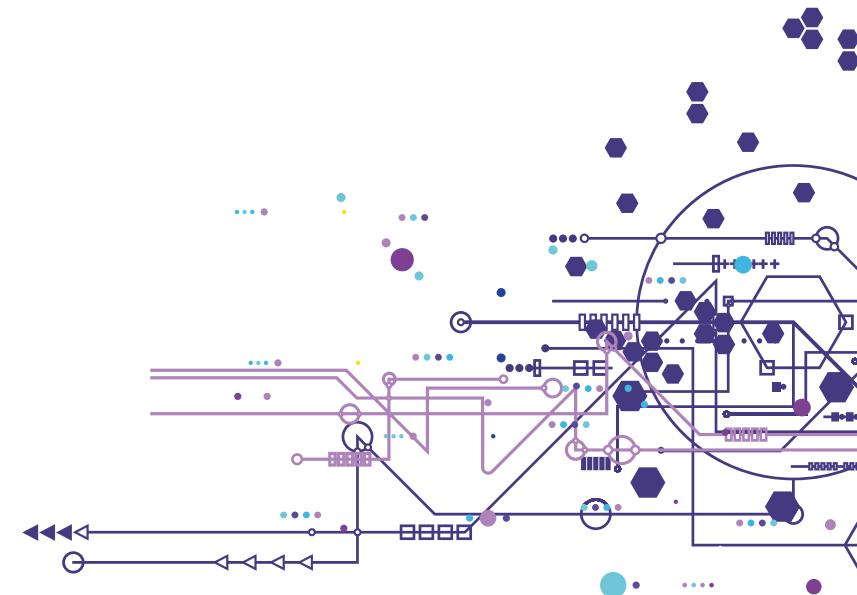


E-INFRA CZ

National plans for EGI and EOSC

Luděk Matyska

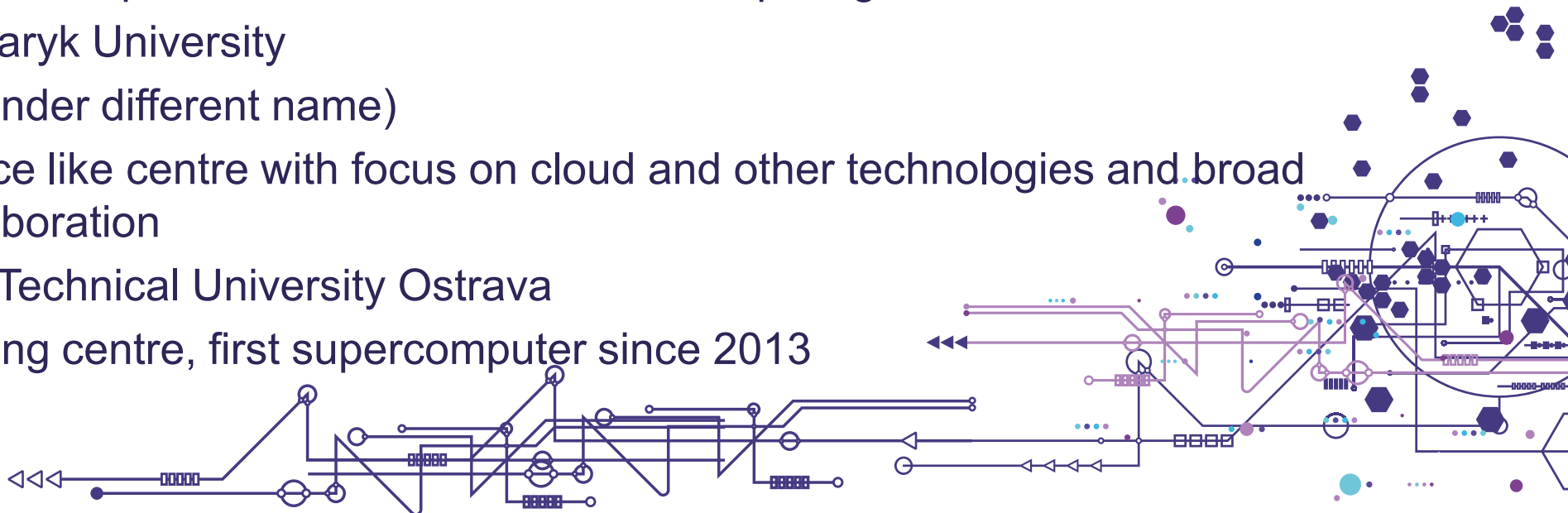
e-INFRA CZ



e-INFRA CZ – who we are?



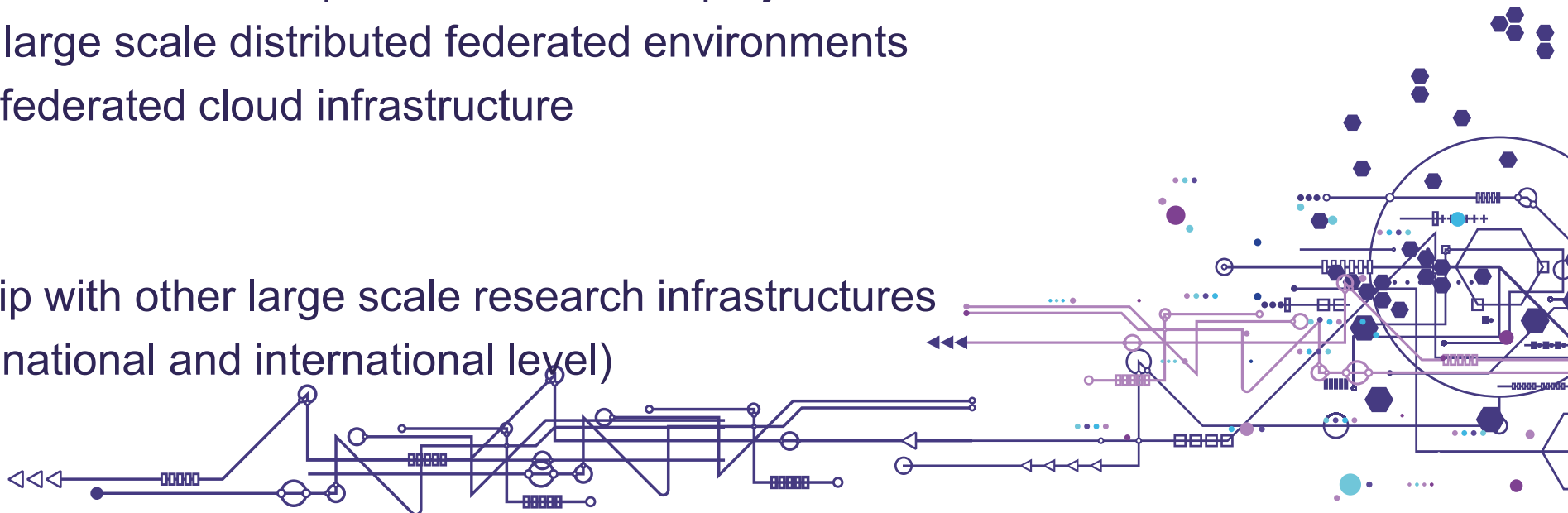
- Large national research e-infrastructure
- A consortium of three originally independent national e-infrastructures
- CESNET
 - Legal body since 1996
 - Owned by all public universities and Academy of Sciences
 - Originally NREN, expanded into Grid and Cloud computing and related services
- CERIT-SC at Masaryk University
 - Since 1994 (under different name)
 - Agile e-Science like centre with focus on cloud and other technologies and broad scientific collaboration
- IT4Innovations at Technical University Ostrava
 - Supercomputing centre, first supercomputer since 2013



e-INFRA CZ and EGI

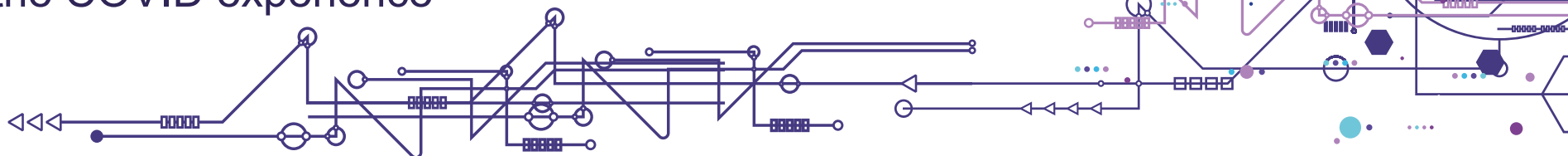


- CESNET a founding member of the EGI federation
 - A coordinator of the EGI Design Study project to setup EGI.eu
- Extensive involvement in EGI activities
 - CESNET and CERIT-SC involvements
 - Partner in (many) EC co-funded projects
 - Even a reverse relationship in the EOSC-Life project around LS AAI
- Shared interest in large scale distributed federated environments
 - From grids to federated cloud infrastructure
 - AAI
 - Applications
 - partnership with other large scale research infrastructures at national and international level)



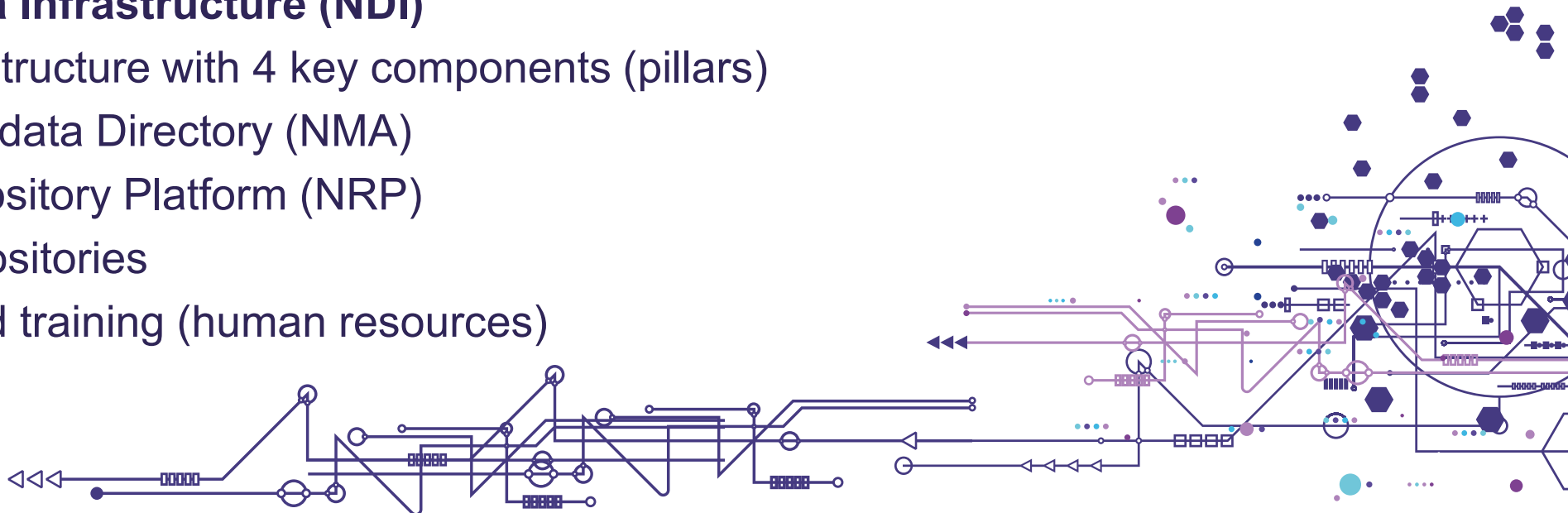
Recent focus towards EOSC

- New workplan for the whole e-INFRA CZ consortium (2023-2028)
- Focus on work with research data
 - Aligned with both external and internal motivation of our partners
 - Also e-INFRA CZ internal plans around role of HPC, clouds and other e-infrastructure elements
- Open Science principles
 - FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) data
 - Funding agencies pushing towards openness
 - Open Access
 - Data Management Plans
 - Large distributed research collaborations
 - e.g. the COVID experience



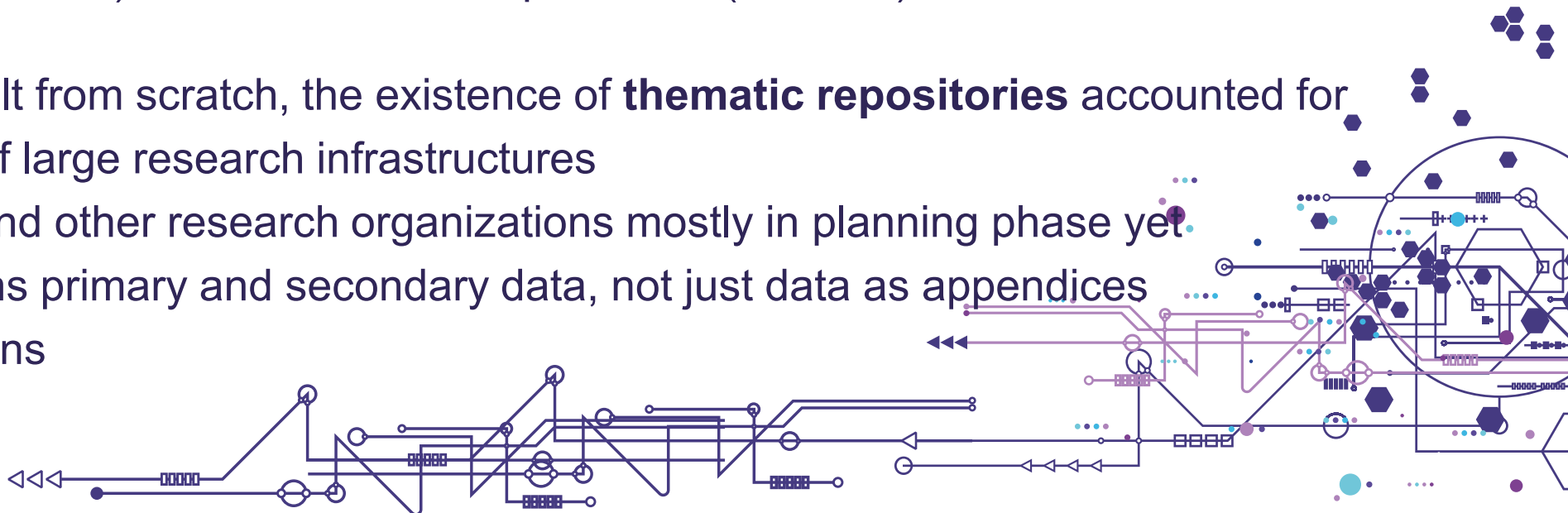
EOSC CZ Principles

- 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
- **Architecture of EOSC Implementation in the Czech Republic**
 - **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)



National Data Infrastructure

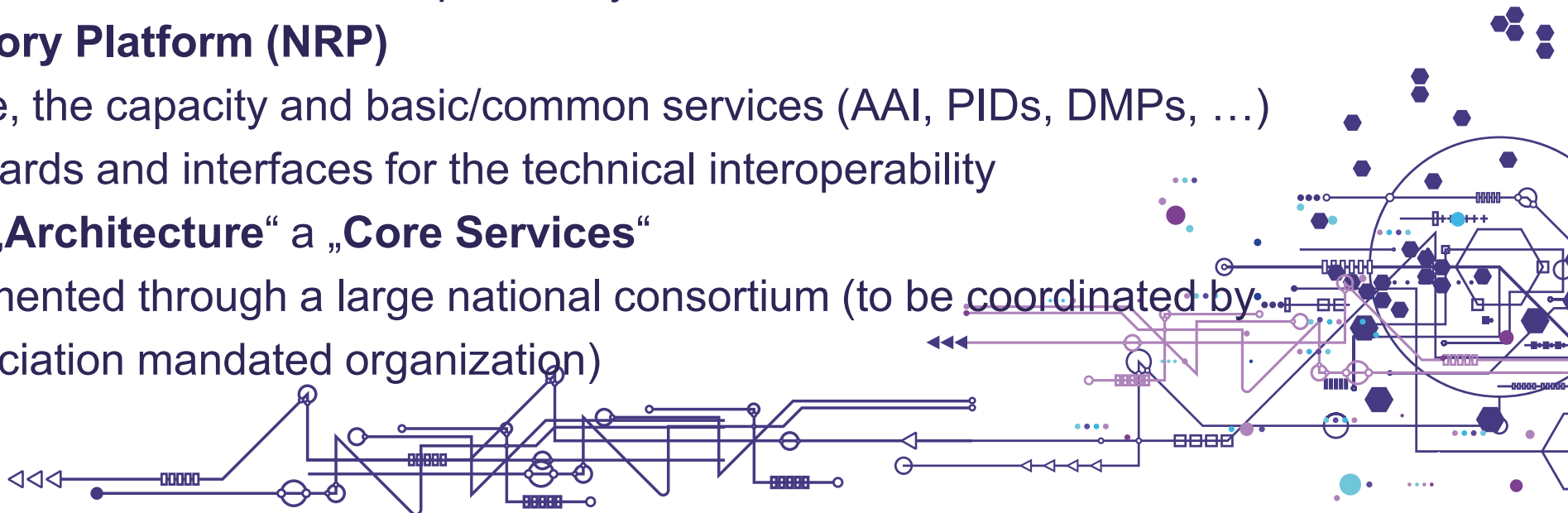
- Complex system with the following aims:
 - Basic **infrastructure** with **sufficient capacity** to store research data in Czechia
 - Open set of **key and other services** for data manipulation
 - Interaction with processing resources (HPC, clouds, ...)
- With parallel support of **human resources**
 - 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

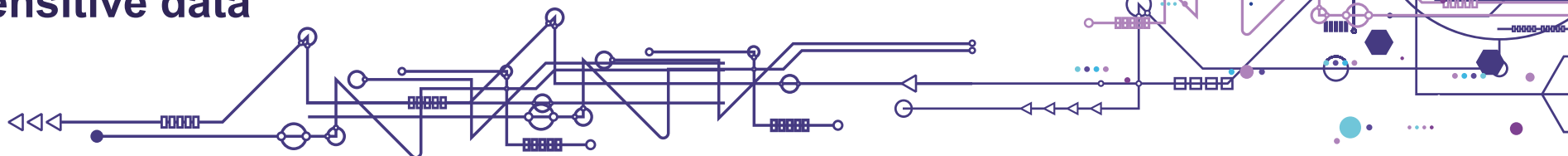
- **National Metadata Catalogue (NMA)**
 - Key component, guaranteeing findability and interoperability
 - Work group „**Metadata**“
 - Interacting directly with all other NDI components
 - Centrally managed NDI component
 - Still distributed; e-INFRA CZ responsibility
- **National Repository Platform (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

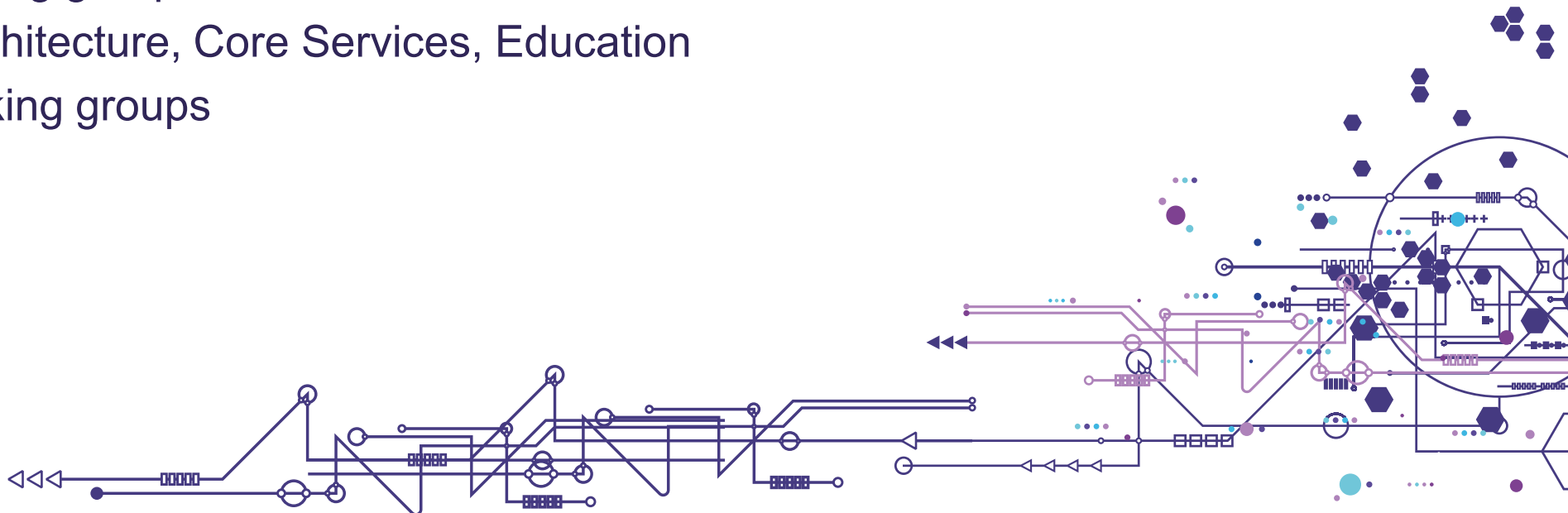
Other parts

- **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**



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
- 7+1 thematic working groups

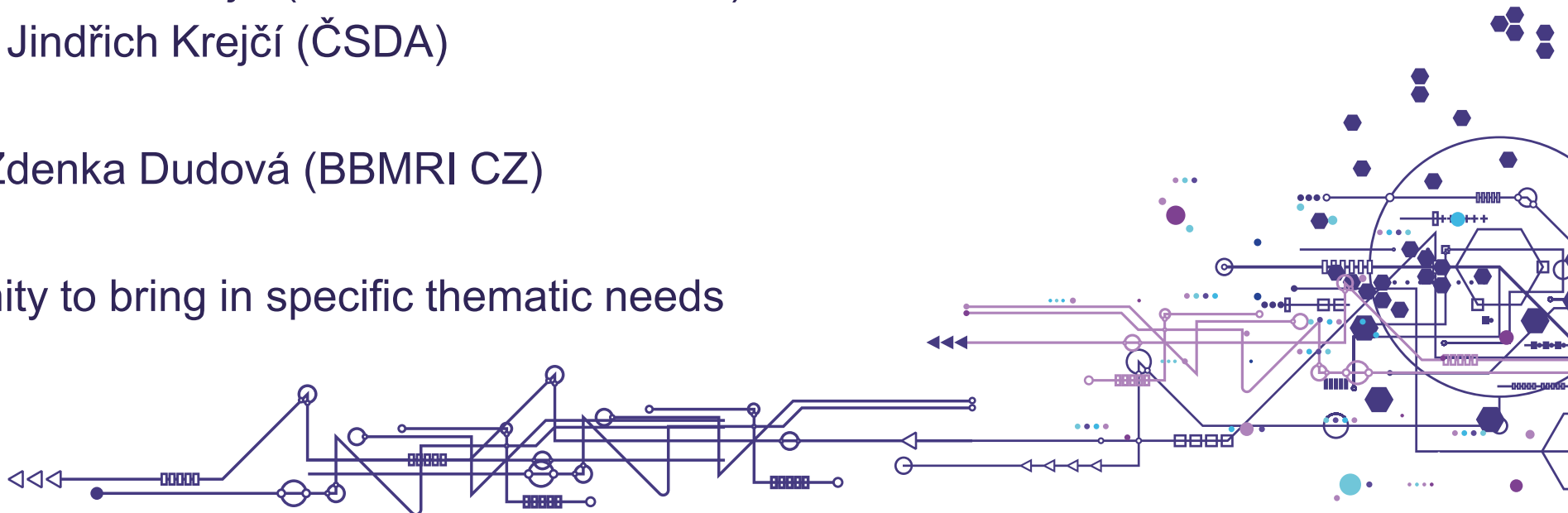


Thematic Working Groups

- 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)

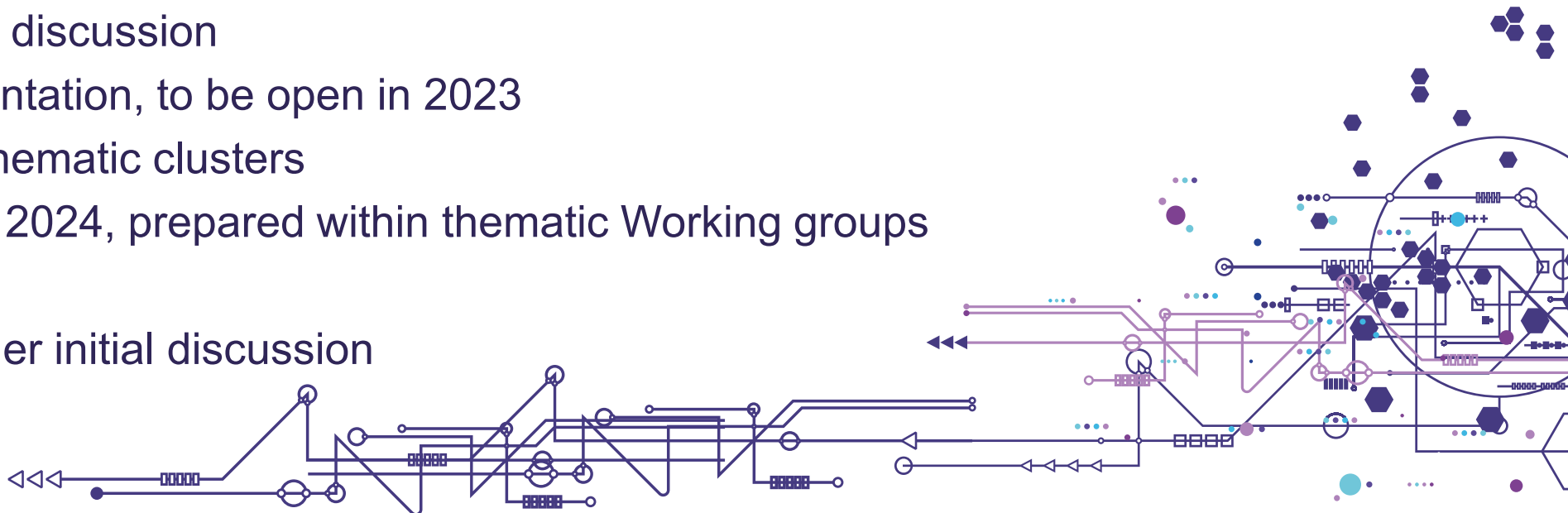
and

- Sensitive Data – Zdenka Dudová (BBMRI CZ)
- Provides opportunity to bring in specific thematic needs



Actual Plans

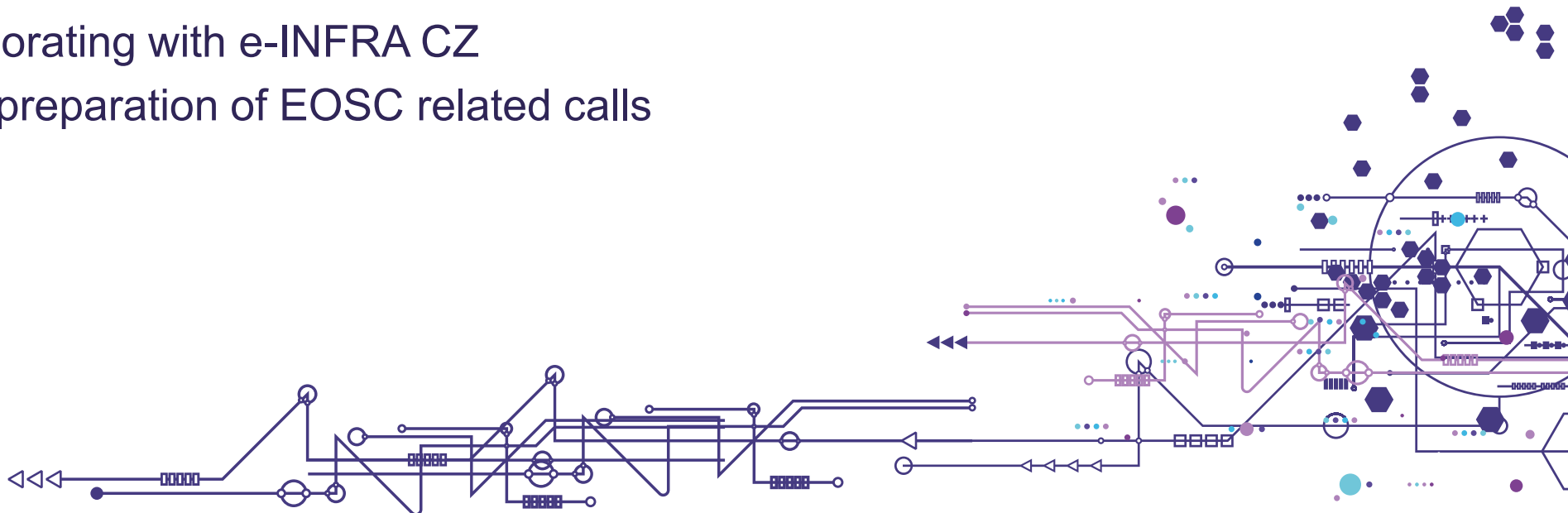
- EOSC implementation to be supported through structural funds
- Preliminary allocation of some 120 M€ for the 2023-2028 period
- Several calls under discussion
- First call already out
 - Projects **EOSC CZ** (EOSC Secretariat, NMA, training coordination) and **CARDS** (Metadata, PIDs)
- Second call 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



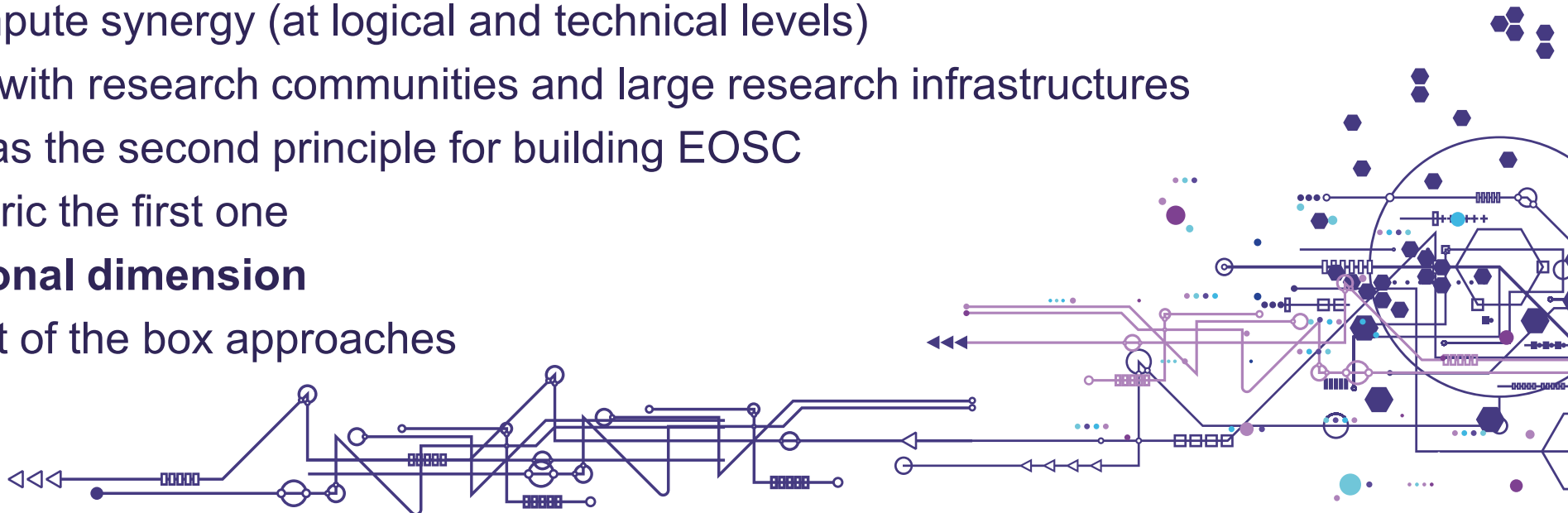
e-INFRA CZ Engagement



- e-INFRA CZ partner in all the activities
 - And coordinator of most of them
- Active contribution to the architecture
- Overall coordination of the working groups
 - Plus leadership of two core ones
- MEYS established an EOSC CZ Coordination Board
 - Closely collaborating with e-INFRA CZ
 - Input into the preparation of EOSC related calls



- 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 work in and with EGI federation
 - 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**
 - Feedback, out of the box approaches



Summary

- EOOSC 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 EOOSC CZ alone
 - Large consortium for NRP
 - Extensive collaboration with thematic clusters (>100 persons already)
- Feedback towards EGI
- **Can't prepare EOOSC implementation without EGI experience**





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>

e-infra.cz