(Research) Infrastructures in the Czech Republic

Tomáš Rebok
CERIT-SC a MetaCentrum NGI
Masarykova Univerzita a CESNET z.s.p.o.

Roman Leontovyč
MetaCentrum NGI
CESNET z.s.p.o. a Univerzita Karlova

Lecturer

- **Tomáš Rebok**
  - CERIT-SC Centre @ Institute of Computer Science (ICS), MU
    - Head of the *Tom Rebok's Research Group*
  - *MetaCentrum @ CESNET z.s.p.o.*
  - senior researcher, head of applied-research projects
  - long-term activity in the field of computing and data infrastructures
    - supporting high-performance computations and data processing
  - primary orientation in the field of data infrastructures and analytics
    - data infrastructures and analysis
    - innovative approaches to specific data-analysis tasks
  - interdisciplinary research collaborations with
    - security forces & Police of the CR
    - environmental research groups

A bit of theory…

Supercomputers

- (extremely) powerful computers used for complex calculations and simulations
  - their performance is commonly measured in floating-point operations per second (FLOPS) instead of million instructions per second (MIPS)
- introduced in the 1960s (UNIVAC supercomputer)
- architecture – single, highly integrated systems
  - typically use massive number of specialized processors (up to millions) and high-speed interconnects to achieve their performance
- **pros:** high processing power, used for scientific research, weather forecasting and climate modelling, etc.
- **cons:** very expensive (specialized HW), limited accessibility, and high maintenance costs

EL CAPITAN – the fastest supercomputer (2025)

Location: Lawrence Livermore National Laboratory (California, US)

Performance: 1,742 petaFLOPS (1.742 exaFLOPS)

First online: November 2024

(Computing) Clusters

- consist of multiple interconnected “contemporary” computers (nodes)
  - they work together to perform tasks
- individual nodes may not be as powerful as a supercomputer
  - but the combined power of all the nodes can be significant
- architecture – clusters use standard, off-the-shelf hardware interconnected via a network
  - need for powerful & low-latency interconnection – Infiniband, Asterfusion, etc.
- **pros:** generally, more cost-effective (than supercomputers)
  - because they use standard hardware and can be scaled by adding more nodes (“horizontal scaling”)
- **cons:** powerful, but may not match the peak performance of a supercomputer (reliance on standard hardware and network communication), higher communication latency, more complex to maintain, not as energy-efficient (performance per watt)

(Computing) Cluster (long days ago ☺)

(Computing) Cluster (nowadays)

Supercomputers vs. (Computing) Clusters

Supercomputers – single, highly specialized systems designed for maximum performance

Computing clusters – collections of standard computers working together to achieve high performance through parallel processing

Remember: Neither supercomputers nor computing clusters can make your job to run significantly faster if it is not implemented for parallel/distributed processing.

(Super)Computing history in the Czech Republic

Early Developments (1990s):

- **1994:** the **Supercomputing Center Brno (SCB)** was established at **Masaryk University**
  - aimed to introduce high-performance computing technologies to the Czech academic community, providing advanced computational resources to researchers and students
  - predecessor of the **Institute of Computer Science MU** and the **CERIT-SC Centre @ MU**

SGI PowerChallenge XL supercomputer with 12 MIPS R10000 processors and 2GB RAM. The first supercomputer at SCB MU. 1994.

(Super)Computing history in the Czech Republic

Early Developments cont'd (1990s):

- **1995:** deployment of **five high-performance computers** across major Czech universities
  - Masaryk University and University of Technology in Brno, Charles University and Czech Technical University in Prague, and the University of West Bohemia in Pilsen
- **1996:** the **MetaCentrum project** was launched to integrate these dispersed computing resources into a cohesive national grid.
- **1999:** the MetaCentrum had become a strategic project under the **CESNET association**
  - focuses on creating a unified computational infrastructure to support diverse scientific research across the country
- since then, MetaCentrum operates the National Grid Infrastructure
  - and collaborates with the CERIT-SC and IT4Innovations on the computing/storage services provided to the Czech research community

(Super)Computing history in the Czech Republic

The core of the **initial supercomputer team** at SCB:

- Luděk Matyska (head)
- Aleš Křenek
- Zdeněk Salvet
- Martin Černohorský
- Miroslav Ruda

In front of the SGI Origin + Onyx supercomputer (2000):
40x CPU MIPS, R10000/195MHz, 18 GB RAM, 2x InfiniteReality2 graphics subsystem, 2x raster manager, 160MB frame buffer, 155 GB internal disks, 5x Fast Ethernet, 4x ATM, 270 GB on disk array.

(Super)Computing (Data)Centers and Infrastructures

To become a **(grid, cloud, container, ...) infrastructure**, the HW and SW components of the (Data)Centers have to be **managed by a solution** that supports distributed computing, resource pooling, and dynamic scalability, tailored to the specific type of the infrastructure.

For example:

- **Grid infrastructure** requires technologies like OpenPBS or UNICORE, which enable resource sharing and task distribution across geographically dispersed systems
- **Cloud infrastructure** relies on platforms such as OpenStack, AWS, or Azure, which provide virtualization, automation, and on-demand resource provisioning
- **Container infrastructure** depends on orchestration tools like Kubernetes or Docker Swarm, which manage containerized applications and ensure efficient deployment, scaling, and networking

These technologies ensure that the hardware and software components work together, and deliver:

- the desired infrastructure type
- computing and storage services
- flexibility, efficiency, and scalability

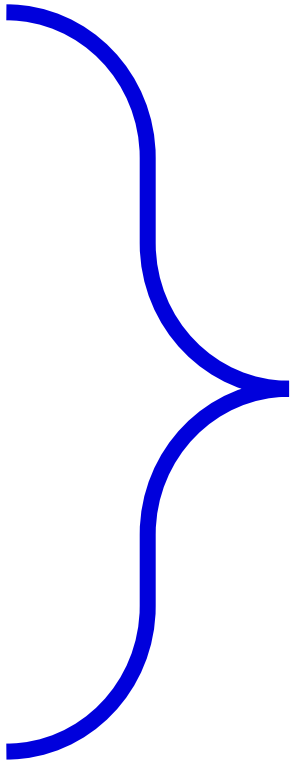
National Computing Infrastructures

supporting research communities

(Super)Computing Centres in the CR

available in 3 infrastructures (centers) in the Czech Republic:

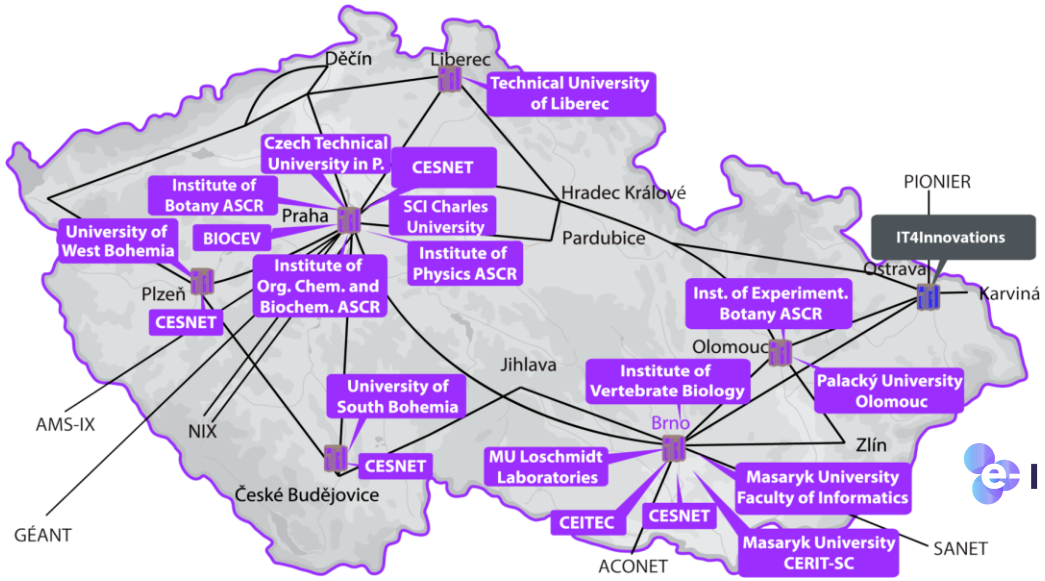
- Cesnet/**MetaCentrum**
  - **grids/PBS**
  - **clouds**
  - **specialized computing**
- MUNI/**CERIT-SC**
  - **grids/PBS**
  - **clouds and containers**
  - **specialized computing**
- VŠB-TUO/**IT4Innovations**
  - **grids/PBS**
  - **clouds**



e-INFRA CZ
https://www.e-infra.cz

MetaCentrum @ CESNET

- activity of the **CESNET Association**
  - CESNET – an association founded (and supported) by Czech universities and Czech Academy of Science
    provides services to universities + own research
- since 1996, the **Coordinator of the National Grid Infrastructure (NGI)**
  - **originally established at MUNI** (Supercomputing Centre Brno, SCB, 1994)
- **integrates large/medium-sized HW centers** (clusters, powerful servers and storage) **of several universities/organizations** within the CR
  - → provides an environment for (collaborative) computing and data processing
- Integrated into the European Grid Infrastructure (EGI.eu)



MetaCentrum NGI

available to all the employees and students of Czech universities/universities, the Academy of Sciences of the Czech Republic, research institutes, etc.

- commercial entities for public research only

http://metavo.metacentrum.cz

offers:

- Computational resources
- Storage capacities
- Application programs

After registration available completely free of charge

"payment" in the form of publications with acknowledgements

NGI – available computing hardware

computing resources: ca 46750 cores (x86_64)

- Nodes with a lower number of performance cores:
  - 2x4-8 cores
- Medium core nodes (SMP machines):
  - 32-80 cores
- Up to 10 TB memory per node

Nodes with a high number of cores: SGI UV 2000

- 504 cores (x86_64), 10 TB RAM
- 384 cores (x86_64), 6 TB RAM

Other "exotic" hardware:

- nodes with GPU accelerators (**Nvidia DGX with H100s**), etc.

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

NGI – available storage hardware

approx. 33 PB for working data

- repository in Brno, Pilsen, České Budějovice, Liberec, Prague
- user quota of 1-3 TB on each of the storages

approx. 80+ PB for long-term/archive data

- HSM – tape libraries, MAIDs (massive array of idle drives)
- CEPH object storage (analogy to Amazon S3)

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

NGI – available software

~500 different apps / ~3000 modules (installed on request)

- see https://docs.metacentrum.cz/software/alphabet/

continuously maintained development environment

- GNU, Intel, PGI, debugging and optimization tools (TotalView, Allinea), ...

generic math software

- Matlab, Maple, Mathematica, gridMathematica, …

commercial and free software for application chemistry

- Gaussian 09, Gaussian-Linda, Gamess, Gromacs, Amber, …

material simulations

- ANSYS Fluent CFD, Ansys Mechanical, Ansys HPC…

structural biology, bioinformatics

- CLC Genomics Workbench, Geneious, Turbomole, Molpro, …

a range of freely available packages

…

NGI – basic characteristics

after registration, all the resources are available without any administrative burden

- → ~immediately (depending on actual usage)
- no applications for resources

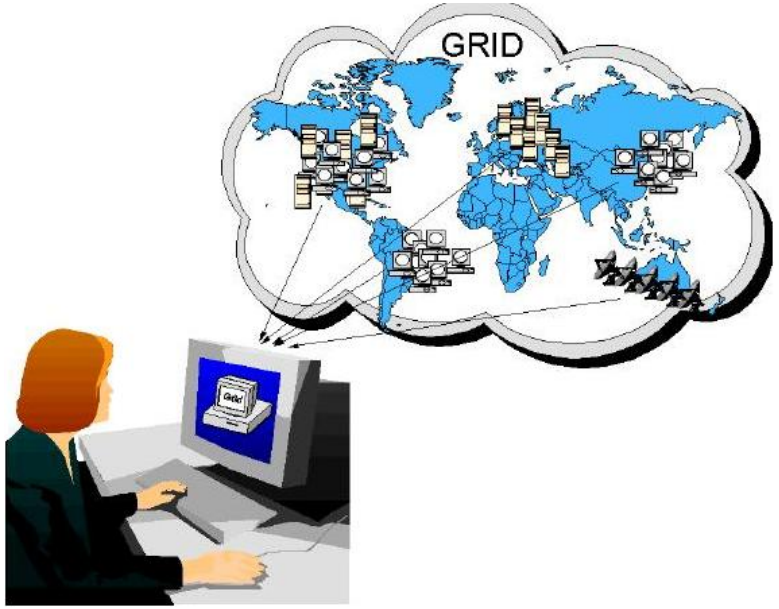
user accounts extensions every year

- validates users‘ relationship to an academic institution
  - federated infrastructure eduID.cz used for minimalization of users‘ burden
- reports of user publications with acknowledgements to MetaCentrum/CERIT-SC
  - used as a proof of infrastructure benefits for Czech research area

best-effort service

- however, still usually 24x7

NGI – how to compute?



Batch jobs

- descriptive task script
- job start and end notifications

Interactive jobs

- text and graphic mode

Cloud computing

- users don't run jobs, but virtual machines
  - for research only

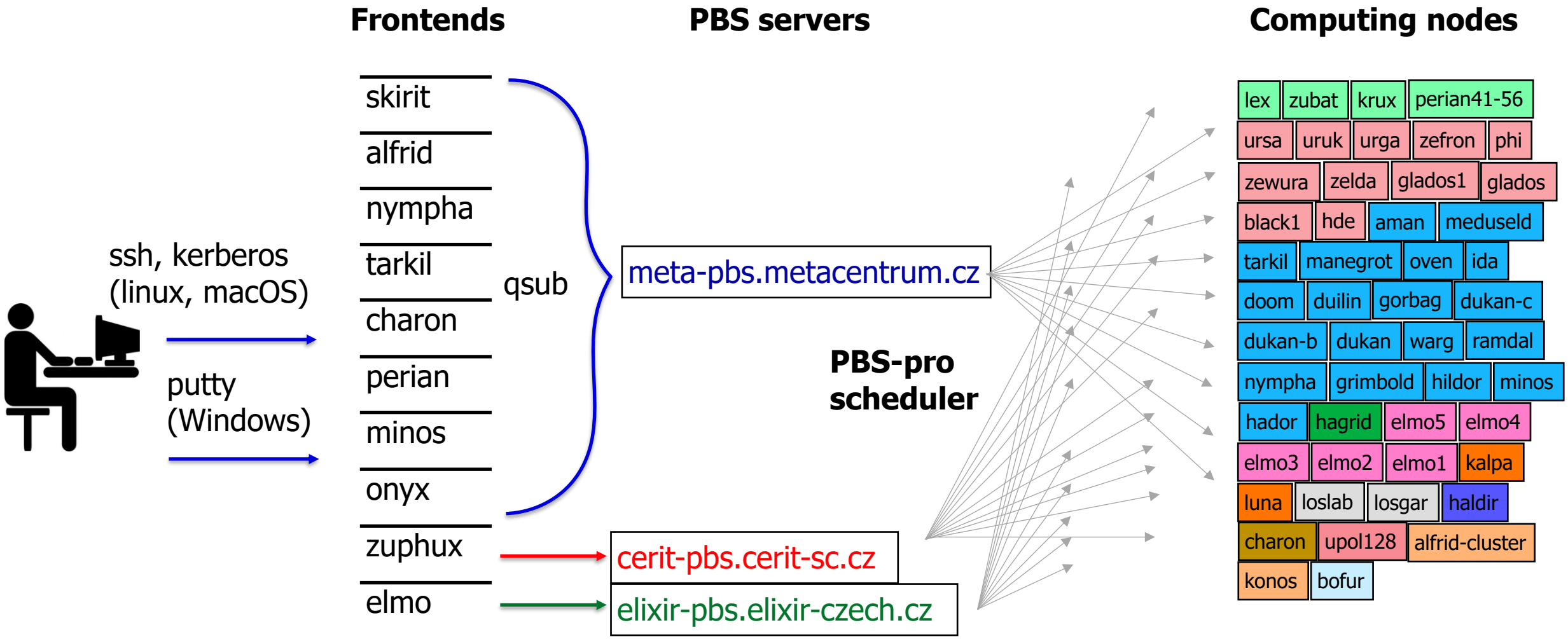
graphical applications and virtual desktops in the browser environment

- Open OnDemand, Rancher, JupyterHub, …

specialized environments

- Apache Hadoop, Galaxy, …

NGI under the hood – how to connect?



NGI under the hood – **in numbers** ...

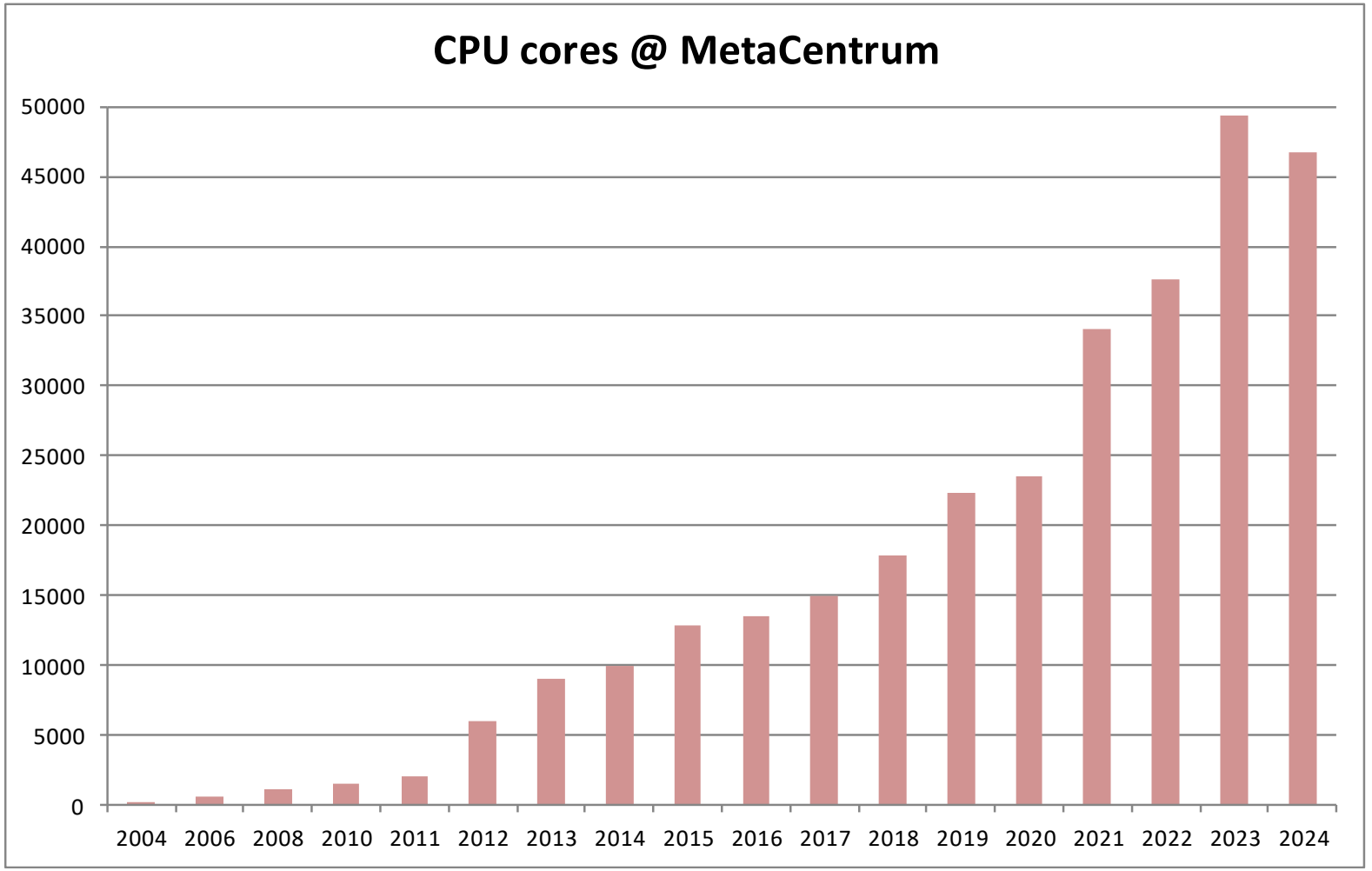
approx. 46750 computing cores, approx. 700 nodes

- and 462 of GPU accelerators

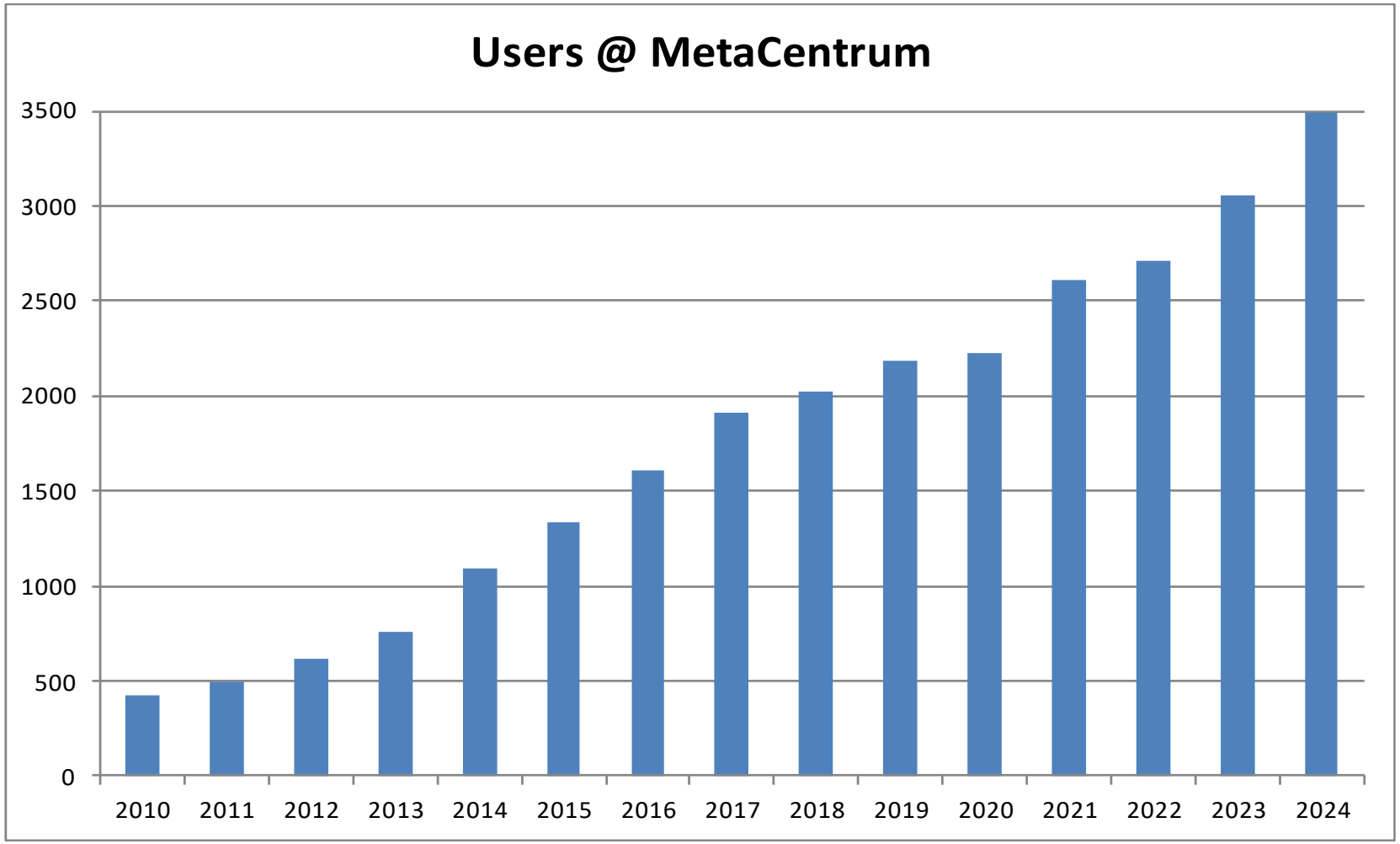
year 2024:

- 3490 users (31.12.2024)
- ca 15 million of jobs running
  - ca 41100 jobs per day
  - ca 4300 jobs per user
- approx. 37.5 thousands of CPUyears computed in total
  - and 357 of GPUyears

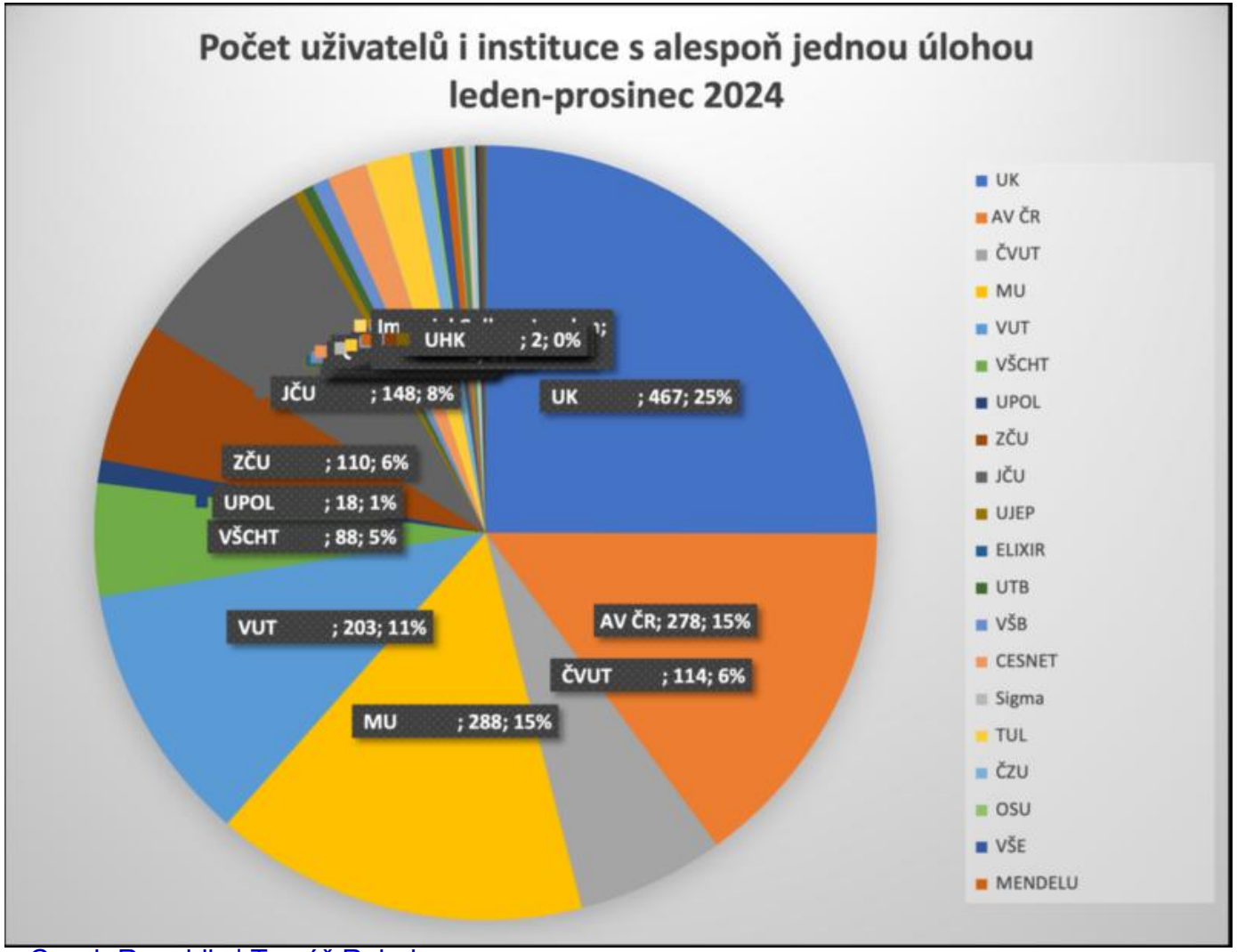
NGI under the hood – **and graphs** ...



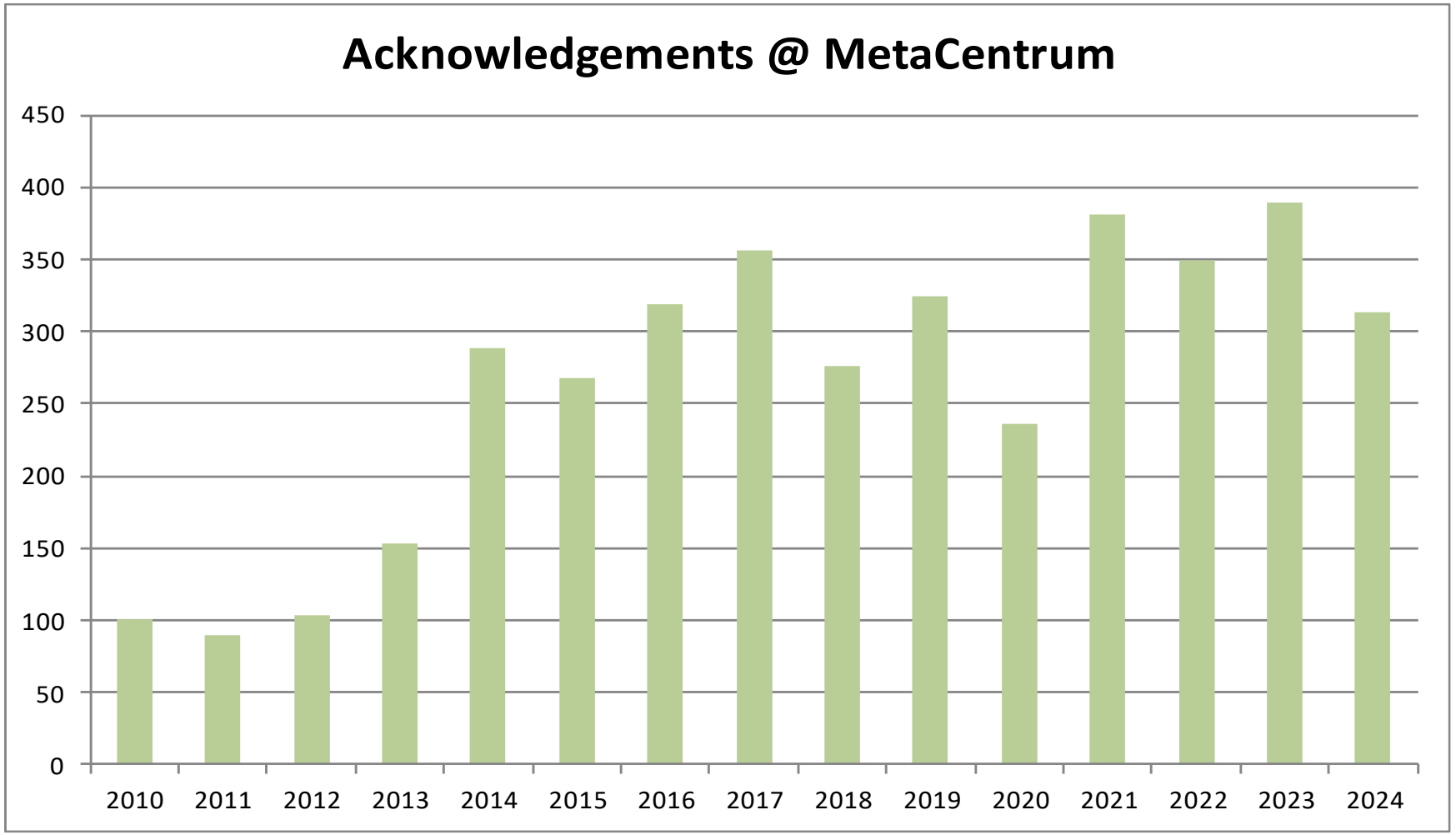
NGI under the hood – **and graphs** ...



NGI under the hood – **and graphs** ...



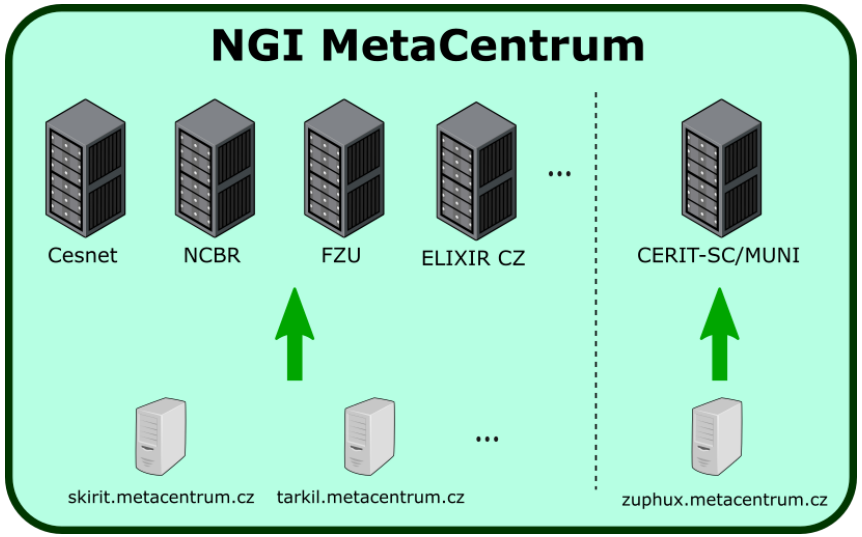
NGI under the hood – **and graphs** ...



MetaCentrum NGI & **Resource integration I.**

MetaCentrum and CERIT-SC

- MetaCentrum provides own HW resources (CESNET) and integrates resources of external providers
  - CERIT-SC/MUNI is one of them
  - others are CEITEC/NCBR, FZU, ČVUT, JČU, ZČU, UPOL, MU, TUL, etc.
    as well as global projects like ELIXIR CZ



+ shared storages

and shared SW apps

MetaCentrum NGI & **Resource integration II.**

resource owners (usually) have priority access to their HW resources

- under agreed conditions
- technically accomplished using specific scheduler queues
  - more details later

MetaCentrum NGI & **Support of external partners**

Assistance with:

- purchase and integration computational resources into NGI
- selection, installation and maintenance of the clusters
- software maintenance
- maintenance of user accounts
- priority/exclusive access to owned clusters

CERIT-SC Centre

CERIT-SC Centre – a research centre built at ICS MU

- originally Supercomputing Centre Brno (SCB)

1. provider of HW and SW resources

- esp. the special ones (DGX, UV 2000)
- part of the MetaCentrum NGI

2. services beyond the scope of a “common" HW centre

- **interdisciplinary research**
  - cooperation of IT researchers and partners from other fields

CERIT-SC Centre

main objectives of the CERIT-SC @ MUNI Centre:

- flexible infrastructure, own research in infrastructure areas
- three main **research pillars**:
  - *High-performance computing* – acceleration of calculations, GPU computing, ...
  - *Artificial Intelligence* – application of artificial intelligence and machine learning methods (esp. in biology)
  - *Data Infrastructures and Big Data analytics* – design and implementation of data analytics infrastructures and

    Tom Rebok's Research Group

CERIT-SC Centre – key research activities

High-Performance Computing (HPC) – accelerating research by providing powerful computing resources

- supporting a wide range of scientific fields including artificial intelligence, data science, and computational biology
- including sensitive data processing and GPU acceleration support

Data Science – unified analysis of vast, heterogeneous data (data collection, processing, analysis, storage, and sharing)

- development of (both national and international) data and analysis infrastructures

Artificial Intelligence (AI) – application of AI / ML techniques in various applied solutions (biology, crime & security, data management, etc.)

EOSC / FAIR Data – promoting Findable, Accessible, Interoperable, and Reusable (FAIR) data principles (see later)

Digital Identity and Access Management – management of digital identities and controlled access to services (Perun AAI ecosystem)

CERIT-SC Centre – research collaborations

How are the research collaborations performed?

- the work is carried via a doctoral/diploma 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

Collaborations through (international) projects

- CERIT-SC participates on several projects, usually developing IT infrastructure supporting the particular research area
  - ELIXIR-CZ, BBMRI, Thalamoss, SDI4Apps, Onco-Steer, CzeCOS/ICOS, …
  - KYPO, 3M SmartMeters in cloud, MeteoPredictions, …

Strong ICT expert knowledge available

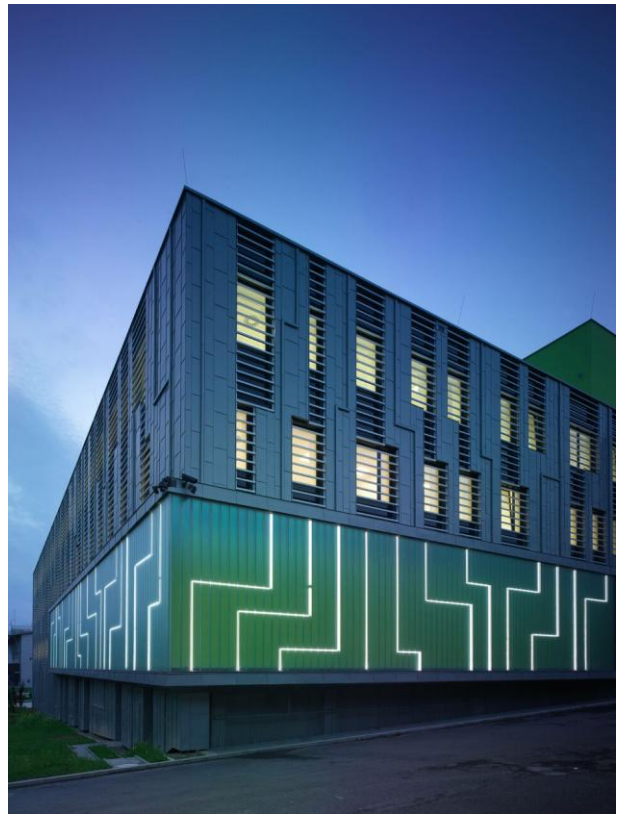
- long-term collaboration with Faculty of Informatics MU
- long-term collaboration with CESNET
  - → consultations with experts in particular areas

https://www.it4i.cz/en/for-users/computing-resources-allocation

IT4Innovations

IT4Innovations – national supercomputing centre at VSB TUO in Ostrava

- established in 2012
- provides cutting-edge computational resources and support for research and innovation in HPC, artificial intelligence, big data analytics, and related fields
  - **one of the leading supercomputing centers in Europe**
- currently available supercomputers
  - Karolina, Barbora, LUMI, NVIDIA DGX-2
- available to academic staff as well as the commercial bodies



services: HW provider and interdisciplinary research

- own research laboratories
- research cooperation with the centre's users

computing time has to be officially requested

- so-called grant competitions (every 6 months)
  - subsequently, dedicated computing time
  - (might require financial participation)

	NVIDIA DGX-2	Barbora	Karolina	LUMI
Put into operation	Spring 2019	Autumn 2019	Summer 2021	Winter 2023
Theoretical peak performance	130 TFlop/s	849 TFlop/s	15.7 PFlop/s	531.5 PFlop/s
Operating system	CentOS 7	RHEL 8	Rocky Linux 8.x	HPE Cray OS
Compute nodes	1	201	831	5,042
Types of compute nodes	**1 GPU node** 2x Intel Xeon Platinum 8168, 24 cores, 2.7 GHz, 1.5 TB RAM, 16x NVIDIA Tesla V100, 32 GB HBM2	**192 CPU nodes** 2x Intel Cascade Lake 6240, 18 cores, 2.6 GHz, 192 GB RAM **8 GPU nodes** 2x Intel Skylake 6126, 12 cores, 2.6 GHz, 192 GB RAM, 4x NVIDIA Tesla V100, 16 GB HBM2 **1 SMP node** 8x Intel Xeon 8153, 16 cores, 2.0 GHz, 6 TB RAM	**756 CPU nodes** 2x AMD EPYC 7h12, 64 cores, 2.6 GHz, 256 GB RAM (of which 36 nodes used for Cloud services) **72 GPU nodes** 2x AMD EPYC 7763, 64 cores, 2.45 GHz, 1 TB RAM, 8x NVIDIA A100, 40 GB HBM2 **1 data analytics node** 32x Intel Xeon-SC 8628, 24 cores, 2.9 GHz, 24 TB RAM **2 visualisation nodes** 2x AMD EPYC 7452, 32 cores, 2.35 GHz, 256 GB RAM, 1x NVIDIA RTX 6000 GPU	**2,048 CPU nodes** 2x AMD EPYC 7763, 64 cores, 2.45 GHz, 256 – 1,024 GB RAM **2,978 GPU nodes** 1x AMD EPYC 7A53, 64 cores, 2.45 GHz, 512 GB RAM, 4x AMD Instinct MI250X GPUs, 128 GB HBM2e **8 data analytics nodes** 2x AMD EPYC 7742, 64 cores, 2.25 GHz, 4 TB RAM **8 visualisation nodes** 2x AMD EPYC 7742, 64 cores, 2.25 GHz, 2 TB RAM, 8x NVIDIA A40 GPU
Accelerators in total	16x NVIDIA Tesla V100	32x NVIDIA Tesla V100	576x NVIDIA Tesla A100, 2x NVIDIA RTX 6000	11,912x AMD Instinct MI250X, 8x NVIDIA A40
CPU cores in total	48	7,232	106,880	454,784
Storage	30 TB NVMe	29 TB / home 310 TB / scratch (28 GB/s)	30 TB / home 1,275 TB / scratch (NVMe, 730 GB/s SWP, 1,198 GB/s SRP)	81 PB / (home + project + scratch) (240 GB/s)
Interconnection	Infiniband EDR 100 Gb/s	Infiniband HDR 200 Gb/s	Infiniband HDR 200 Gb/s	200 Gb/s Slingshot-11

IT4Innovations – statistics

Distribution of computational resources in 2023

- 82% Open Access projects
  6,038,771 of used node hours*
- 15% EuroHPC JU Grant Competitions
  1,076,824 of used node hours
- 3% Access to Thematic HPC Resource Utilisation, including the rental of computational resources
  228,545 of used node hours

Computational resources allocated within the Open Access Grant Competitions in 2023 by institutions

- 34% Czech Academy of Sciences (50 projects)
- 18% VSB – Technical University of Ostrava (30 projects)
- 11% Brno University of Technology (24 projects)
- 8% Masaryk University (11 projects)
- 8% Czech Technical University in Prague (38 projects)
- 7% Charles University (25 projects)
- 4% Palacký University in Olomouc (13 projects)
- 3% University of Chemistry and Technology in Prague (8 projects)
- 3% CESNET (1 project)
- 2% University of Ostrava (4 projects)
- 1% CEITEC (7 projects)
- 1% Others (ELI Beamlines, University of South Bohemia in České Budějovice, Technical University of Liberec, University Hospital Hradec Králové, University of West Bohemia in Pilsen, Czech Aerospace Research Centre, University Hospital Ostrava)

e-INFRA CZ – how to become a user?

submit an application

- https://docs.e-infra.cz/cs/account/ , section „Account creation“
- EduID.cz => verification of your academic identity will be made using your home institution

learn about the documentation and basics of Linux OS

- http://metavo.metacentrum.cz , section „Documentation“
- practical seminars: https://metavo.metacentrum.cz/cs/seminars/index.html
- https://www.abclinuxu.cz/ucebnice/zaklady

compute

- *NGI and CERIT-SC:* no need for submitting any requests for computing time
- *IT4Innovation:* participate in the grant competitions

e-INFRA CZ – selected services for end users

FileSender – web service for sending large files

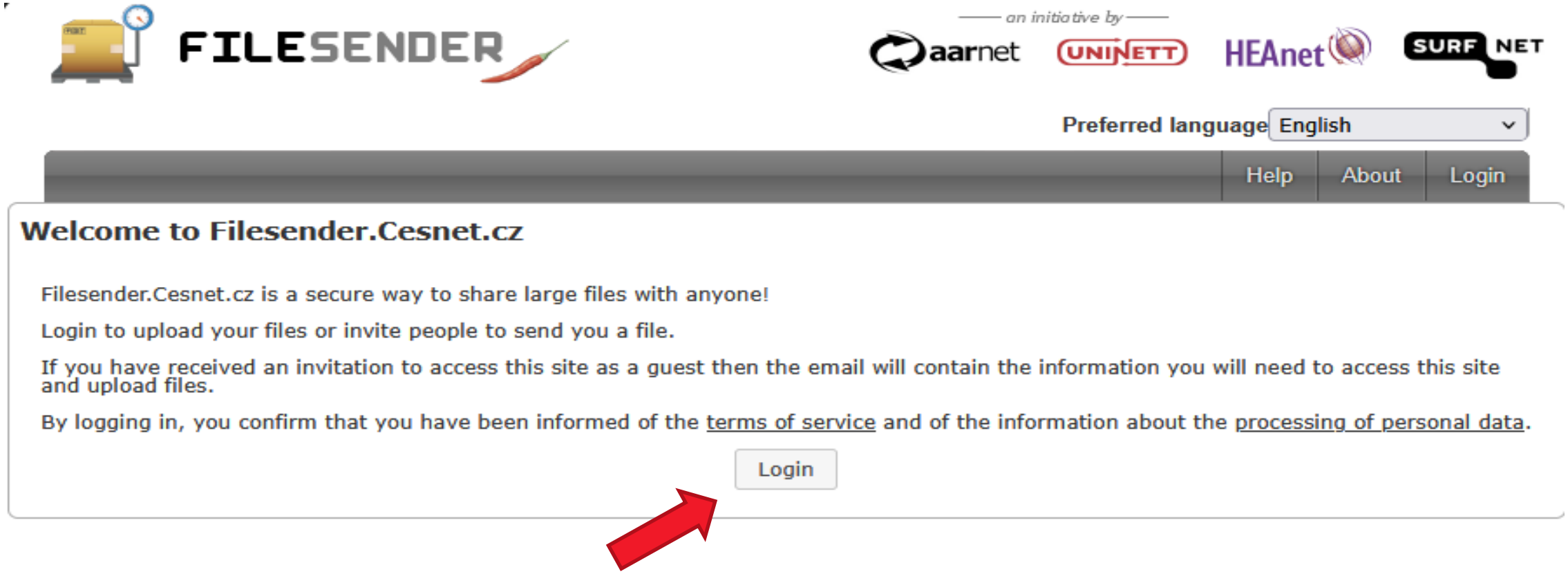
- current limit: 2 TB (~2000 GB)
- expiration time: up to 1 month

http://filesender.cesnet.cz

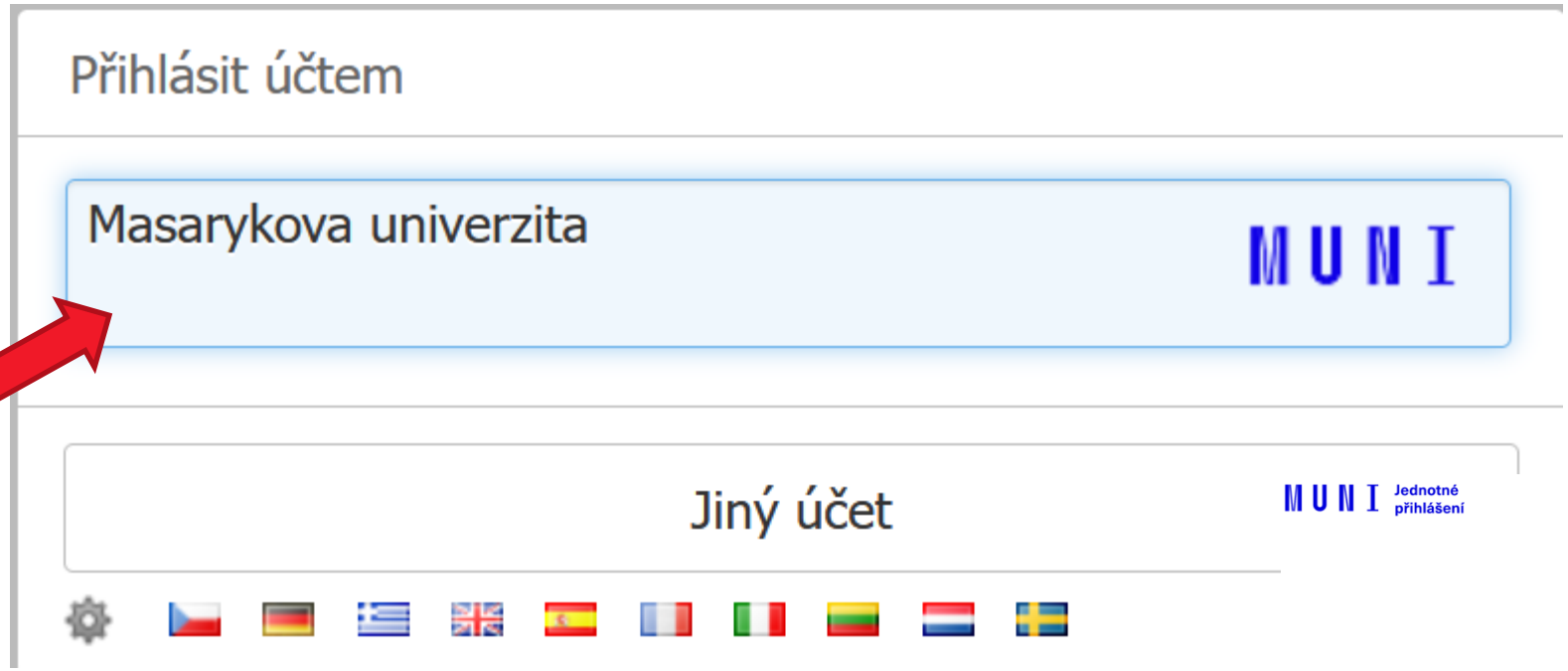
Either the sender or recipient must be an authorized academic staff member

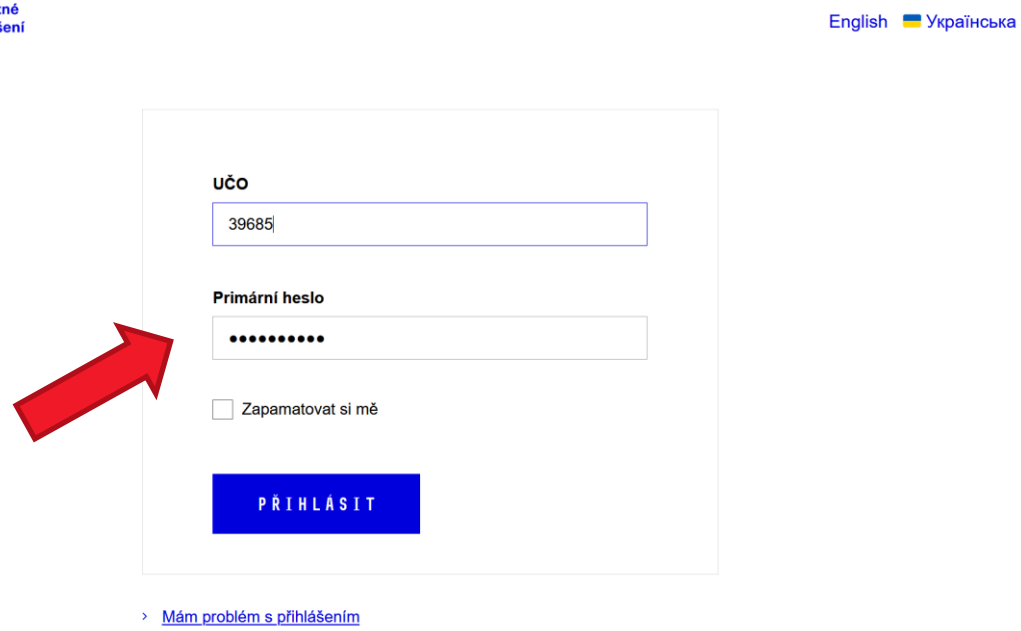
- an authorized user can **send data files to any user**
  - may include email notifications about the data lifecycle
- an authorized user can **send an invitation to receive data files from any user**

FileSender – how to use



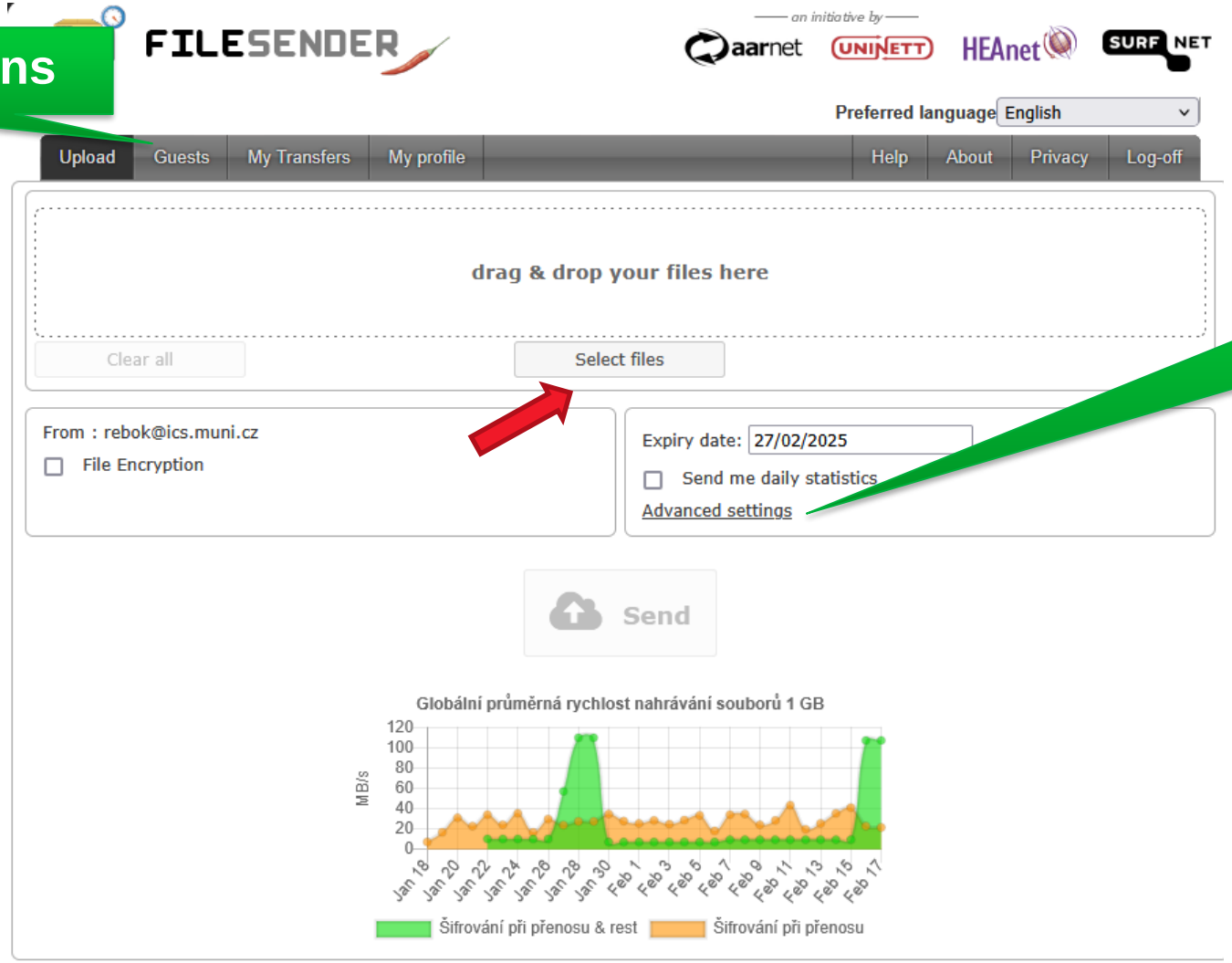
FileSender – how to use





FileSender – how to use

vouchers/invitations



advanced
notifications,
reference links, etc.

e-INFRA CZ

e-INFRA CZ – selected services for end users

OwnCloud – cloud storage a la Google Drive or Dropbox

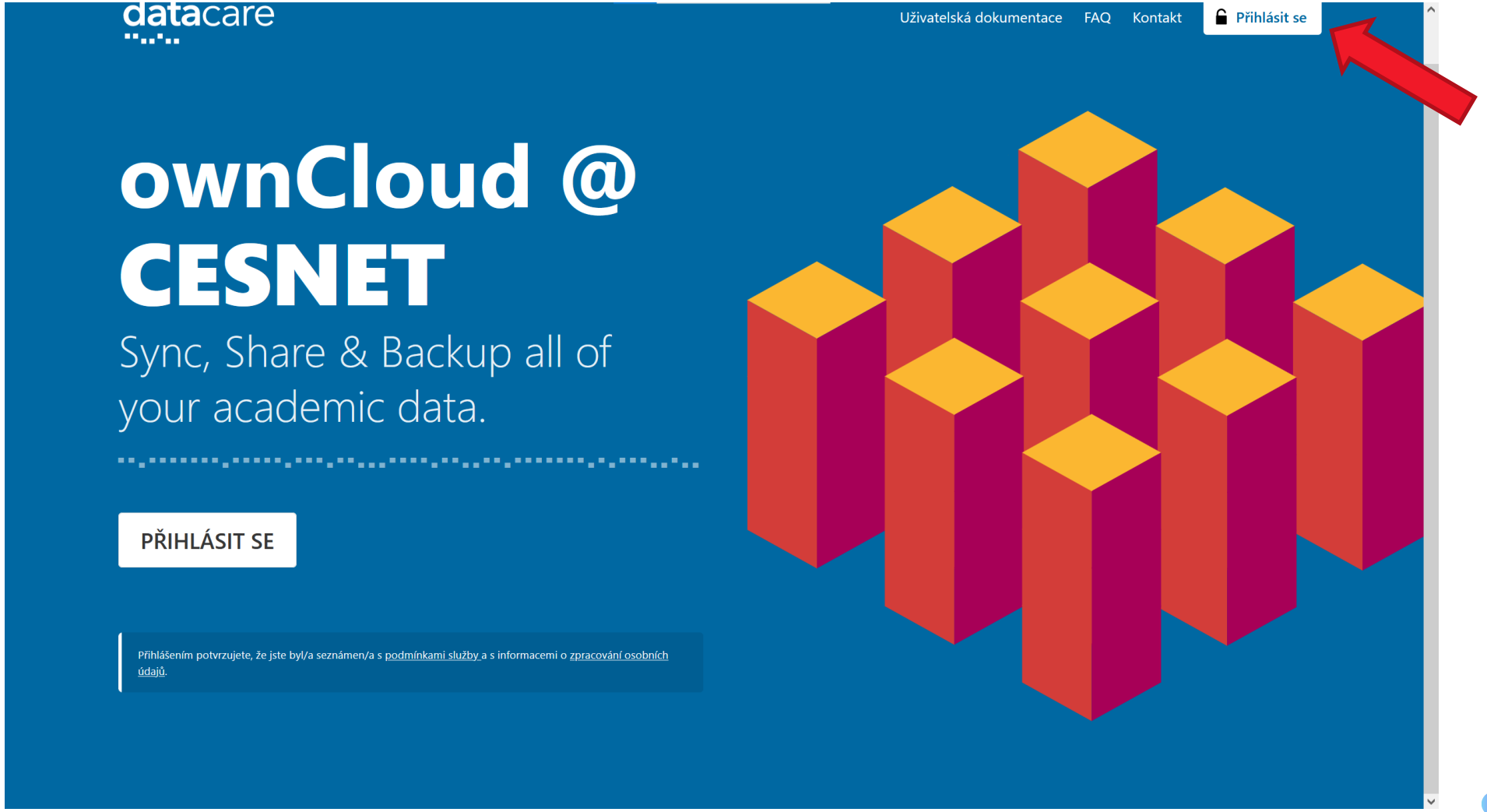
- current quota: 100 GB / user

https://owncloud.cesnet.cz/

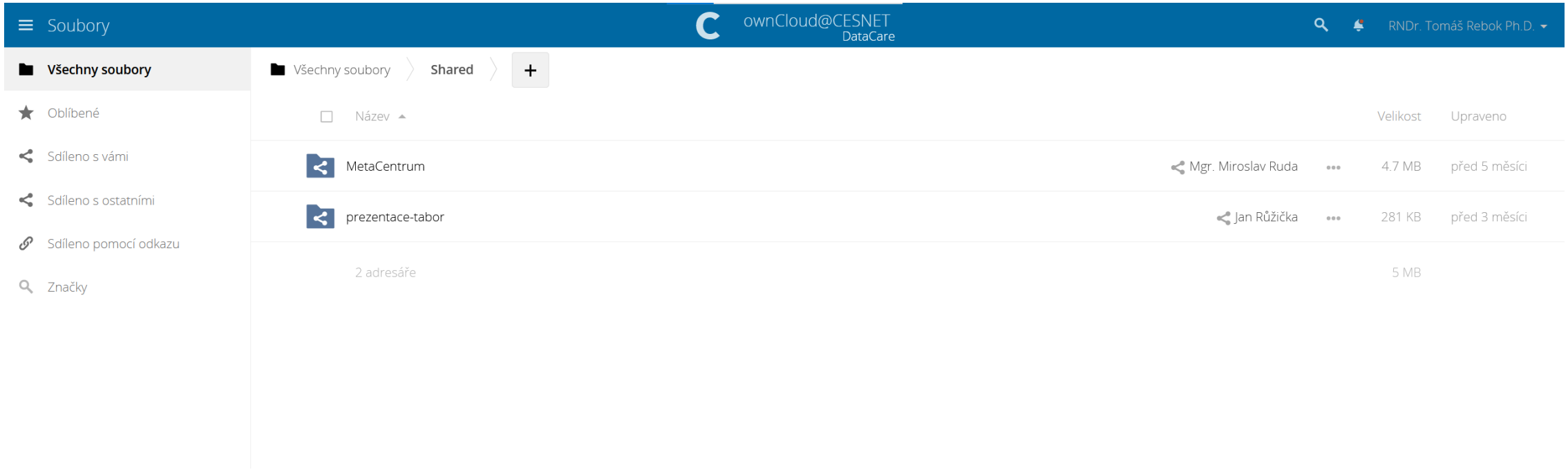
Synchronization and availability of data between devices

- clients available for OS Windows, Linux, OS X
- available for smartphones and tablets as well
- enables data sharing among users
- provides backups
- etc.

OwnCloud – how to use



OwnCloud – how to use



e-INFRA CZ – selected services for end users

Open OnDemand – a user-friendly UI interface to interactive computing

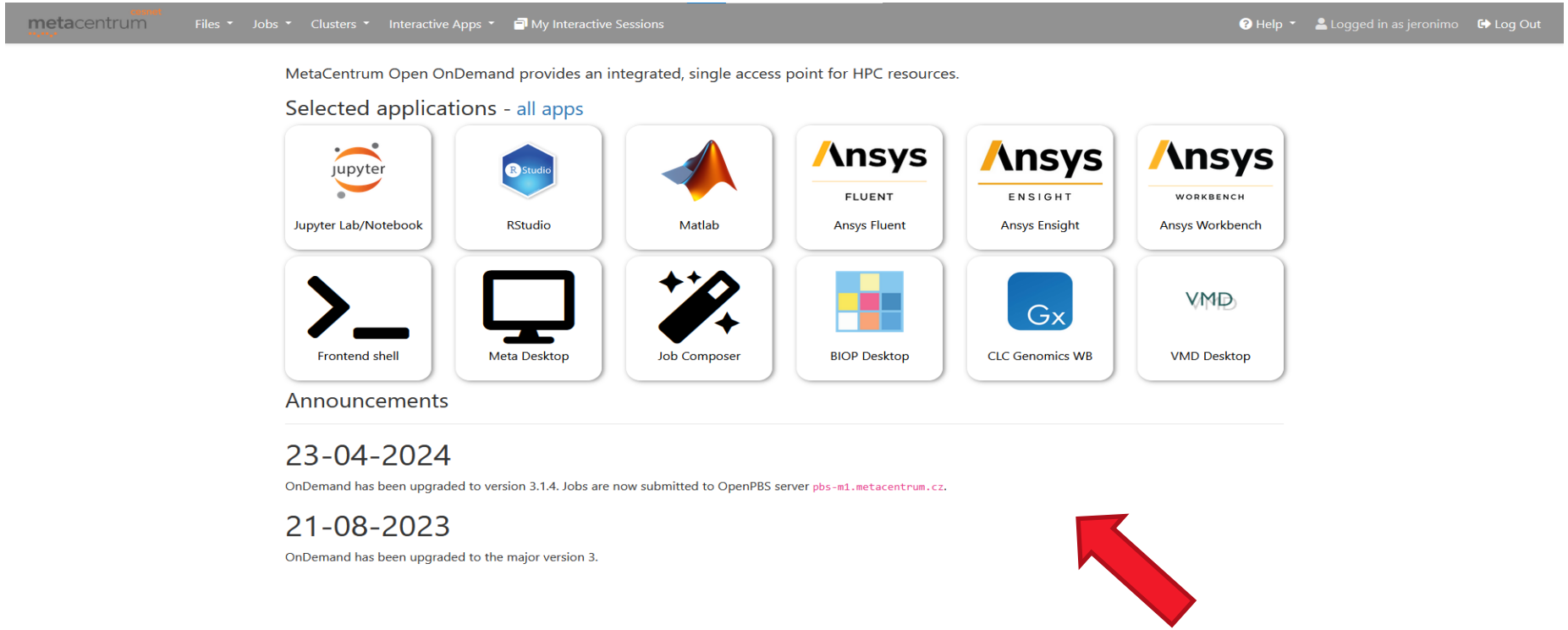
- simplifies interaction with supercomputing systems

https://ondemand.metacentrum.cz

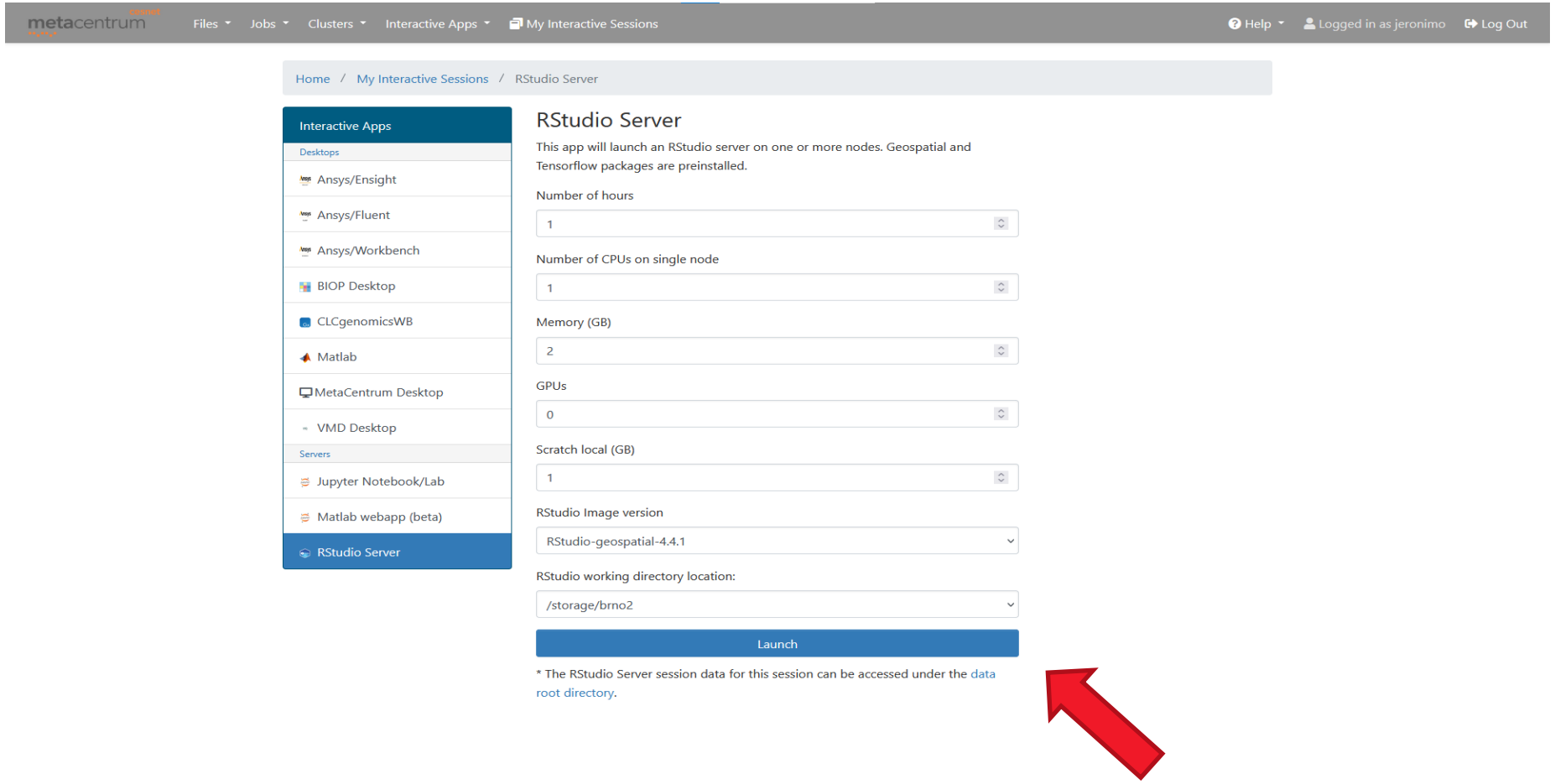
Key Features:

- **User-Friendly Interface –** no command-line expertise required
- **File Management –** easily upload, download, and manage files
- **Job Submission –** submit, monitor, and manage HPC jobs via a web browser
- **Interactive Apps –** run GUI-based applications (e.g., Jupyter Notebook, MATLAB, RStudio)
- **Remote Visualization –** visualize data and results directly in the browser

OnDemand – how to use (after federated login)

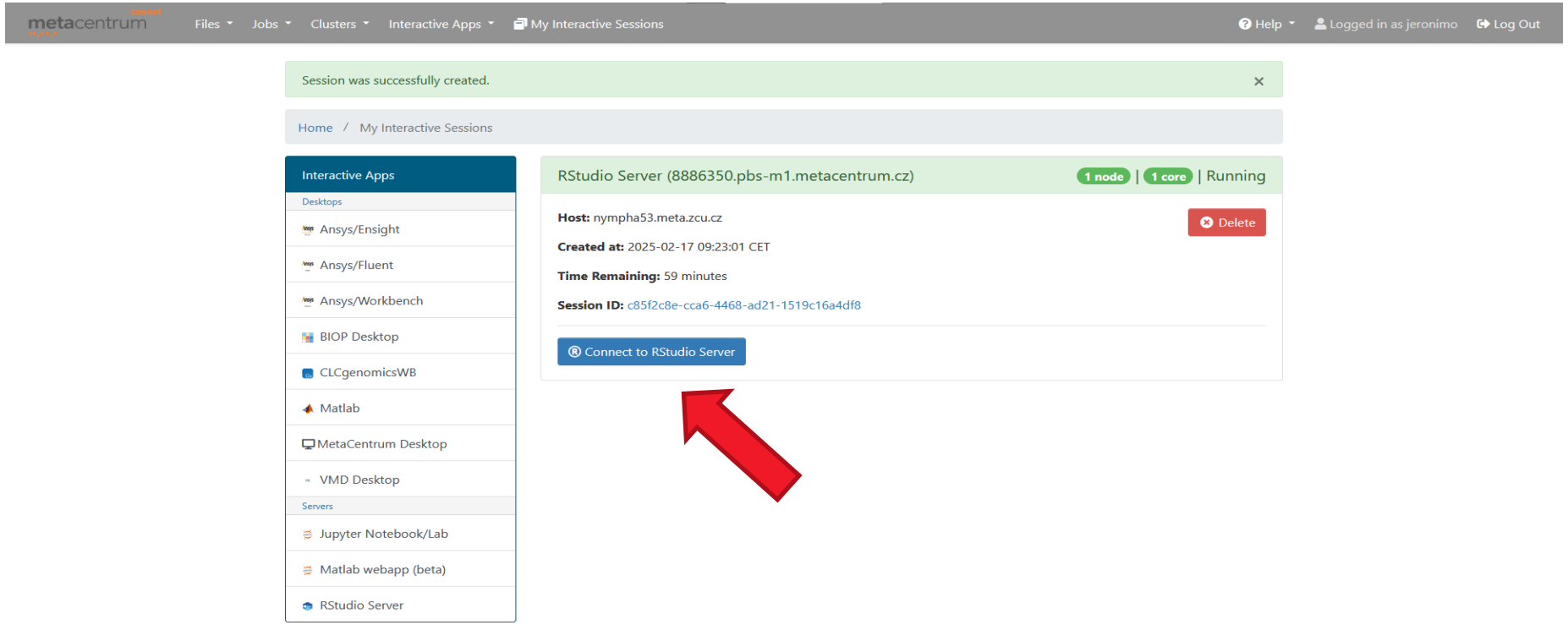


OnDemand – how to use



e-INFRA CZ

OnDemand – how to use



e-INFRA CZ – selected services for end users

Rancher (Kubernetes) – open-source platform for managing Kubernetes clusters

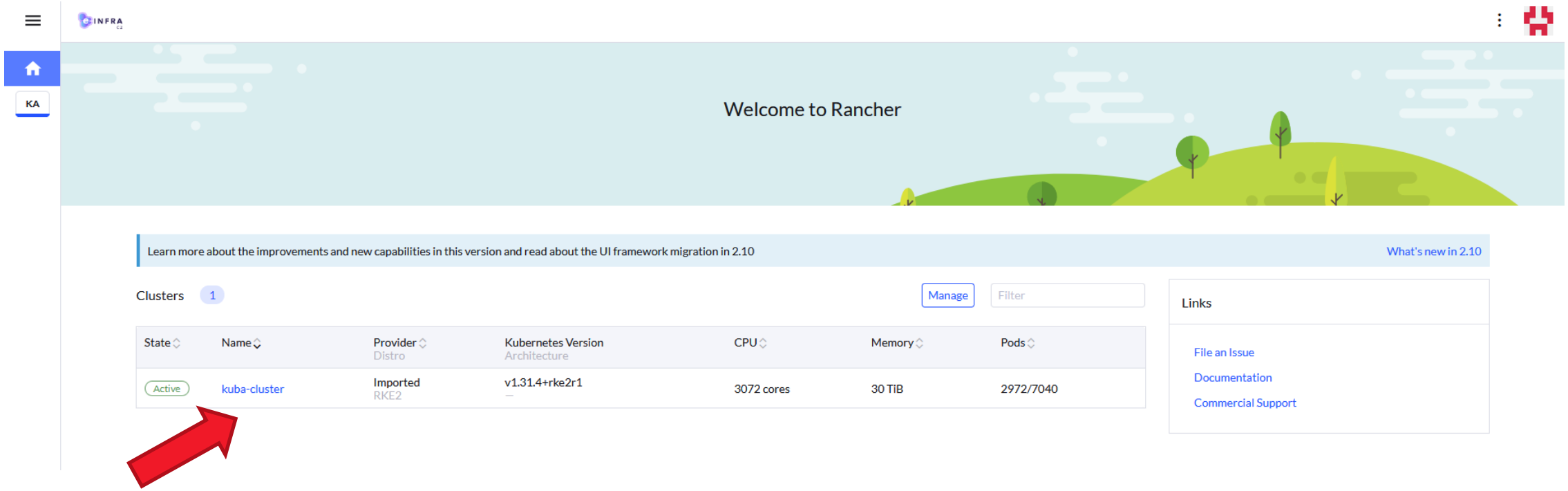
- simplifies deployment, scaling, and management of containerized applications

https://rancher.cloud.e-infra.cz

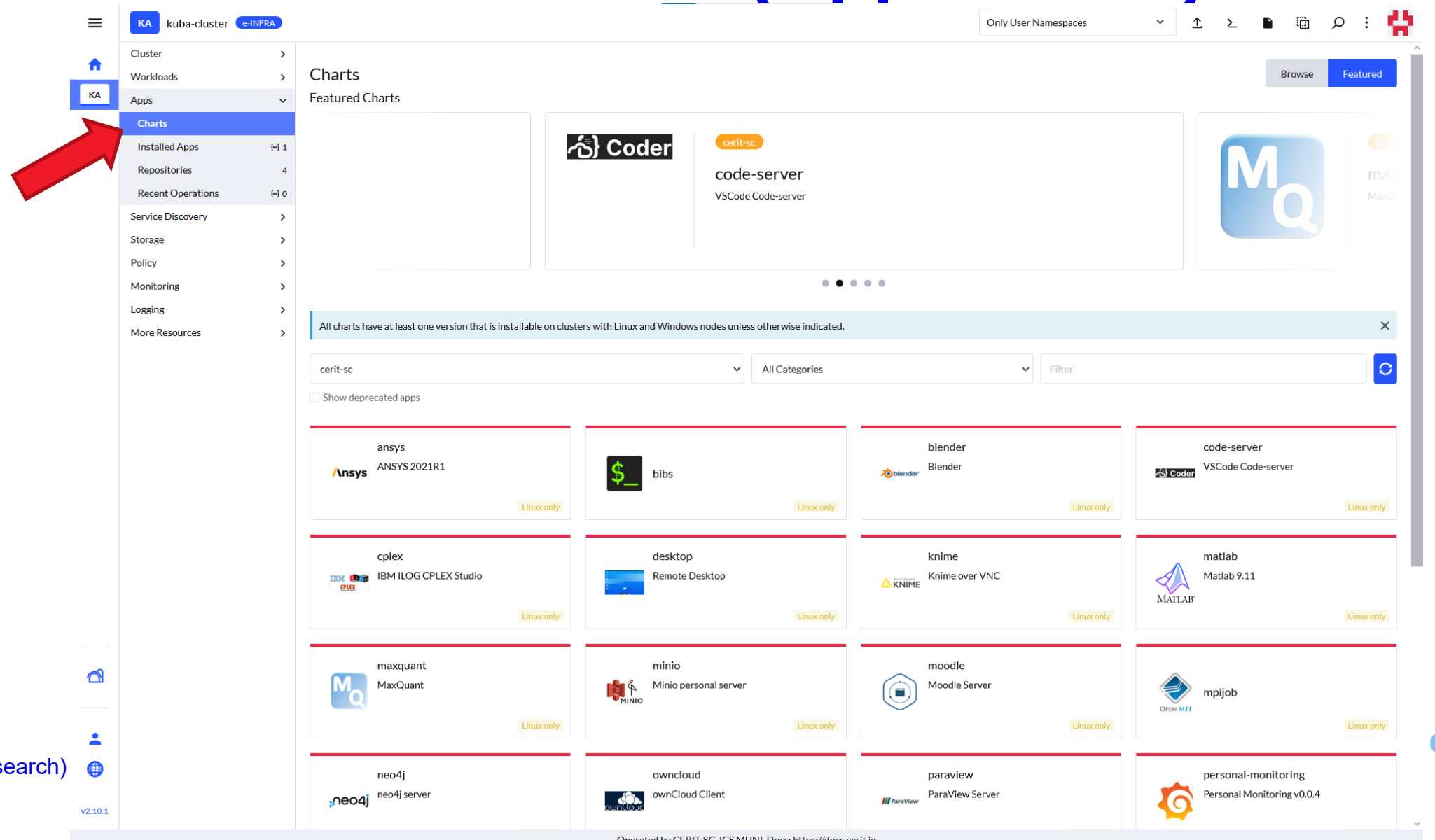
Key Feature:

- reduces complexity of Kubernetes management as well as container deployment

Rancher – how to use (after federated login)



Rancher – how to use (Apps menu)



e-INFRA CZ – selected services for end users

Foldify – a web application based on **AlphaFold** tools

- predicting **protein structures** with high accuracy

https://foldify.cloud.e-infra.cz/

AlphaFind – a web-based search engine that allows for structure-based search of the entire AlphaFold Protein Structure Database

- developed by FI MUNI in cooperation with CERIT-SC

https://alphafind.fi.muni.cz/

European Computing & Data Research Infrastructures

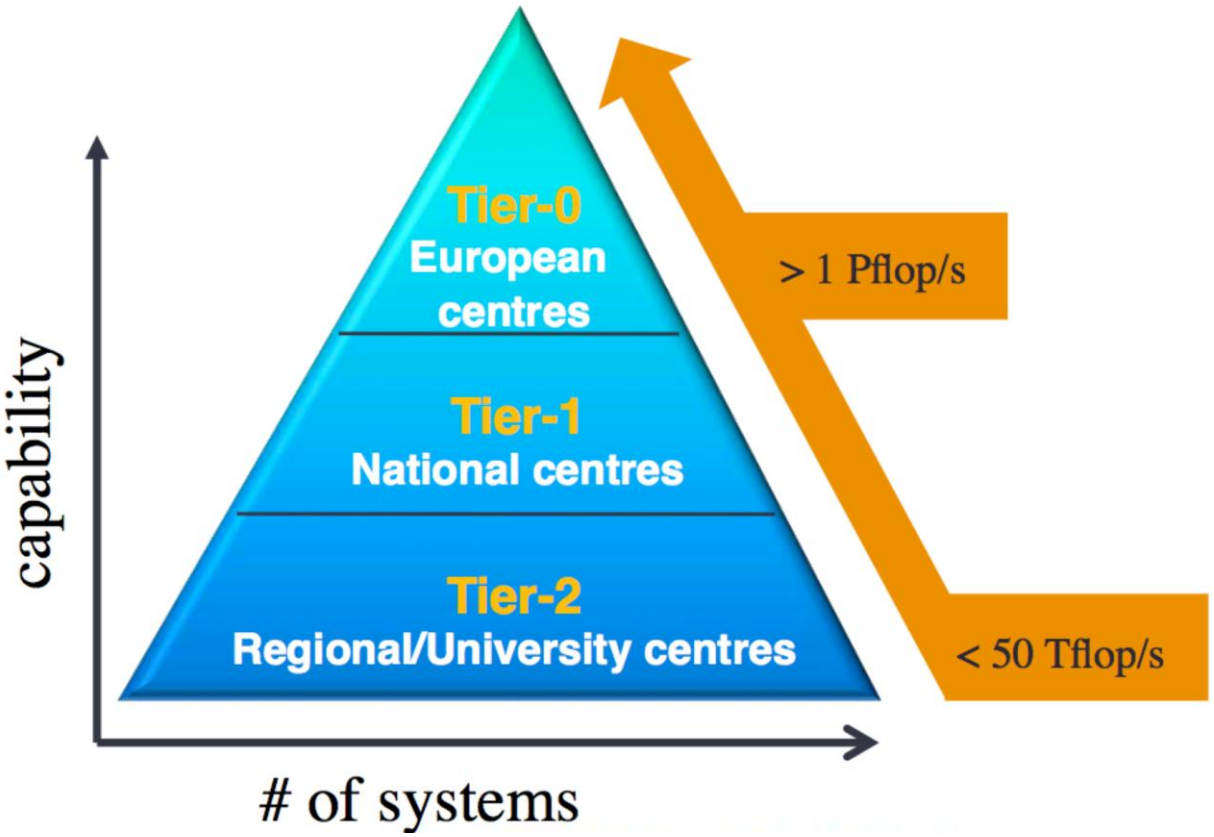
European HPC Infrastructures

Structured provision of European HPC facilities:

- Tier-0: European Centres (> petaflop machines)
- Tier-1: National Centres
- Tier-2: Regional/University Centres

Tiers planned as part of an EU Research Infrastructure Roadmap

This is coordinated through "PRACE" – http://prace-ri.eu



PRACE

Partnership foR Advanced Computing in Europe

- international non-profit association (HQ office in Brussels)
- established in 2010 following ESFRI roadmap to create a persistent pan-European Research Infrastructure (RI) of world-class supercomputers
  - ESFRI = European Strategy Forum on Reseach Infrastructures

Mission: enable high-impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society

- provide access to leading-edge computing and data management resources and services for large-scale scientific and engineering applications at the highest performance level

PRACE members

Currently 25 members:

- Austria
- Belgium
- Bulgaria
- Cyprus
- Czech Republic
- Denmark
- Finland
- France
- Germany
- Greece
- Hungary
- Ireland
- Israel
- Italy
- The Netherlands
- Norway
- Poland
- Portugal
- Slovakia
- Slovenia
- Spain
- Sweden
- Switzerland
- Turkey
- UK

CR participates through IT4Innovations.

e-INFRA CZ

EGI.eu (European Grid Infrastructure)

EGI.eu

- a federation of computing and storage resource providers united by a mission to support data-intensive research with a wide range of advanced computing services
- established to create a persistent pan-European Research Infrastructure (RI) that facilitates scientific discovery and innovation

Key Features:

- **Federated e-Infrastructure:** comprises national computing and data centers
- **Technologies:** Grid/PBS computing, cloud computing, containers
- **Services:** HPC, cloud computing platforms, data management, and analytics
- **Community Support:** engages with diverse user communities to identify needs, provide support, and drive innovation

CR participates through MetaCentrum NGI.

EUDAT (European Data Infrastructure)

EUDAT

- a comprehensive set of research data services, expertise, and technology solutions to all European scientists and researchers
- established to create a persistent pan-European Research Infrastructure (RI) that supports data stewardship and management

Key Features:

- **Federated Data Infrastructure:** distributed across 15 European nations, integrating data storage with some of Europe's most powerful supercomputers
- **Technologies Used:** data management tools, cloud storage, federated identity management
- **Services:** solutions for finding, sharing, storing, replicating, and computing with research data, including the B2 Service Suite (B2SAFE, B2SHARE, B2DROP, etc.)
- **Community Support:** engages with diverse research communities to provide tailored data services and support

European Open Science Cloud (EOSC) – core ideas

1. **Research data are valuable**
   - And large part is lost after initial processing
2. **Only properly annotated data are valuable (even in mid term)**
   - The value goes beyond groups that create the (primary) data
3. **The data are critical for the research reproducibility**
4. **Not only data, but the processing tools and environments must be maintained**
   - ie., the other digital artefacts

⇒ besides the HPC infrastructures, **data infrastructures are also of high importance**

European Open Science Cloud (EOSC)

What is EOSC?

- a pan-European initiative to create a virtual environment for open science
- aims to provide seamless access to research data and services
- supports European researchers and innovation through data sharing

<u>Key Goal:</u> Build a federated and interoperable ecosystem for research data and services.

- EU started with EOSC around 2016

European Open Science Cloud (EOSC) – main principles

Open Science: promotes transparency, accessibility, and collaboration

Federation: connects national and international data repositories

Interoperability: ensures seamless integration across disciplines and services

Trust and Security: provides reliable access to research data with ethical considerations

FAIR Data Principles

- **Findable:** Data should have unique identifiers and metadata.
- **Accessible:** Data should be retrievable via standard communication protocols.
- **Interoperable:** Data should use standard formats and vocabularies.
- **Reusable:** Data should have clear licensing and provenance information.

European Open Science Cloud (EOSC) – FAIR data

Why FAIR data?

- **F**indable, **A**ccessible, **I**nteroperable, **R**eusable

Benefits:

- You don’t loose your data (Findable)
- You will know where they are and how to get them (Accessible)
- You have your data properly described/annotated (Interoperable)
- You will be able to use them again (Reusable)

Additional benefits

- You can share such data without (much) work on your side
- You can (easily) combine such data with data from other sources
- And you can publish your data

European Open Science Cloud (EOSC) – basic components

Metadata Repositories: store and manage metadata for research data

- metadata? key-value information associated with the data files

Data Storage & Management: secure and scalable storage solutions for research data

- place for storing the data

Computational Services: cloud computing and HPC resources

- close interconnection with HPC infrastructures

Identity and Access Management (IAM): ensures secure access control for researchers

- AAI – authentication, authorization, identities

Collaboration Platforms: virtual research environments and shared workspaces

- place to analyze the data (and place to cooperate)

European Open Science Cloud (EOSC) – summary

- EOSC aims to transform European research through open science and data sharing
- FAIR principles ensure accessibility and reuse of research data
- cooperation with HPC infrastructures enhances computational capabilities
- EOSC is considered as a key driver for the future of scientific innovation in Europe

Czech EOSC node: https://www.eosc.cz/

- designed, managed and operated by the e-INFRA CZ
  - close cooperation of CERIT-SC and CESNET


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
bedankt
dziękuje
go raibh maith agat
sukriya
kop khun krap
terima kasih
mochchakkeram
tapadh leat
спасибо
dank je
arigatô
dakujem
grazie
takk
taiku
sagolun
mauruuru
hvala
рахмат
Баярлалаа
faafetai lava
vinaka
спасибі
blagodaram
mersi
kia ora
barka
welalin
tack
spas
misaotra
matondo
paldies
grazzi
mahalo
хвала
asante
manana
obrigada
tenki
murakoze
chokrane
mamnun
дякую
djiere dieuf
tau
kiitos
dankie
dhanyavad
nanni
nandri
bayarlalaa
gràçie
köszönöm
enkosi
akun
dankon
ačiū
chnorakaloutioun
gratias ago
gràcies
sulpáy
sobodi
děkuji
mèsi
didi madloba
najis tuke
kam sah hamnida
rahmat
ありがとう
tanemirt
rahmet
diolch
dhanyavadagalu
shukriya
mercé
мерси
trugarez
তোমাকে ধন্যবাদ
감사합니다
xièxie
ευχαριστώ

e-INFRA CZ