EOSC Implementation in Czechia

Luděk Matyska

e-INFRA CZ

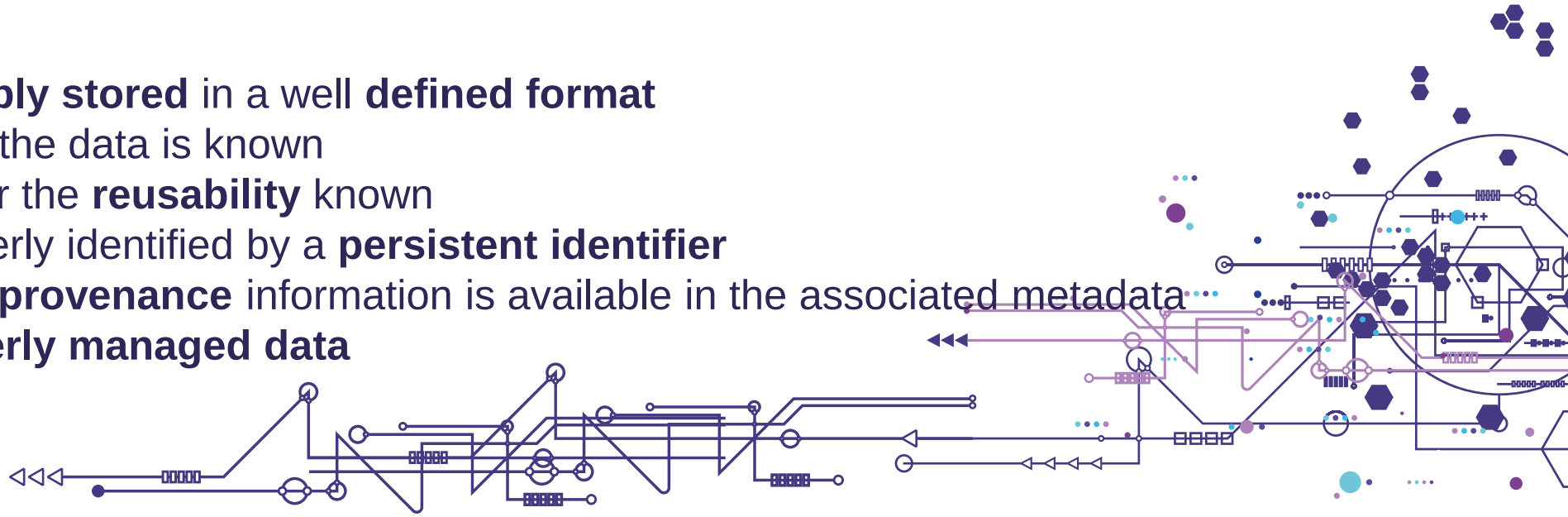
cesnet

MUNI CERIT-SC

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | IT4INNOVATIONS NÁRODNÍ SUPERPOČÍTAČOVÉ CENTRUM

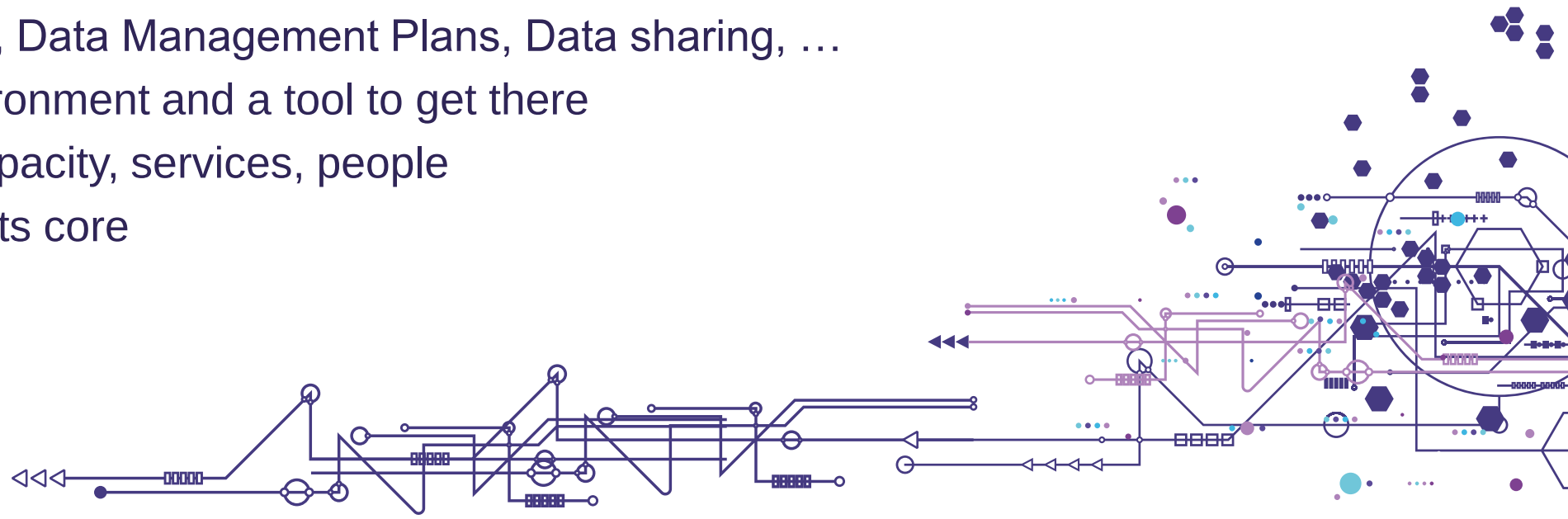
Data and Open Science

- Data in the widest possible interpretation
  - Publications, Research Data, Software, Workflows, …
  - No silos, barriers, …
- Open Data too strong principle → FAIR Data Principle
  - **F**indable
  - **A**ccessible
  - **I**nteroperable
  - **R**eusable
- Meaning
  - Data are **reliably stored** in a well **defined format**
  - **Semantics** of the data is known
  - **Conditions** for the **reusability** known
  - Data are properly identified by a **persistent identifier**
  - Some **quality/provenance** information is available in the associated metadata
- **FAIR data = properly managed data**



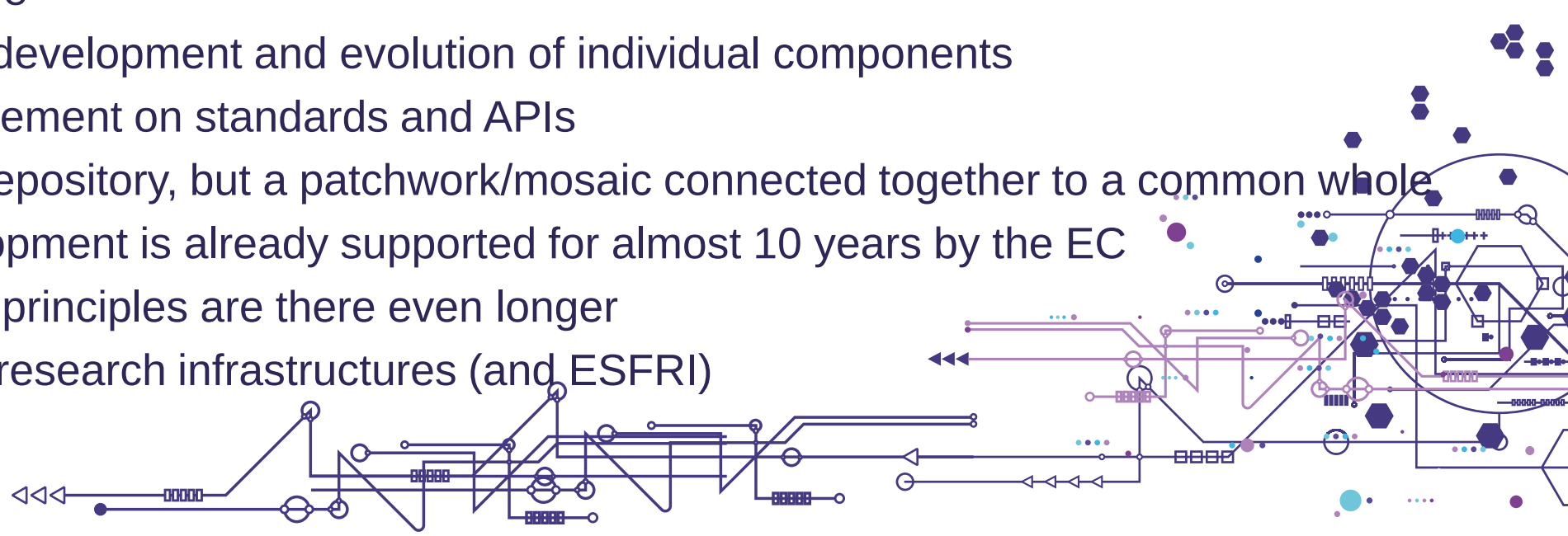
European Open Science Cloud (EOSC)

- One of the pillars of the **Open Science**
  - Together with **FAIR Data principles**
- European Research Data Area and Ecosystem
  - European Research Data Commons
  - Storing, processing/analysing, re-using research data Under FAIR principles
- Funding agencies pushing towards openness
  - Open Access, Data Management Plans, Data sharing, …
- EOSC as an environment and a tool to get there
  - Principles, capacity, services, people
  - FAIR data in its core



European Open Science Cloud (EOSC)

- Hardware, software, services, processes supporting the managed work with data
  - **Capacity**
  - **Data management**
  - **Access control**
  - Access to the processing/analytics tools and environments
- Federated principle
  - Independent development and evolution of individual components
  - Need an agreement on standards and APIs
- Not one storage/repository, but a patchwork/mosaic connected together to a common whole
- The EOSC development is already supported for almost 10 years by the EC
  - However, the principles are there even longer
  - Role of large research infrastructures (and ESFRI)

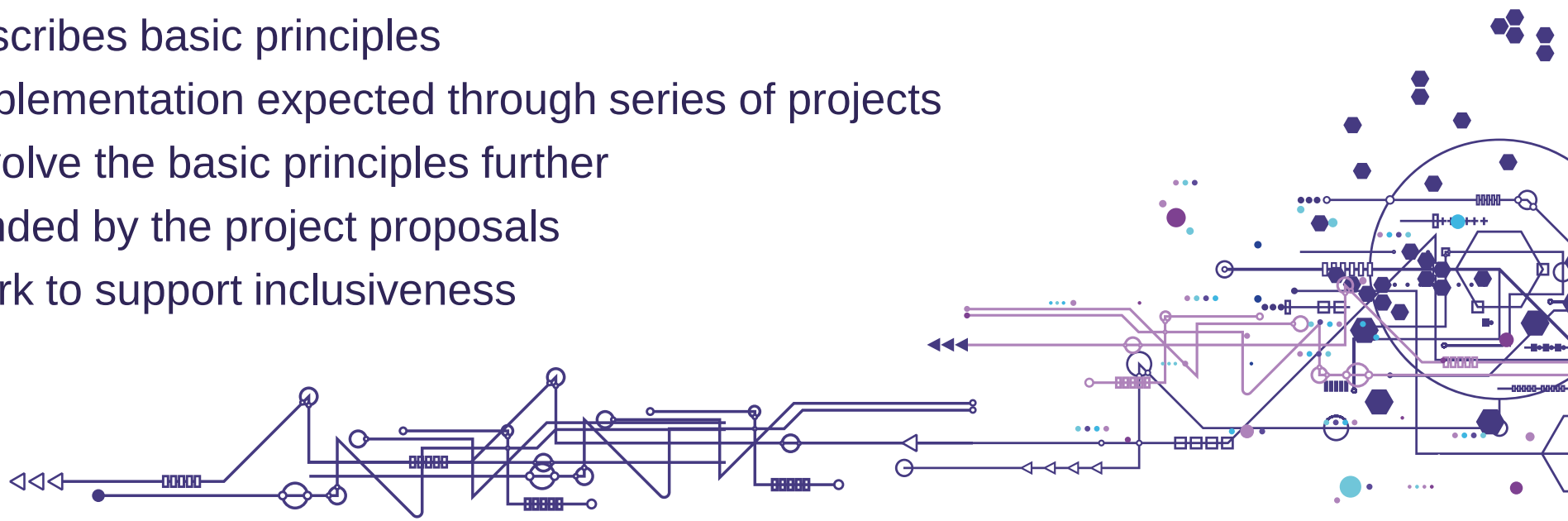


EOSC CZ Principles

- EOSC understood as a federation of FAIR data and related services
  - Federation of existing and future data resources (FAIR data repositories)
  - Open ecosystem of services
- EOSC CZ architecture discussion in 2021
  - Under the auspices of the MEYS
- **Output: Architecture of EOSC Implementation in the Czech Republic**
- Governance: **EOSC CZ Coordination Board**
  - Established 2021/2022
  - Advisory board of MEYS, chaired by two vice-ministers
- Direct financial support through OP JAK

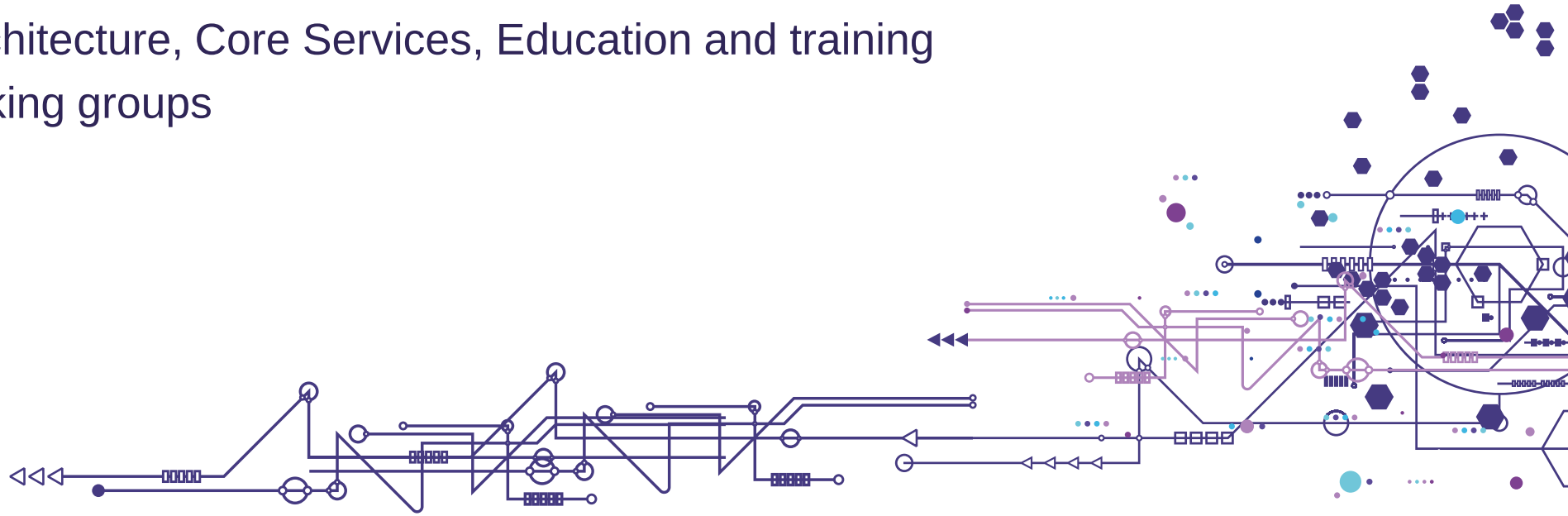
Architecture of the EOSC CZ Implementation

- **National Data Infrastructure (NDI)**
- Data-centric infrastructure with 4 key components (pillars)
  - **National Metadata Directory (NMA)**
  - **National Repository Platform (NRP)**
  - Thematic repositories
  - Education and training (human resources)
- The document describes basic principles
  - The actual implementation expected through series of projects
- Work groups to evolve the basic principles further
  - First phase ended by the project proposals
  - Continued work to support inclusiveness



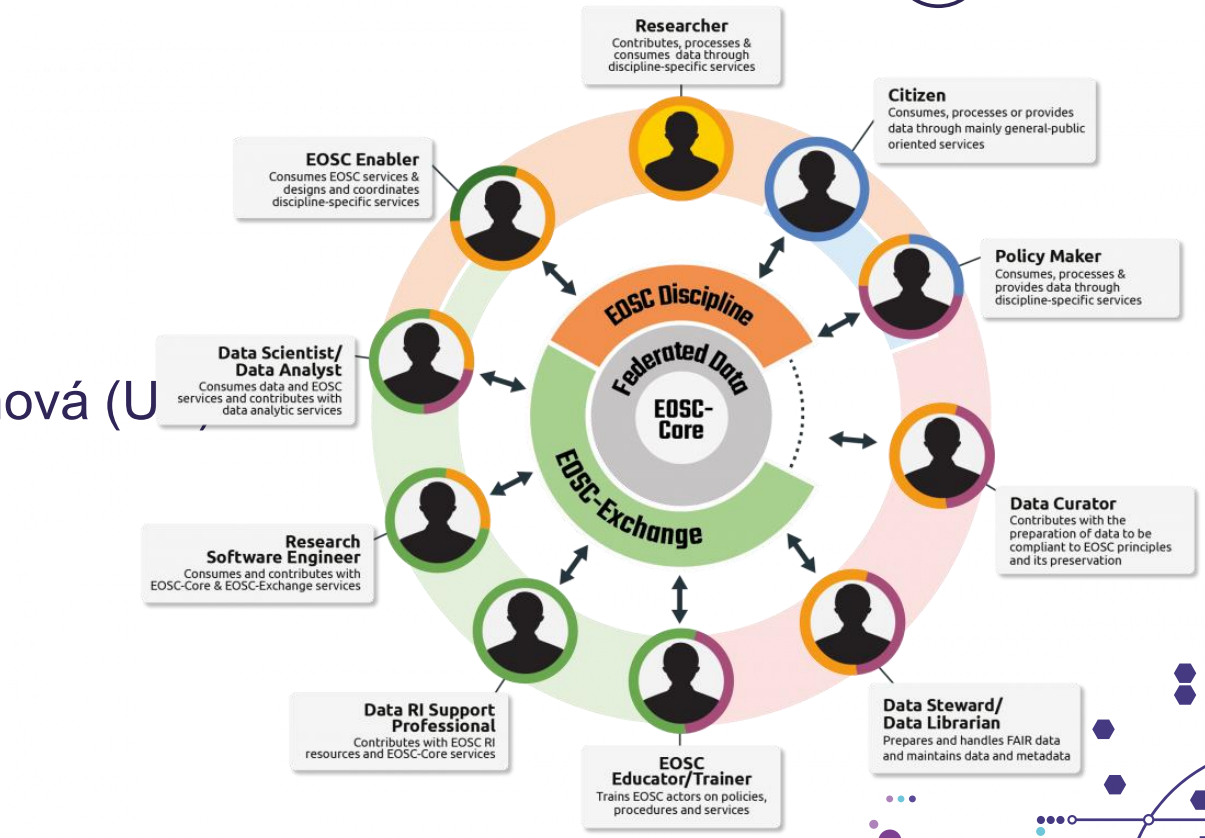
Preparatory steps

- Working groups established since Autumn 2021
  - Open platform, with fluid membership
  - Discussion of basic principles of all aspects of the EOSC CZ implementation
    - What is needed by researchers and their communities
    - And how to best fulfill these expectations
- 4 foundation working groups
  - Metadata, Architecture, Core Services, Education and training
- 7+1 thematic working groups



Foundation EOSC CZ working groups

- **Metadata** – Petra Černohlávková (NTK)
- **Architecture** – David Antoš (CESNET)
- **Core services** – Michal Růžička (MU)
- **Education and training** – Radka Římanová (U
  - Education and support for researchers
    - New competencies in the area of **data management and sharing**
  - Education for support roles
    - New carriers
    - Study programmes
    - Training



Thematic EOSC CZ Working Groups

- **Bio/Health/Food** – Jiří Vondrášek (ELIXIR CZ)
- **ENVRI** – Jana Klánová (RECETOX a EIRENE)
- **Physics** – Jiří Chudoba (CERN)
- **Material Sciences and Technology** – Marek Cebecauer
- **AI and Digital Science** – Jan Šivic
- **Humanities and Arts** – Jan Hajič (LINDAT/CLARIAH-CZ)
- **Social Sciences** – Jindřich Krejčí (ČSDA)

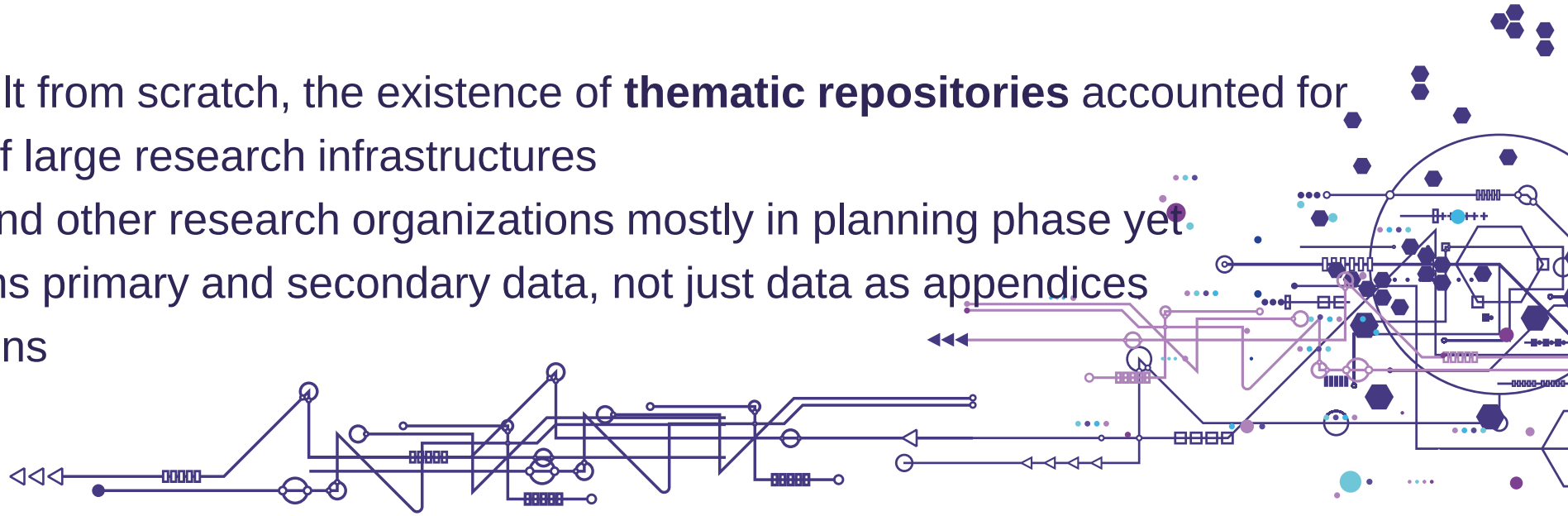
and

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

- Provides opportunity to bring in specific thematic needs

National Data Infrastructure

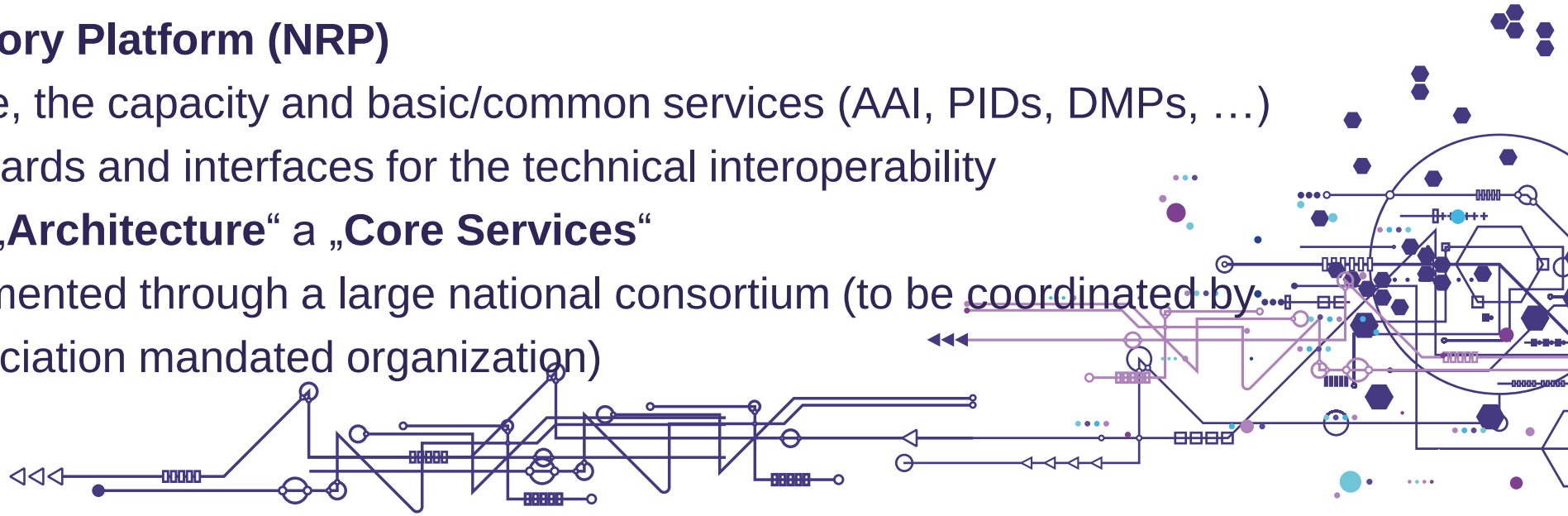
- Complex system with the following aims:
  - Basic **infrastructure** with **sufficient capacity** to store research data in Czechia
  - Open set of **key and other services** for data manipulation
  - Interaction with processing resources (HPC, clouds, …)
- With parallel support of **human resources**
  - Education of (future) scientists, new competencies (curation), data stewards, data scientists
- Not everything built from scratch, the existence of **thematic repositories** accounted for
  - Usually part of large research infrastructures
  - Universities and other research organizations mostly in planning phase yet
- Data-centric means primary and secondary data, not just data as appendices to publications



National Data Infrastructure

Key components

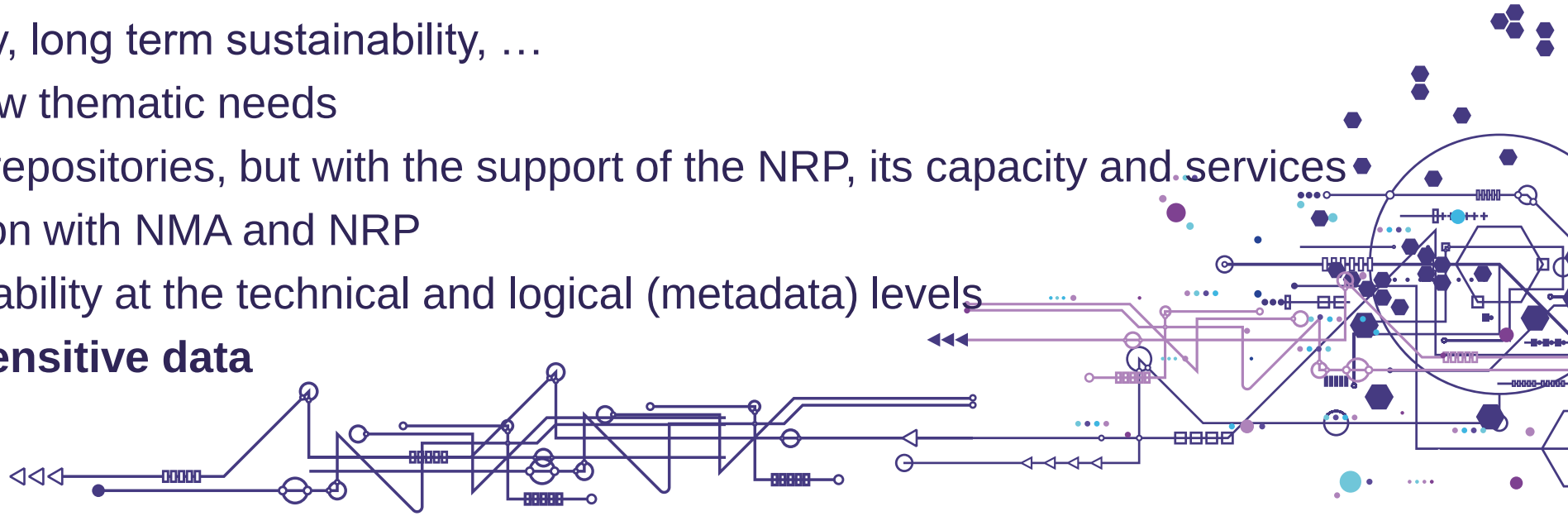
- **National Metadata Catalogue (NMA)**
  - Key component, guaranteeing findability and interoperability
  - Work group „**Metadata**“
  - Interacting directly with all other NDI components
  - Centrally managed NDI component
    - Still distributed; e-INFRA CZ responsibility
- **National Repository Platform (NRP)**
  - NDI backbone, the capacity and basic/common services (AAI, PIDs, DMPs, …)
  - Defines standards and interfaces for the technical interoperability
  - Work groups „**Architecture**“ a „**Core Services**“
  - Will be implemented through a large national consortium (to be coordinated by EOSC Association mandated organization)



National Data Infrastructure

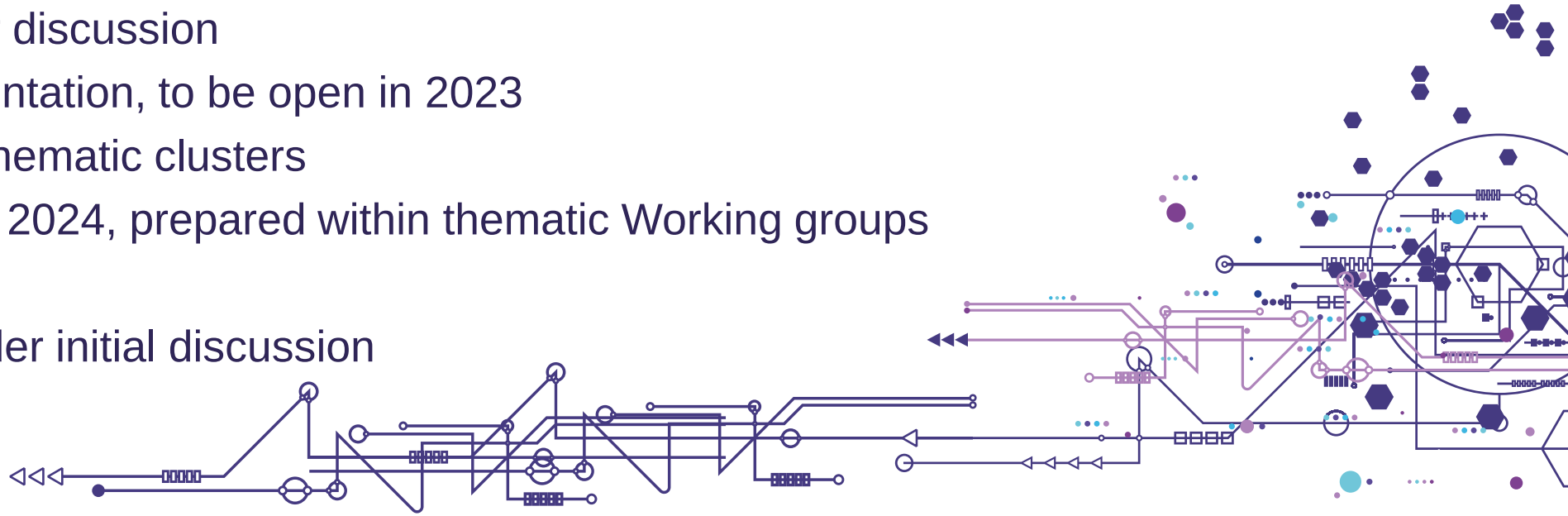
Other parts

- **Human Resources**
  - Horizontal Activity
  - Work group „**Education**“
  - To be supported through the whole implementation of NDI
- **Existing thematic repositories**
  - Support of on-going activities and repositories
    - Continuity, long term sustainability, …
  - Support of new thematic needs
    - Not new repositories, but with the support of the NRP, its capacity and services
  - Interconnection with NMA and NRP
    - Interoperability at the technical and logical (metadata) levels
- Specific care of **sensitive data**



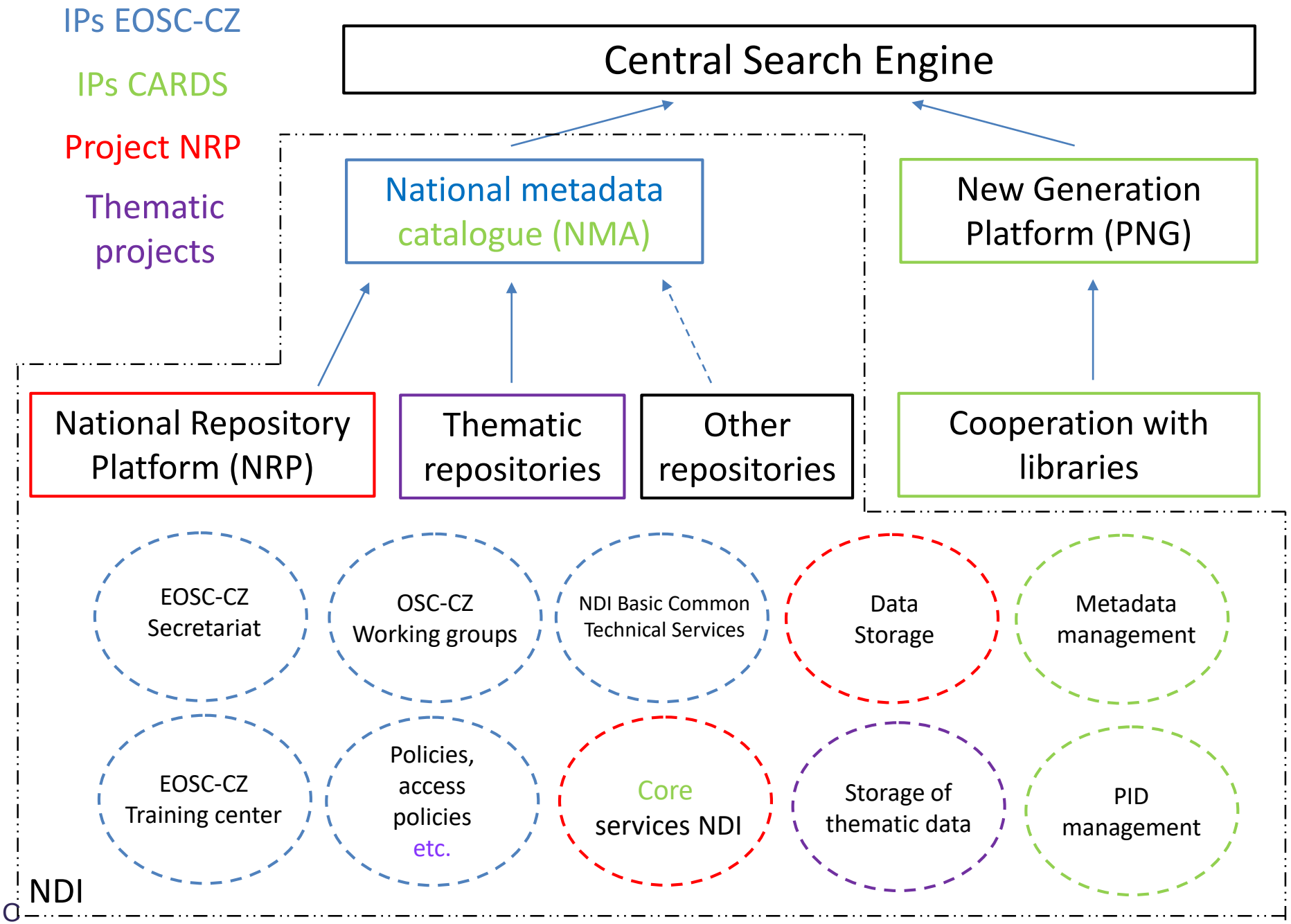
Actual Plans

- EOSC implementation to be supported through structural funds (**OP JAK**)
- Preliminary allocation of some **120 M€** (2.5 billion Czech crowns) for the 2023-2028 period
- Several calls under discussion
- First call already out
  - Projects **EOSC CZ** (EOSC Secretariat, NMA, training coordination) and **CARDS** (Metadata, PIDs)
- Second call under discussion
  - NRP implementation, to be open in 2023
- Third call for the thematic clusters
  - To be open in 2024, prepared within thematic Working groups
- Uptake/Upskilling
  - For 2025, under initial discussion



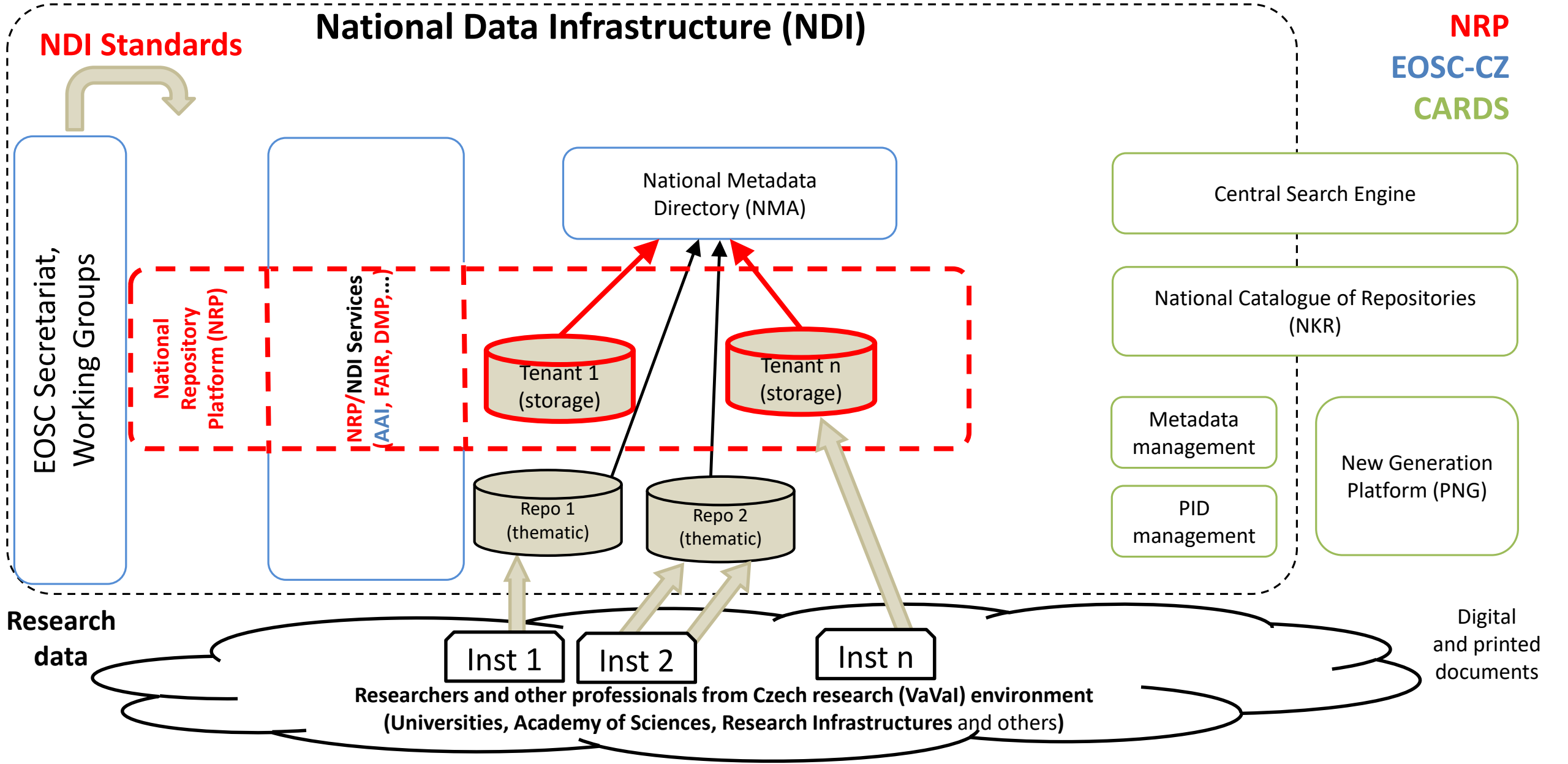
General structure of the EOSC CZ implementation

- **Coordination and general environment (background)**
  - EOSC CZ Secretariat, NMA setup and operation plus directly related services, PIDs, …
  - Information and training hub
  - Other EOSC related projects and activities monitoring
  - IPs EOSC CZ (e-INFRA CZ) and CARDS (NTK), call already open, projects almost finalized
- **National Repository Platform (NRP)**
  - Call in preparation, implementation start expected Q3 2023
  - One joint project expected, coordinated by ESOC Association mandated organization
- **Thematic and other specific repositories**
  - Call will be prepared during 2023, implementation to start Q3 2024
  - Several projects around thematic clusters, including sensitive data management
- **Uptake/Upskilling** call expected 2025


IPs EOSC-CZ
IPs CARDS
Project NRP
Thematic projects
Central Search Engine
National metadata catalogue (NMA)
New Generation Platform (PNG)
National Repository Platform (NRP)
Thematic repositories
Other repositories
Cooperation with libraries
EOSC-CZ Secretariat
OSC-CZ Working groups
NDI Basic Common Technical Services
Data Storage
Metadata management
EOSC-CZ Training center
Policies, access policies etc.
Core services NDI
Storage of thematic data
PID management
NDI

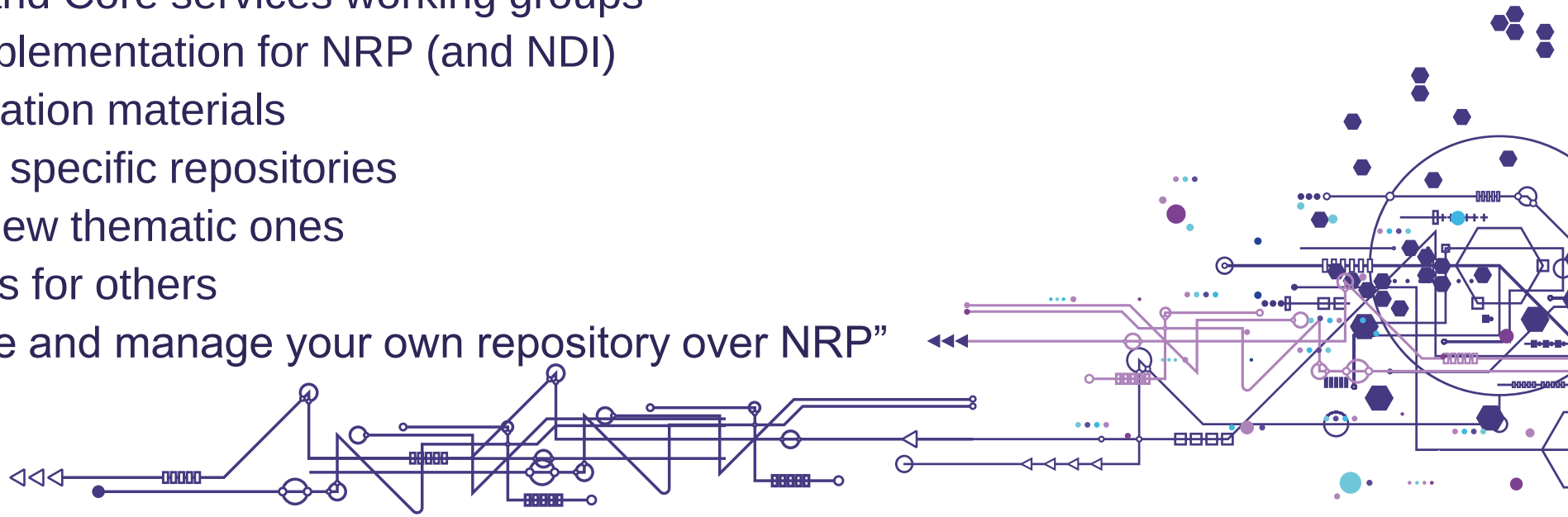
Structure of the Czech intervention

Realized via operational programme Open Science OP JAK (context of the NRP call)



NRP Project – expected outputs

- Continuously updated NDI and NRP architecture with connection to the thematic repositories
- Detailed architecture of NRP
  - Hardware, software
  - Physically distributed, logically one platform
- Continuously updated list of NDI/NRP services and implementation of the selected ones
  - Architecture and Core services working groups
- Access control implementation for NRP (and NDI)
- Training and education materials
- Implementation of specific repositories
  - Institutional, new thematic ones
  - Also templates for others
  - “How to create and manage your own repository over NRP”



e-INFRA CZ and EOSC

- Believe that **data-centric** approach is the core of the next generation federated computed and data infrastructures
- Extensive use of experience and know how gained through the previous work
  - Federated principles
  - Cloud computing in a distributed environments
  - Data and compute synergy (at logical and technical levels)
- The collaboration with research communities and large research infrastructures
  - **Partnership** as the second principle for building EOSC
    - Data centric the first one
- And the **international dimension**
  - EC co-funded project participation
  - Feedback, out of the box approaches

Summary

- EOSC implementation as an opportunity to evolve
- From compute to data-centric focus
  - Reshaping some of the core business of the e-INFRA CZ
- Huge potential, but also extreme challenge
  - Collaboration and partnership a key
- e-INFRA CZ can’t built EOSC CZ alone
  - Large consortium for NRP
  - Extensive collaboration with thematic clusters (>100 persons already)
- e-INFRA CZ prepared to extend the collaboration
- **EOSC implementation is a collaborative work**

Thanks and questions?

info@e-infra.cz, eosc-info@infra.cz

More information on

https://www.e-infra.cz

https://www.e-infra.cz/eosc