e-infrastructure usage

- **big focus on internal research in e-infrastructure services**

we collaborate with several EU projects, including the ESFRI ones

- participation in the preparation of EU H2020 projects

however, we're also interested in collaboration with smaller groups/individuals

- currently, the interest exceeds our (personal) capacities (we have to choose among the collaboration proposals)

Conclusions III.

CERIT-SC didn't grow on a green meadow ...

... and doesn't operate on an isolated island

- long-term history & experience (SCB established in 1994)
- strong interconnection with European infrastructures
 - 10 Gbps connection to NREN academic network (core 100 Gbps)
 - NREN directly connected to European 10 Gbps GÉANT network

Centre location in Brno, CZ is highly beneficial:

- Brno city provides a strong academic & IT background
 - 5 universities (→ intellectual background, sustainability)
- many worldwide IT companies reside in Brno:
 - we cooperate with Red Hat, IBM, Microsoft, NetSuite, ...
 - further companies in Brno: Honeywell, AVG, Avast, Solarwinds, GoodData, 2K, ...
- *“Brno ~ Mekka IT in the CR”*

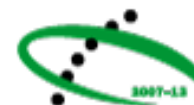


Thank You!



EUROPEAN UNION

EUROPEAN REGIONAL
DEVELOPMENT FUND
INVESTING IN YOUR FUTURE



OP Research and
Development for Innovation

The CERIT Scientific Cloud project (reg. no. CZ.1.05/3.2.00/08.0144) is supported by the *Operational Program Research and Development for Innovations*, priority axis 3, subarea 2.3 *Information Infrastructure for Research and Development*.

<http://metavo.metacentrum.cz>

<http://www.cerit-sc.cz>