



e-Infrastructure in the Czech Republic

Luděk Matyska

GCCP2011, October 24, 2011



EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI



OP Výzkum a vývoj
pro inovace

e-Infrastructure explicitly defined in the *Roadmap of the large infrastructures for research, development and innovation in the Czech Republic*

- ▶ Considered equally important as other large infrastructures for R&D
- ▶ Understood as a combination of vertical and horizontal responsibilities
 - ▶ However, no specific action defined (e.g., projects in other disciplines may have their own IT budget)
- ▶ Combination of resource provisioning, coordination, and own research
 - ▶ Directly or indirectly attached to universities
- ▶ All major e-Infrastructure aspects covered
- ▶ Total funding above 160 MEuro for next 5 years
 - ▶ Additional minor funding under discussion

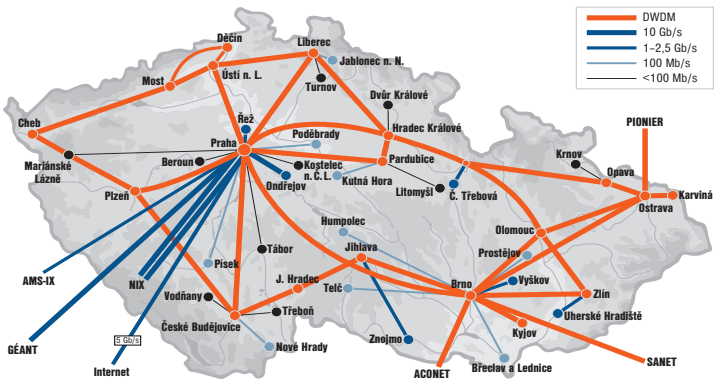
- ▶ **CESNET**
 - ▶ Originally Czech NREN
 - ▶ HQ in Prague, distributed in all regions
 - ▶ Responsible for communication infrastructure
 - ▶ Coordinating role as the Czech NGI in Grids
 - ▶ Contributing to the data infrastructure
- ▶ **IT4Innovations**
 - ▶ Located in Ostrava (Technical University)
 - ▶ Supercomputing centre
- ▶ **CERIT-SC**
 - ▶ Located in Brno (Masaryk University)
 - ▶ Grid, cloud and data centre

- ▶ Internet in the Czech Republic since 1992
 - ▶ CESNET as a department of the Czech Technical University
 - ▶ Transformed into legal body in 1996
 - ▶ Owned by public universities and Academy of Sciences
- ▶ Supercomputing Centre Brno at MU since 1994
 - ▶ Originally HPC interest
 - ▶ Distributed computing infrastructure (MetaCentrum) initiated in 1996 (founded MetaCentrum)
- ▶ DCI became part of CESNET in 1998
 - ▶ Joint activity with SCB MU
- ▶ International collaboration
 - ▶ Since TEN-34 till GEANT3
 - ▶ From DataGrid to EGI InSPIRE and EMI projects
 - ▶ Coordinator in EGI_DS
 - ▶ Many auxiliary projects

- ▶ CESNET became “Large infrastructure for R&D”
 - ▶ Direct governmental subsidy since March
 - ▶ Project *Extension of the national information infrastructure for R&D in regions* (eIGeR, European Structural Funds, 3rd axis, started May 1st)
- ▶ SCB MU transformed into CERIT-SC
 - ▶ Project *Cerit Scientific Cloud* (CERIT-SC, European Structural Funds, 3rd axis, started May 1st)
- ▶ Centre of Excellence IT4Innovations
 - ▶ New activity and project (European Structural Funds, 1st axis, started July 1st)

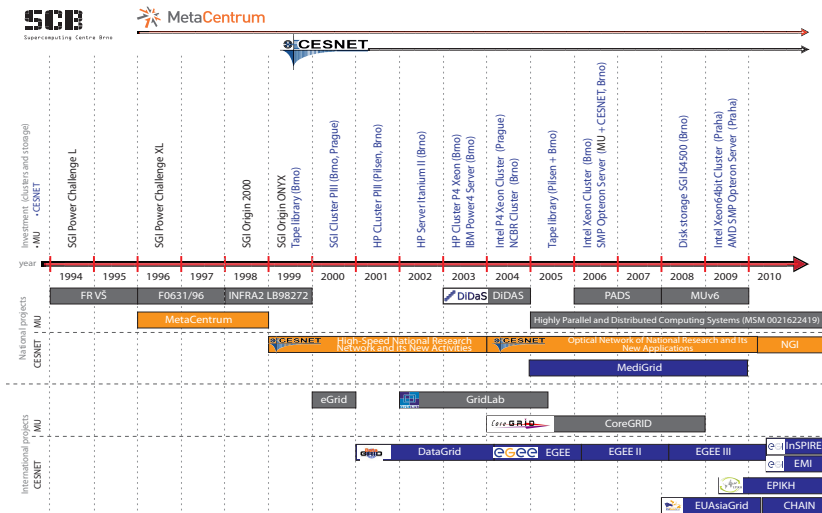
- ▶ High speed communication network
 - ▶ Multi 10 Gbps now, upgrade to 40–100 Gbps in near future
 - ▶ Connecting all major cities in the Czech Republic
 - ▶ All public and state universities, all institutes of Academy of Science
 - ▶ Many private universities, industrial research facilities, faculty and other hospitals, libraries, secondary schools, . . .
- ▶ National Grid coordinator—MetaCentrum
 - ▶ Serves as Czech NGI
 - ▶ Basic resource provisioning
- ▶ Independent data management (new activity)
 - ▶ Three multi-PB installations
 - ▶ Core of the national distributed data infrastructure
- ▶ Close collaboration with other e-Infrastructure components

National Network Infrastructure



- ▶ Provide reasonable computing and data resources
- ▶ Driver for new generation of e-Infrastructure related R&D
- ▶ Novel (including disruptive) use of e-Infrastructure
- ▶ Cooperative R&D with e-Infrastructure users
- ▶ Controlled playground for R&D on the boundary between Informatics and other scientific areas
 - ▶ With computing and storage needs that are not satisfied with standard approaches
- ▶ Collaborate with scientific communities
 - ▶ Not just users, *partners*
- ▶ Adapt and *evolve* the e-Infrastructure
 - ▶ To suit actual and foreseen needs of scientific communities
- ▶ Personal involvement in EGI, EMI, and other EU projects

- ▶ Supercomputing centre
 - ▶ First resources 2012
 - ▶ Full size in 2014, targeting 50th to 100th position in TOP500
- ▶ Centre of excellence
 - ▶ IT4People
 - ▶ Multimedia and risk management
 - ▶ SC4Industry
 - ▶ Numerical models, parallel computing
 - ▶ Theory4IT
 - ▶ Soft computing, knowledge management, security
- ▶ Represents Czech Republic in PRACE



- ▶ Close collaboration of CESNET and CERIT-SC
 - ▶ CESNET: Broad regional coverage, neutral body
 - ▶ CERIT-SC: Access to students, PhD school, closer interaction with partners
- ▶ Conservative versus innovative
 - ▶ Grid infrastructure serves as stable computing and data processing environment
 - ▶ CERIT-SC resources will serve primary for research and development of new methods and protocols
 - ▶ Close partnership with other scientific communities
 - ▶ e-Infrastructure related R&D results will be transferred through NGI/MetaCentrum to other resource centers

- ▶ Has been part of Grid related activities (MetaCentrum)
 - ▶ This is still part of CERIT-SC vision
- ▶ CESNET added a construction of an independent data facility
 - ▶ To serve also non-computing oriented requirements
- ▶ Data facilities built also within other projects
 - ▶ IT4Innovations and CERIT-SC will have several PB each
 - ▶ Large national projects like CEITEC will also have their own data depots
- ▶ Distributed Data Infrastructure
 - ▶ Again CESNET & CERIT-SC close collaboration

- ▶ IT4Innovations
 - ▶ “Standard” supercomputing process
 - ▶ Easy access to small capacity
 - ▶ Aiming for fast turnaround
- ▶ Network and Grids
 - ▶ Primary best effort
 - ▶ Dynamic priority assessment and assignment
 - ▶ Joint projects, collaboration, . . .
 - ▶ But also past results with acknowledgment
 - ▶ Support scientific excellence without bureaucracy
- ▶ Data infrastructure
 - ▶ Under discussion
 - ▶ Aiming for the open best effort access
 - ▶ But SLD/SLA for data preservation

- ▶ Coordinated by CESNET
- ▶ Five major resource providers:
 - ▶ Masaryk University (SCB, now CERIT-SC), Brno
 - ▶ Charles University, Prague
 - ▶ West Bohemia University, Pilsen
 - ▶ Institute of Physics, Prague
 - ▶ CESNET, Prague (but resources also elsewhere)
- ▶ Many smaller resource centres
- ▶ Power
 - ▶ Computing: > 5,000 cores
 - ▶ Data: around 1 PB

- ▶ Virtualized physical layer
 - ▶ SMP and GPGPU equipped machines still experimental
 - ▶ Sits under the “standard” grid middleware
 - ▶ Magrathea/Torque used for the scheduling
- ▶ Majority of resources (thanks to HEP) on EGI infrastructure
- ▶ However, national grid infrastructure simpler
 - ▶ Torque (previously PBSPro) as the central scheduler
 - ▶ Storage through combination of AFS (metadata, software distributions) and NFS (locally v3, globally v4)
 - ▶ Currently moving into a distributed peer to peer scheduler infrastructure
 - ▶ More flexibility
 - ▶ No dependency on central services
- ▶ Virtualization opens space for experimental use of the Grid resources

- ▶ CESNET
 - ▶ Stable infrastructure, serving all scientific communities
 - ▶ Joint projects with selected users
- ▶ CERIT-SC
 - ▶ Experimental/Discovery use
 - ▶ Development and optimization of algorithms, methods and their scalability, ...
 - ▶ Unstable/development OSES, specialized environments, map-reduce, ...
 - ▶ Intensive collaboration leading to joint publications
- ▶ Jointly: Keep balance among conflicting requirements
 - ▶ Helps users to make proper estimates of their real needs
 - ▶ Buy and use own resources
 - ▶ Move long-term planned computations elsewhere (e.g., IT4I)
 - ▶ No bureaucratic procedures, user prioritization based on recent results

e-Infrastructure usable (shared) by various *scientific* communities

- ▶ Computer networks are closest to this ideal
 - ▶ But moving 100TB is still fastest (and cheapest) by physical transporting of disks
- ▶ Batch processing on grids suitable only for *some* users
 - ▶ Interactive access more convenient
 - ▶ Limited support for parallel jobs
- ▶ Data consolidation, access, sharing, ... still rather primitive
 - ▶ Commonalities hard to find
- ▶ Collaborative infrastructure and tools
- ▶ Security
 - ▶ Orthogonal to other activities
 - ▶ From authentication and authorization (AA) to secured data and computation environments
 - ▶ Mutual trust between users and resource providers

- ▶ Managing Virtual LANs at the network backbone
 - ▶ Part of L2 setup
 - ▶ Private IP networks running at the national network infrastructure
- ▶ Coordinated with virtual cluster setup
 - ▶ Part of the job scheduling
 - ▶ Internal IP addresses, not exposed outside
 - ▶ Running virtual clusters without any internal authentication
 - ▶ Even nodes/clusters with the same IP addresses can run concurrently
- ▶ No impact on the network performance
- ▶ Access to storage under development
 - ▶ Multihomed data depots

- ▶ Czech Republic is building complex e-Infrastructure
- ▶ All layers/components present
- ▶ Close collaboration between university R&D and infrastructure operations
 - ▶ Duality CESNET and CERIT-SC
- ▶ Both conservative and innovative components
- ▶ Focus is on the *novelty*
 - ▶ in the way the e-Infrastructure is built and operated
 - ▶ in the way the e-Infrastructure is used
 - ▶ in the way why the e-Infrastructure is used
- ▶ The goal is to provide a meeting place for research

More information can be found on the corresponding web pages:

- ▶ CESNET: <http://www.cesnet.cz>, <http://meta.cesnet.cz>
- ▶ CERIT-SC: <http://www.cerit-sc.cz>
- ▶ IT4Innovations: <http://www.it4innovations.cz>



EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI



2007-13
OP Výzkum a vývoj
pro inovace

Projects CERIT Scientific Cloud, reg. no. CZ.1.05/3.2.00/08.0144,
and eIGeR, reg. no. CZ.1.05/3.2.00/08.0142,
are supported under the 3rd priority axis of Operation Programme
“Research and Development for Innovations”.
CESNET is supported as the large infrastructure of the Czech
Republic.