

National Czech
Programme



EOSC CZ and FAIR Data Implementation in Czechia

Luděk Matyska

EOSC Tripartite event
Prague, 1 June 2023



Spolufinancováno
Evropskou unií



Content

- A little bit from the history
- EOSC-CZ principles introduction
- NDI and its components
- Funding and governance
- Reflections and summary

A Brief History of Open Science in Czechia

- Long tradition in the Open Access support
 - Policies at the national and institutional (university) levels
 - Transformation activities led by National Library of Technology
- Rather limited activities towards other pillars of the Open Science
- Very fragmented research data landscape
 - Open Data policy for “public” data, but not for research data
- Limited to no knowledge of FAIR principles
 - Mostly limited to infrastructures with international outreach (ESFRI nodes)
- Insufficient and fragmented data storage capacity

EOSC CZ history

- Since 2021 EOSC activities closely followed by members of the e-INFRA CZ consortium plus few large research infrastructures (LINDAT/CLARIAH and ELIXIR CZ)
- MEYS coordinating country-wide EOSC “awakening” in 2021
- EOSC Forum, to steer the initial phase
- **Architecture of the EOSC Implementation in the Czech Republic**

EOSC CZ basic principles

- Primary focus on the data
 - Storage capacity, fragmentation and hidden data, unclear access conditions, ...
- Support FAIR principles to dealing with data
 - Raise awareness
 - Provide support for FAIRification
- Make FAIR dealing with data a common practice of researchers' work
 - Provide environment
 - Provide tools and services
 - Encourage, teach and train people

Czech National Data Infrastructure

- Decision to fulfil the expectation through the concept of **National Data Infrastructure**
- Data centric infrastructure with four key components
 - **National Metadata Directory (NMD)**
 - **National Repository Platform (NRP)**
 - **Thematic repositories**
 - **Education and training (human resources)**
- Store as complete as possible/sensible data, not just appendices to publications

Czech National Data Infrastructure

- Complex systems with the following aims:
 - **Consolidated and sufficient storage capacity** to store research data in CZ
 - Open set of **key and auxiliary services** to support data repositories' management and data manipulation
 - Extensive **access control** mechanisms
 - Connected to **processing and analytical resources** (HPC, clouds, ...)
- Human resources
 - Training, education of (future) scientists
- Existing (thematic) data repositories
 - Mostly part of the infrastructures, institutions rather in planning phase

More background

- EOSC CZ Working groups established since Autumn 2021
 - Open platforms, fluid membership
 - Discussion of basic principles and aspects of the EOSC implementation in CZ
 - What is need by the researchers and research communities
 - How to best fulfil their expectations
 - But also to help “unfragment” the national landscape
 - Communities’ building
 - Combine expertise
- 4 core (cross-cutting) working groups
- 7+1 thematic (discipline) working groups

EOSC CZ Core WGs

- Metadata – Petra Černohlávková (NTK)
- Architecture – David Antoš (CESNET)
- Core Services – Michal Růžička (MU)
- Education and Training – Radka Římanová (UK)
- Cross-disciplinary, defining the foundation of the EOSC CZ implementation

EOSC CZ Thematic WGs

- **Bio/Health/Food** – Jiří Vondrášek (ELIXIR CZ)
- **ENVRI** – Jana Klánová (RECETOX a EIRENE)
- **Physics** – Jiří Chudoba (CERN)
- ***Material Sciences and Technology*** – Marek Cebecauer (ÚFCHJH)
- ***AI and Digital Science*** – Jan Šivic (Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague)
- **Humanities and Arts** – Jan Hajič (LINDAT/CLARIAH-CZ)
- **Social Sciences** – Jindřich Krejčí (ČSDA)

and

- **Sensitive Data** – Adam Svobodník, Věra Franková (BBMRI CZ, CZECRIN, EATRIS CZ)

Components of NDI

- **National Metadata Directory (NMD)**
 - Key component, guaranteeing **Findability and Interoperability**
 - Interacting directly with all other NDI components
 - Mandatory use
 - Physically distributed (redundancy), centrally managed
- **National Repository Platform (NRP)**
 - NDI backbone, providing capacity and basic/common services
 - **Accessibility and Reusability**
 - Defines standards and interfaces for the whole NDI
 - Physically distributed, managed by a national (larger) consortium
 - Coordinated by EOSC Association Mandated Organization

National Metadata Directory

- Specific repository, index to all the data in NDI
 - In principle a metadata database
- High availability and throughput
 - Accessible 24/7, no interruptions
 - Metadata can't be lost
 - Potential large number of simultaneous accesses (reads, but also writes)
 - Support to automatic upload of metadata (e.g.. Instruments)
- Distributed architecture
 - 5 instances (3-5 cities)
- Access control support even for metadata access

National Repository Platform

- Robust platform to support creation/instantiation of repositories
 - Physically distributed nodes, data stored in synchronized object storage
 - Provides high availability, data replication, data movement, ...
- Repositories as tenants on this platform
 - Usually thematic or institutional (not recommended)
 - Conditions/policies for creating new repository
 - Mandatory conditions guaranteeing long term sustainability
 - Access control at the level of whole repositories, metadata, datasets, ...
 - Direct link to NMD and the catalogue of repositories
 - Metadata export

Components of NDI

- **Existing thematic repositories**

- Continuity, support for long-term sustainability, ...
- Interconnection with NRP and NMD
 - Interoperability at the technical level, general findability
- Also support to communities with new/cross disciplinary needs
 - Either new physical repositories or instances in the NRP

- **Human Resources**

- Horizontal activity
- Broad ambitions, from training to (university) education
- New positions and carrier paths

Access Control

- Extremely important aspect
- Data (and even most of the metadata) in NDI are under direct and full control by their owners
 - Access is granted to fulfil the conditions sets by them
- Even the information about the existence of specific data (sets) may be sensitive
 - Access control must accommodate fine grained access to specific parts of metadata and data sets (not only the whole (large) datasets)
- Federated AAI support

Funding

- Series of several dedicated call of OP JAC (structural funds)
 - Plus support from other call with broader context
- Mandatory requirements in research projects calls
 - DMPs
- 4 dedicated calls between 2022 and 2025
 - Special allocation for 2022 call (covers more than just EOSC)
 - Dedicated allocation of 2,5 billion Czech Crowns (approx. 100 Million Euro) for calls since 2023

Funding – calls

- **Background of the implementation**

- First specific call in 2022
- Two projects: **IPs CARDS** and **IPs EOSC-CZ**
 - Metadata, PIDs, ...
 - EOSC CZ Secretariat, Virtual Education and Training Centre
 - NMD implementation

- **National Repository Platform**

- Call open (2023), one project expected
- Consortium coordinated by EOSC Association Mandated organization (CESNET)
- Currently under preparation

Funding – calls

- **Thematic and other repositories**

- Call under preparation, expected opening early 2024
- Expected several projects (consortia)
- Support for consolidating the thematic repositories space
- Mandatory connection to NMD, technical interoperability (through APIs) with NRP

- **Uptake and Human resources**

- Call expected in 2025
- To support uptake of EOSC services and capacity
- Broader education activities

- Other calls, e.g. Research Environments expected in 2024

Aim of the projects

- The projects aim to build the NDI and its components either from scratch (e.g. NMD, NRP) or through connecting existing repositories
 - By-product is cross-country experts' community building and support
- The results will be available for the whole national research community
 - NDI as a background for Open Science implementation
 - No one should be excluded
- Sustainability aspects also part of the outputs
 - A dedicated deliverable of IPs EOSC-CZ project due in 2025

Governance

- EOSC Coordination Board for the EOSC implementation
 - Minister of MEYS Advisory Board
 - Stakeholders coordination
 - Discussion and consultation platform
 - Accepts recommendations, EOSC-CZ standards, EOSC related funding call, ...
 - Contributes to the evolution of EOSC-CZ architecture and NDI
- Provides feedback to the EOSC related projects
 - Primary interaction through the **IPs EOSC-CZ** project – **OP JAC Strategic project**
 - Role of EOSC-CZ Secretariat

Expected outputs

- NDI – Easily usable complex environment to support FAIR work with FAIR data
 - Complex but easy to use Access control environment (EOSC-CZ AAI)
 - NRP – Robust infrastructure to create and manage repositories to store and access data compliant with FAIR principles
 - Life data included
 - Thematic repositories – connection to NMD, NRP, connection to EOSC-CZ AAI, ...
 - Tools like actionable DMPs, DSW and FAIRification support, ...
- NMD and Catalogue of repositories – country-wide entry point, data findability
- Researchers' education and support

Why we are doing it?

- **Support Open Science initiative** as the basic of strategic development of ERA
 - And FAIR data a must for an excellent science with wide societal impacts
- **EOSC as a critical pillar for Open Science** implementation
 - FAIR data, European partnership, ...
- Using this opportunity to **advance Czech research quality/excellence**
 - Support to properly (FAIR) dealing with research data (and other digital objects)
 - Ease data sharing
- **Country-wide collaboration** in all steps, from preparation through building to use

Reflections on some of the Ten things EOSC is not ...

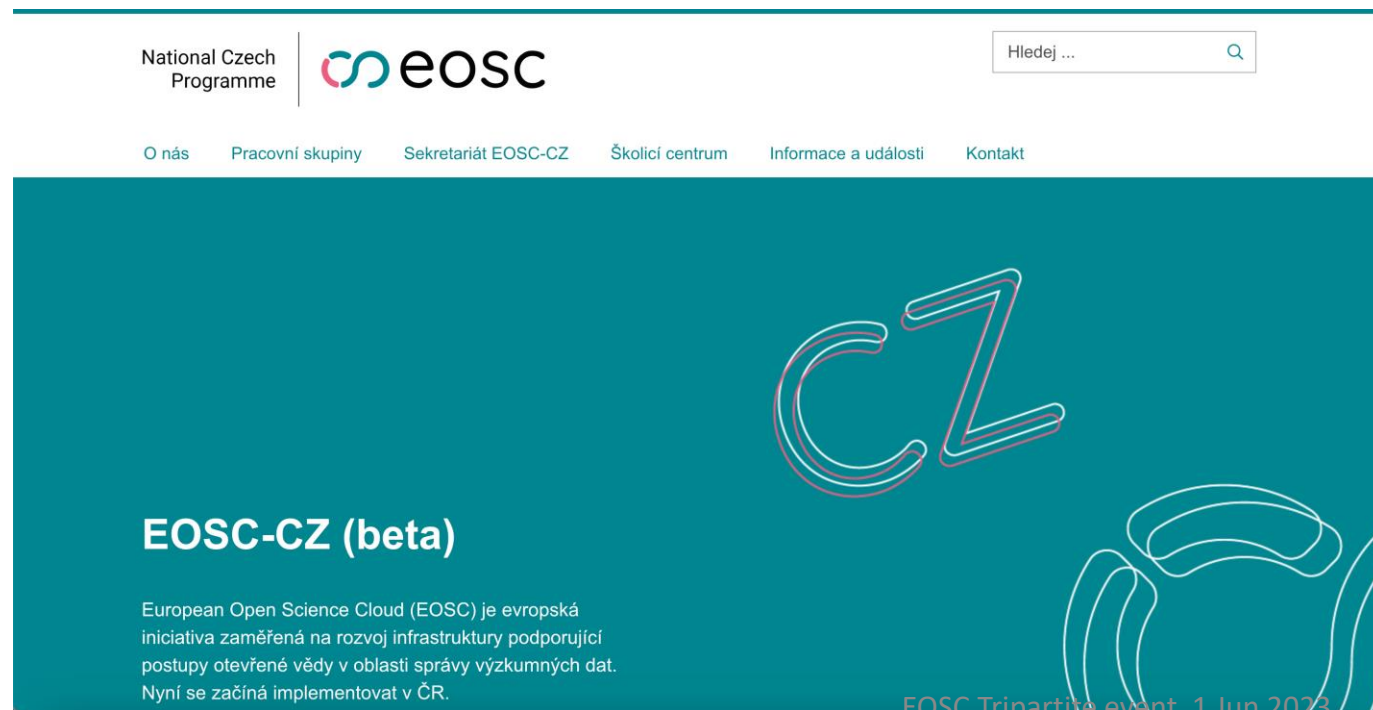
- **a cloud infrastructure**
 - NDI is a data oriented cloud infrastructure to setup and operate repositories
- **substitution of any existing data- or (e-)infrastructures**
 - NDI encompasses existing data infrastructures/repositories but also provides platform for complementary ones through the NRP
- **the EOSC Portal**
 - Not an EOSC Portal, but NMD and Repository catalogue are portal to access data in NDI
- **directly for individual researchers, but it is for research**
 - NRP a platform to create repositories by research communities/institutions/... that are directly used by researchers
 - Tools and services like EOSC-CZ AAI, actionable DMP or supported DSW directly used by researchers

Summary/Conclusion

- EOSC CZ implementation seen as a unique opportunity to improve Czech research landscape
- Move towards data-centric focus
- Huge potential, but also extreme challenge
 - The collaborative aspect is the must for a success
- Ambitious plans
 - Unfragment the research data landscape
 - **“FAIR data in every lab”**
- **Make better science through the better FAIR data management**

WWW.EOSC.CZ

National Portal for the EOSC-CZ Implementation



National Czech
Programme



Thank you for your attention

Contacts and additional info:

info@eosc.cz

www.eosc.cz



**Spolufinancováno
Evropskou unií**

EOSC Tripartite event, 1 Jun 2023


MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



MUNI
CERIT-SC

cesnet
■■■■

VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA

IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ
CENTRUM