National Czech Programme | eosc

EOSC – where are we in Czechia

Ludek Matyska

CESNET and Masaryk University

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

MUNI ICS

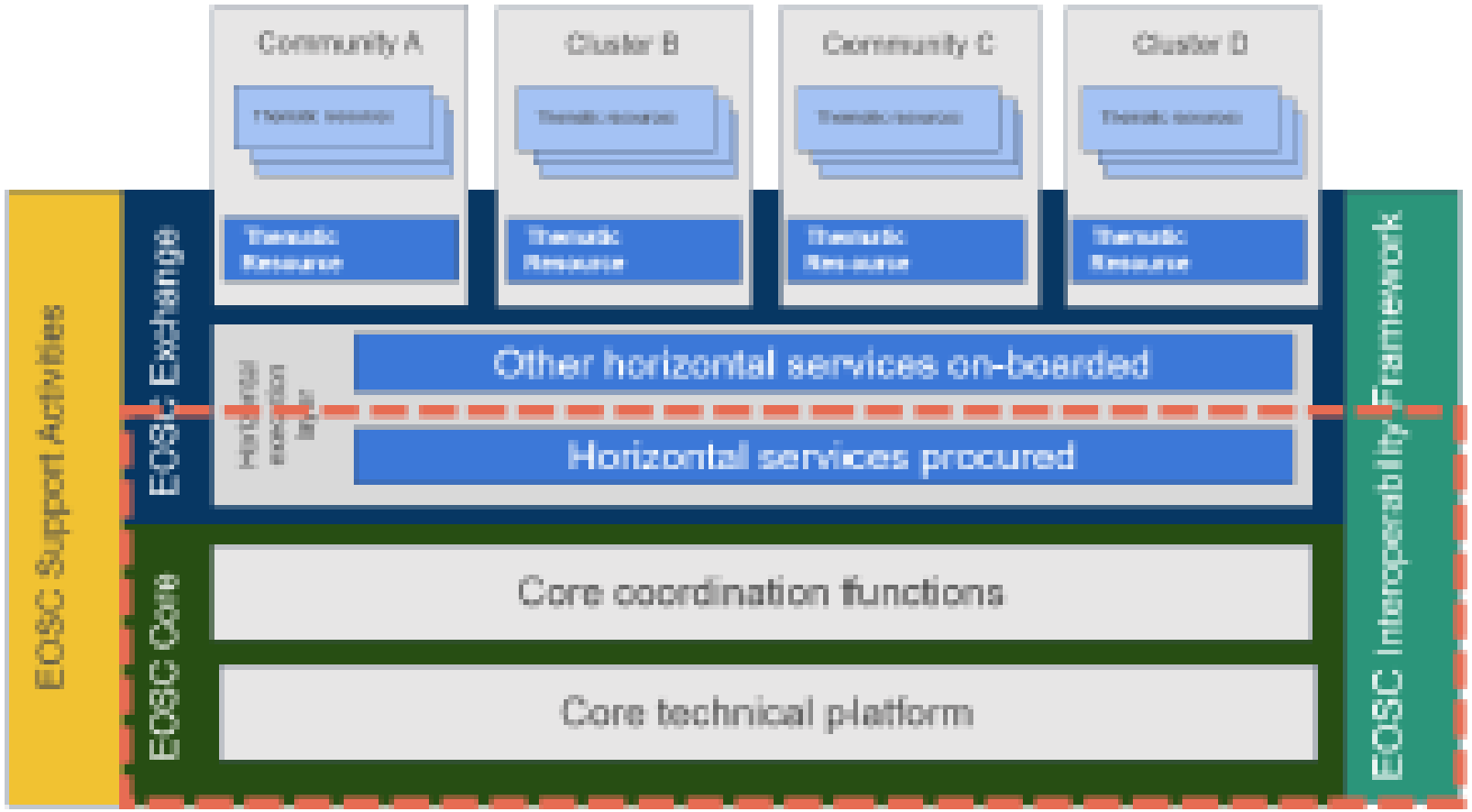
What is EOSC

- **Web of FAIR data and related services**
- Nice, but what does it mean?
- *A complex ecosystem providing seamless access to interoperable research objects and value-added services for the whole research data cycle, from discovery and mining to storage, management, analysis and re-use across borders and scientific disciplines*
- It supports scientists to properly manage their FAIR data and to find and reuse data produced by others in an orderly way; supporting interoperability and collaboration

EOSC – A little bit of international background

- Almost a decade of EC support to build the EOSC ecosystem
- The Tripartite
  - European Commission
  - EOSC Steering Board
  - EOSC Association
- EOSC Procurement
- EOSC Nodes concept

EOSC – Public Procurement Action 2022



EOSC Procurement

- To build and deploy a fully operational enabling infrastructure for EOSC
- LOT1 – Core Federation Services for EOSC EU Node
  - *Open Science Agora Consortium*
- LOT2 – Exchange Infrastructure Services for the EOSC EU Node
  - Poznan Supercomputing and Networking Center (PSNC)
- LOT3 – Exchange Application Services for the EOSC EU Node
  - *Poznan Supercomputing and Networking Center (PSNC)*

EOSC Procurement

- To build and deploy a fully operational enabling infrastructure for EOSC
- LOT1 – Core Federation Services for EOSC EU Node
  - Web Portal Front Office, Resource Catalogues and Registry Services, AAI, Application Workflow Management, Monitoring and Accounting function
- LOT2 – Exchange Infrastructure Services for the EOSC EU Node
  - Managed Container service, Managed Compute (Virtual Machine) service, Managed Bulk Data Transfer service
- LOT3 – Exchange Application Services for the EOSC EU Node
  - Managed File Synchronisation and Sharing service, Interactive Notebooks service, and Managed Large File Transfer service for end-users

EOSC Node Concept

- EOSC as a federation of EOSC Nodes
  - In fact admitting EOSC is a special case of an infrastructure
- EOSC Node is a distributed structure that provides data and/or services to EOSC users as well as to other EOSC Nodes
- Nodes need to conform to the requirements and governance rules of the EOSC Federation as set by the tripartite governance
- Ongoing discussion about the exact role, position and implementation of EOSC Node
  - To provide a structure, governance, and systematic funding for the EOSC Ecosystem

EOSC EU Node

- Prime inter pares
- The core of the EOSC Federation
  - EOSC Nodes expected to connect to it
- Web already setup, a reference node for EOSC
  - Operated by European Commission
- Takes care of the procurement results
  - Individual services from LOTS 1 to 3 will be available through it

European Open Science Cloud - EU Node

(https://open-science-cloud.ec.Europa.eu)

EOSC in Czechia – Preparation and Background

- ***Architecture of EOSC implementation in Czechia***
  - Document accepted in 2021 by Ministry of Education, Youth and Sports
- **Coordination Board** at the Ministry since December 2022
- Structural Funds support
  - Open Science/EOSC CZ series of calls
  - Almost 3 billions Czech Crowns (120 M€) for the period 2023-2028
- Special role of e-INFRA CZ
  - Partner for Ministry to discuss the architecture and support
  - Implementation responsibility
    - Not alone, but the key player

EOSC CZ

- Deliberately narrow focus (compared to the EU position)
- **Primary target**: Data produced by scientists
- **Goal**: Make sure that at least 80% of all data producing scientists knows where to store them permanently

EOSC CZ

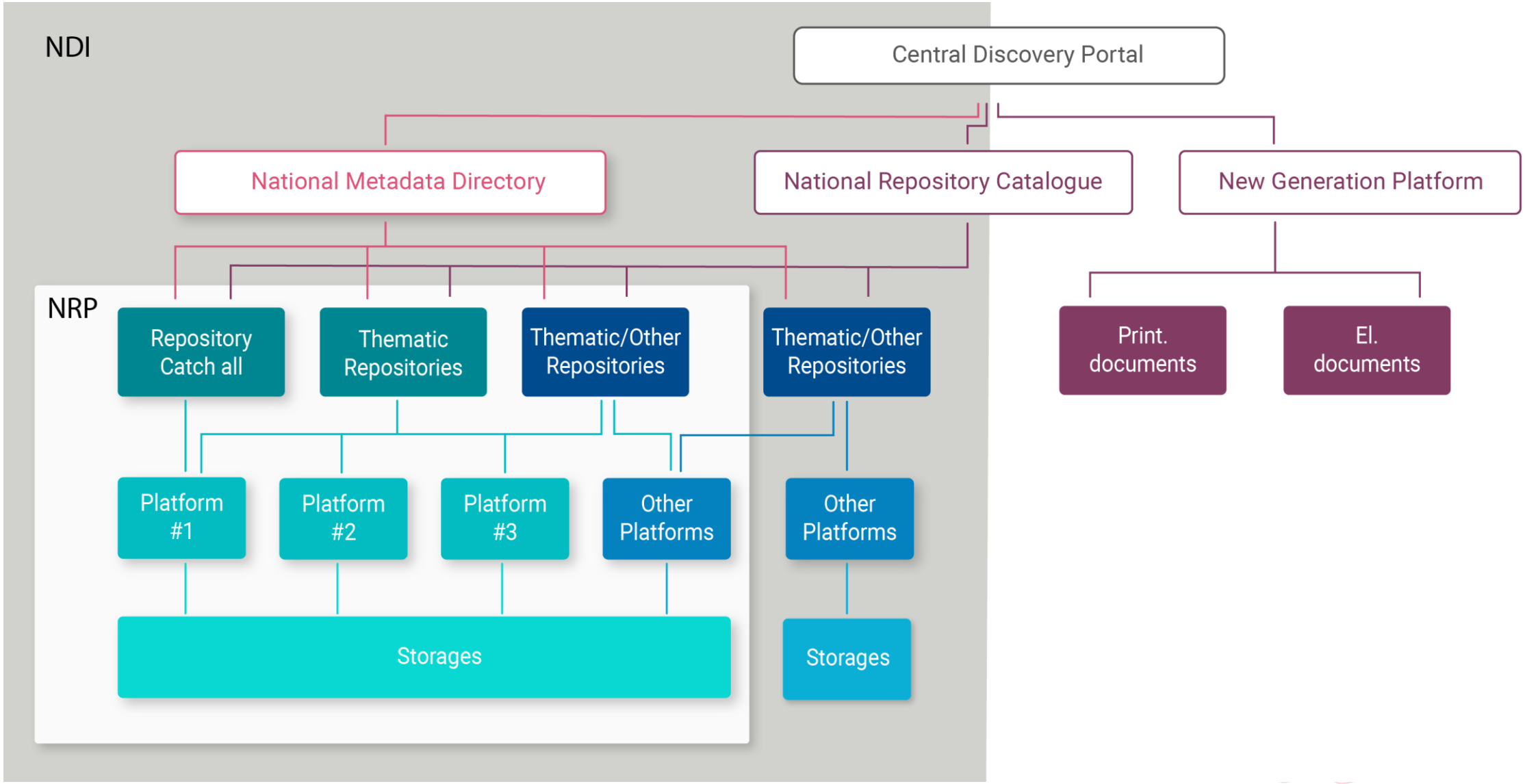
- Deliberately narrow focus (compared to the EU position)
- **Primary target**: Data produced by scientists
- **Goal**: Make sure that at least 80% of all data producing scientists knows where to store them permanently
- **Secondary target**: Scientists looking for data
- **Goal**: All scientists know where to look for the data they are interested in

EOSC CZ

- Deliberately narrow focus (compared to the EU position)
- **Primary target**: Data produced by scientists
- **Goal**: Make sure that at least 80% of all data producing scientists knows where to store them permanently
- **Secondary target**: Scientists looking for data
- **Goal**: All scientists know where to look for the data they are interested in
- **Corollary**: All the data we speak about is FAIR

EOSC CZ Architecture

- **National Data Infrastructure** – A system composed of
- **National Metadata Directory (NMA)**
- **Repository Platform**
  - With several repository systems like Invenio or DSpace
- **Services and tools**
- **Repositories**
  - Esp. those built on top of the repository platform
- **Training, Methodological Support**
- All this as a part of a larger system covering also publications


NDI
Central Discovery Portal
National Metadata Directory
National Repository Catalogue
New Generation Platform
NRP
Repository Catch all
Thematic Repositories
Thematic/Other Repositories
Thematic/Other Repositories
Print. documents
El. documents
Platform #1
Platform #2
Platform #3
Other Platforms
Other Platforms
Storages
Storages

A Little Bit More Concrete Information

- **EOSC CZ Secretariat**
  - Established in 2023
  - A high-level support body (e.g. working groups secretaries)
  - Overseeing and Monitoring EOSC CZ implementation progress
    - Strong collaboration with Ministry
  - Analytical section
- **National Metadata Directory**
  - An Invenio-based high level metadata repository, harvesting metadata
  - High availability, high throughput (geo-distributed, at least two sites)
  - A general Metadata model to speed up data discovery
  - Pilot version just launched, production Q1/Q2 2025

National Repository Platform

- Raw storage capacity, object storage and repository systems
  - The place where data is stored
  - Several layers – from hardware through storage, repository software to repositories and their interfaces
- **Hardware**
  - 4-5 geo-distributed nodes (Prague, Brno, Olomouc, Ostrava)
  - First two nodes operational in Q2/2025
- **Object storage Ceph** (S3)
  - Replication, High availability
- Three repository systems – **Invenio, CLARIN DSpace, ASEP/ARL**
  - To support different needs

Common Services

- **AAI and strong access control support**
  - A uniform AAI system, Mandatory for all services to connect, but not a monopoly
  - Based on the AAI blueprint, fully compatible and compliant with EOSC AAI
- **Metadata profiles management**
- **Licensing support**
- **Data and Repositories FAIRification**
- **Data management plans** support (DSW, **actionable** DMPs)
- **Automatic data collection** (directly from instruments)
- **Integration with compute and analytical infrastructure, workflows**
- **Bulk data transfer**

Cybersecurity, compliance, and user experience

- **Cybersecurity support and overseeing**
  - Integrated with e-INFRA CZ cybersecurity support
  - Intgerated with e-INFRA CZ helpdesk and primary user support
- **Compliance**
  - Legal compliance
  - Work with sensitive data
- **User experience**
  - Special support to monitor and improve user experience accross all the services provided b the EOSC CZ implementation

Training

- Supporting scientists in uptake of the ecosystem and its services
- Documentation
- Methodological support
- Training materials
- Training support (virtual training center)
- Data Stewards’ community

EOSC is about collaboration

- **Thematic Working Groups**
  - Bio/Health/Food
  - Social Sciences
  - Humanities and the Arts
  - Physics
  - Environmental Sciences
  - Material Sciences and technology
  - *Data management for AI and ML*
  - *Sensitive data*
- Strong connection to the large research infrastructures

Role of Thematic Working Group

- To involve researchers into the EOSC definition and implementation
- To collect feedback on plans and realization
- To discuss strong and weak points
- To assure that the support goes to the right places
- To prepare input for the targeted support of FAIR data handling
  - Repositories
  - Interoperability

Thematic Clusters Oriented Call

- Finalized (to be launched in May/June)
- Targeting research communities
- Repository setup and evolution
  - Using NRP infrastructures where appropriate
  - Using NRP services (e.g. mandatory NMA, Catalogue Registry, AAI)
  - FAIRification, visibility, …
- FAIR data management
  - FAIRification, Interoperability, …
- Core partners and miniprojects
  - To give higher flexibility to the implementation

Who Builds What

- EOSC-CZ and CARDS project (2023-2028)
  - NMA including its Metadata model(s)
  - Part of the core services (AAI, data transfer)
  - PIDs
- NRP for research data project (2024-2028)
  - The whole NRP and its core services
  - Three supported platforms (and pilot repositories)
  - Documentation and training materials
- Clusters' project (2025-2028, expected)
  - The repositories and their specific services
  - Also metadata, interoperability, standards, …

How to orient yourself

- If you want to build and take care of a repository
  - the NRP project is your partner
- If you are a scientist looking for place to store your own data
  - The EOSC-CZ and future Clusters' projects are your partners
- If you are a scientist looking for relevant data
  - The EOSC-CZ (and to some extent the Clusters') project is your partner
  - However, the primary partner should be the relevant research community
    - E.g. the relevant large national or better international infrastructure
- If you are just interested
  - The EOSC-CZ is your partner

Schedule

- 2023
  - Setting the scene, EOSC CZ Secretariat, preparatory work, …
- 2024
  - National Metadata Directory operational
  - Early pilots
  - Cybersecurity and compliance setup
  - First large-scale survey of the scientific communities
  - First phase of the preparation of the thematic clusters project
- 2025
  - Fully operational NRP (just reduced scope)
  - First repositories within NRP implemented and starting to receive data
  - Submission and eventual start of the thematic clusters project
  - Institutional support through a separate funding (Research Environments call)

Schedule

- 2026
  - NRP fully operational
  - Training
  - Wide discussion about the sustainability
  - Repositories and FAIR data collections supported through thematic clusters project
  - Additional support from Ministry (last directly related call)
- 2027-2028
  - Continued NRP, extending number of supported repositories and data collections
  - Final survey of the scientific communities (progress evaluation)
  - Training in full scope, education at wider scale

Summary – Where we are with EOSC in CZ

- Strong support
  - Both at the ministerial and scientific communities' levels
- Strong international collaboration
  - Participation in two EOSC Procurement LOTS (and HEU projects)
  - Active EOSC Association Mandated organization
  - Close collaboration with EOSC SB representatives
- Strong focus on management of FAIR data
  - All aspects of the FAIR data life covered
- Strong interaction and collaboration with e-INFRA CZ
  - Wider extension of operations and responsibilities
- Clear architecture and implementation schedule

National Czech Programme | eosc

Thank you
Any questions?

www.eosc.cz, info@eosc.cz

See also https://arxiv.org/pdf/2402.13343

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

MUNI ICS

cesnet

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER