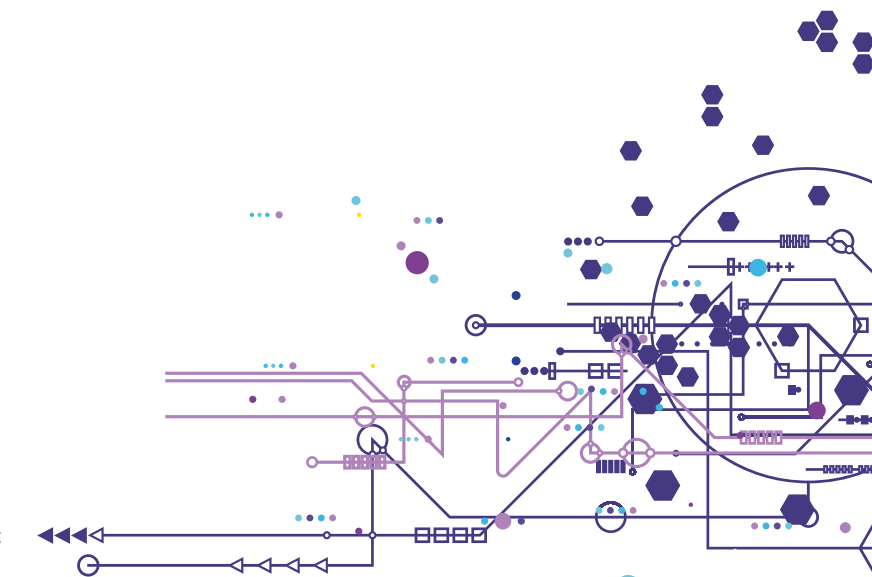
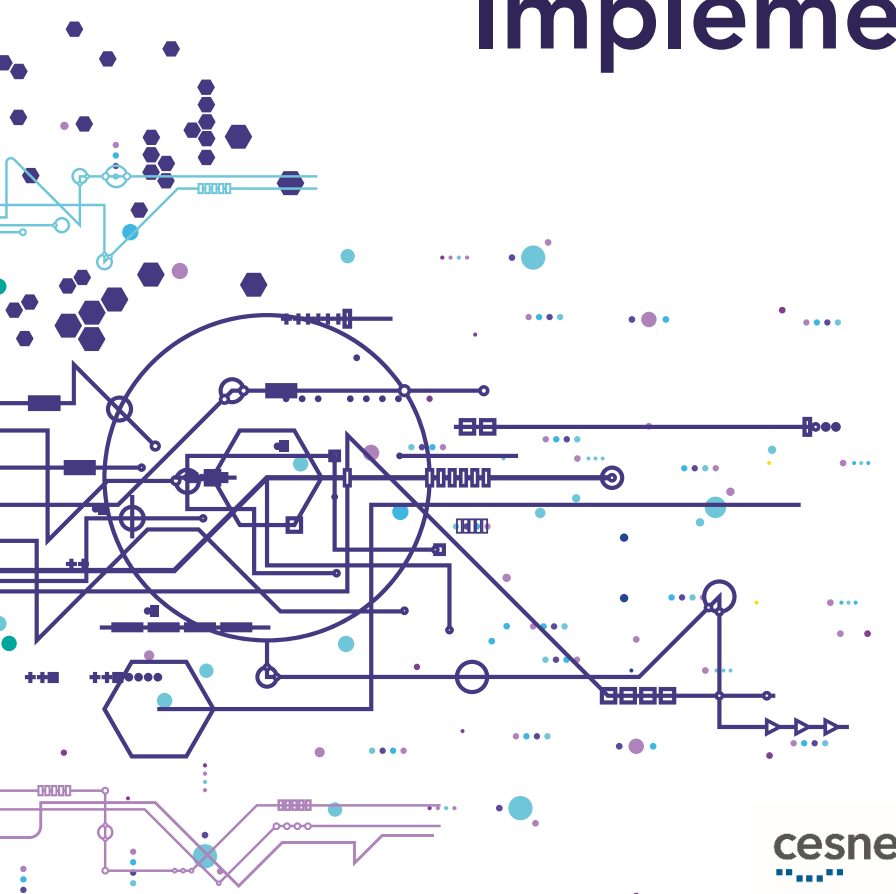


# EOSC CZ Roadshow

## Implementace EOSC v ČR

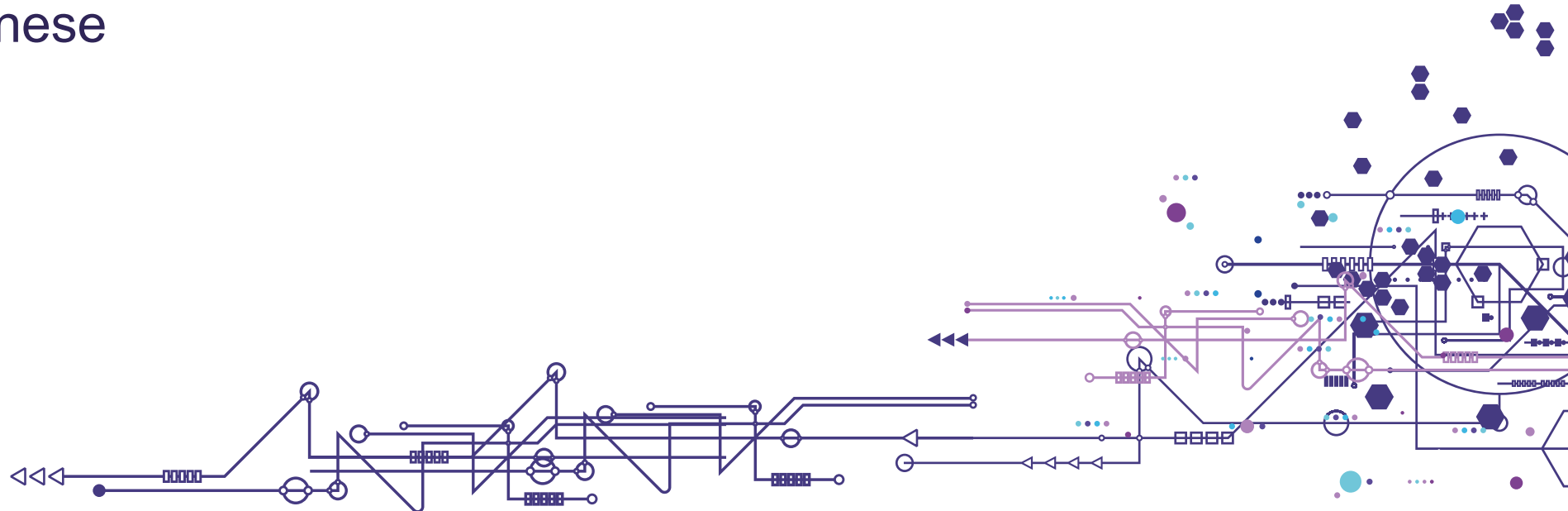
Luděk Matyska

e-INFRA CZ



# Obsah přednášky

- Open Science a European Open Science Cloud
- FAIR data
- EOSC v ČR – Architektura implementace
- Národní datová infrastruktura a její součásti
- Co to komu přinese
- Jak se zapojit



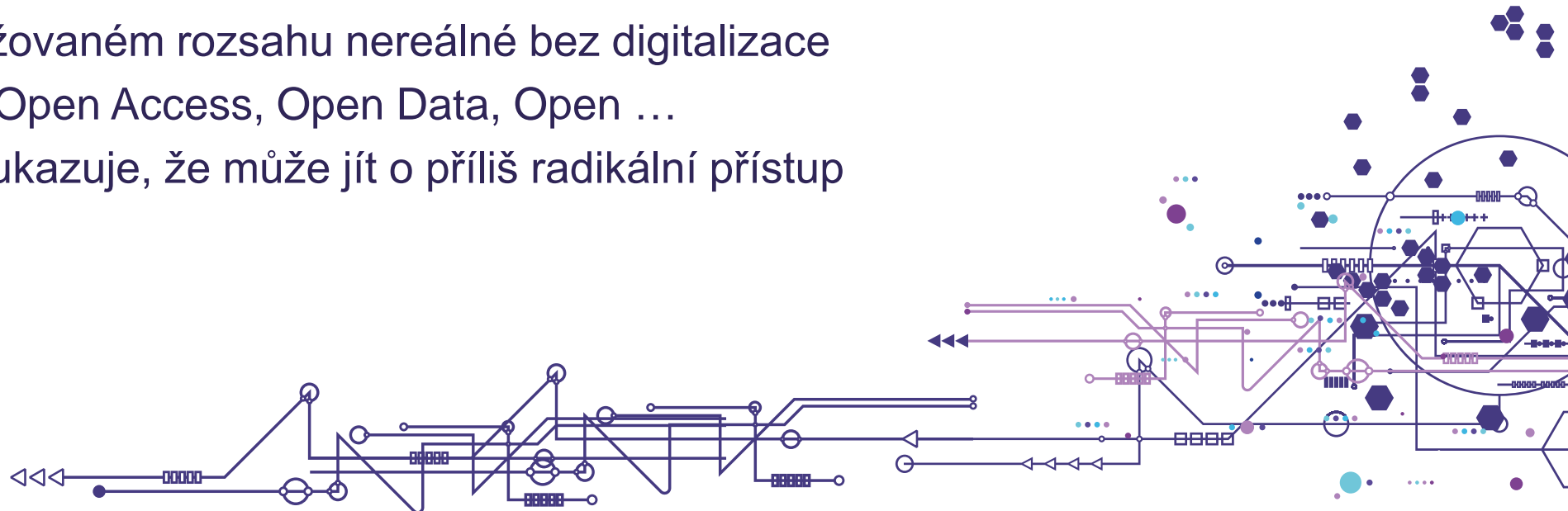
# Open Science, EOSC a FAIR Data

- Otevřená věda – nový způsob vědecké práce
  - Znalosti jsou dostupné a sdíleny
  - Otevřený přístup k výsledkům vědy
  - Otevřený přístup k (primárním) zdrojům a dalším digitálním artefaktům
  - Nově pojaté hodnocení vědeckého impaktu (ne monopol vydavatelů a jejich metrik)
    - Nejen publikace, ale i data a další (digitální) artefakty
  - Multidisciplinarita
    - Řešení problémů, které potřebují spolupráci více disciplín
  - Důraz na spolupráci a znovupoužití dosažených výsledků obecně
- Otevřená věda je zpravidla prezentována stojící na 8 pilířích
  - Jedním je **Open Access**, další dva jsou **EOSC** a **FAIR Data**

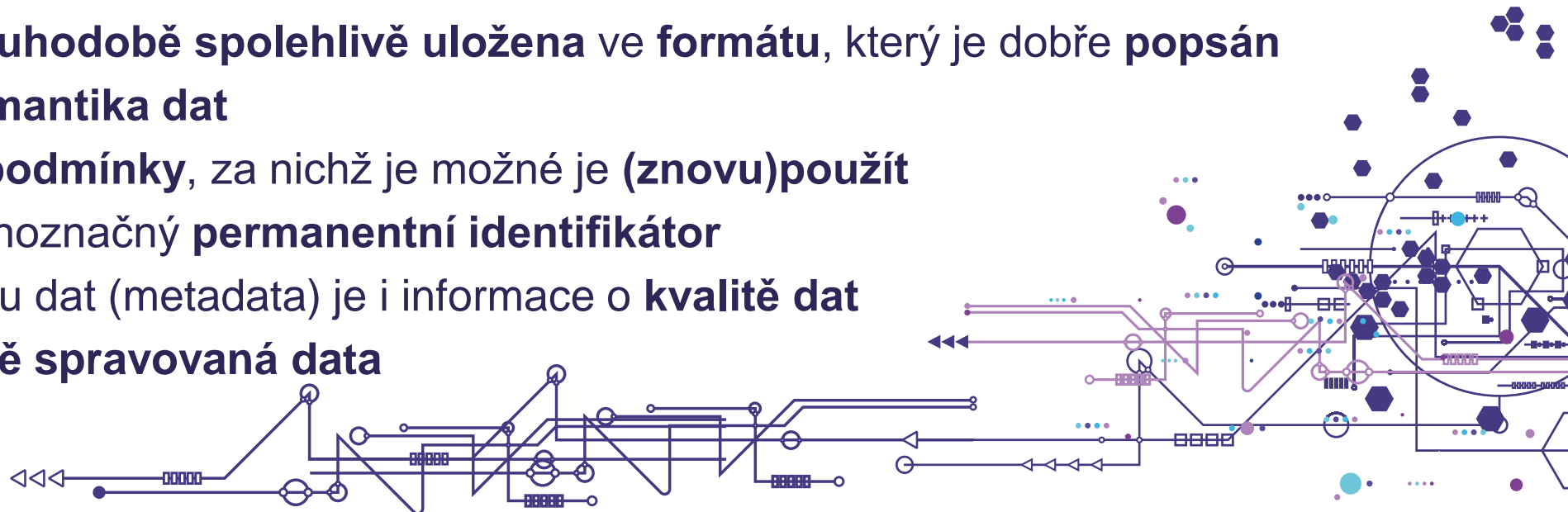


# Data v srdci Open Science

- Data a informace tvoří klíčový prvek konceptu Open Science
  - Data v nejširší možné interpretaci
    - Publikace, výzkumná data, software, workflows, ...
- Důraz na bezbariérový přístup
  - Vyžaduje i schopnost správná data vyhledat a znovupoužít
- IT jako „enabler“
  - Sdílení v uvažovaném rozsahu nereálné bez digitalizace
- Open Science → Open Access, Open Data, Open ...
  - Postupně se ukazuje, že může jít o příliš radikální přístup



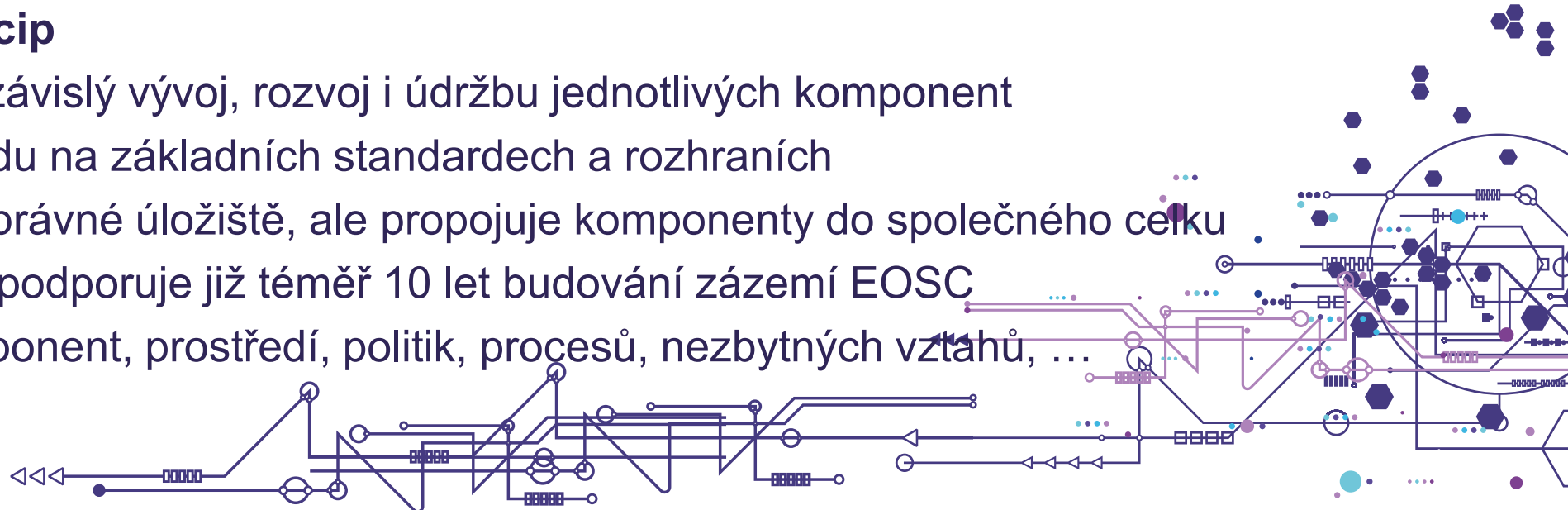
- Původní radikální koncept „Otevřených dat“ transformován do principu FAIR data
  - **F**indable – nalezitelná
  - **A**ccessible – dostupná
  - **I**nteroperable – interoperabilní (metadata)
  - **R**eusable – znovupoužitelná data
- Tedy
  - Data jsou **dlouhodobě spolehlivě uložena** ve formátu, který je dobře **popsán**
  - Je známa **sémantika dat**
  - Jsou známy **podmínky**, za nichž je možné je **(znovu)použít**
  - Data mají jednoznačný **permanentní identifikátor**
  - V rámci popisu dat (metadata) je i informace o **kvalitě dat**
- **FAIR data = řádně spravovaná data**



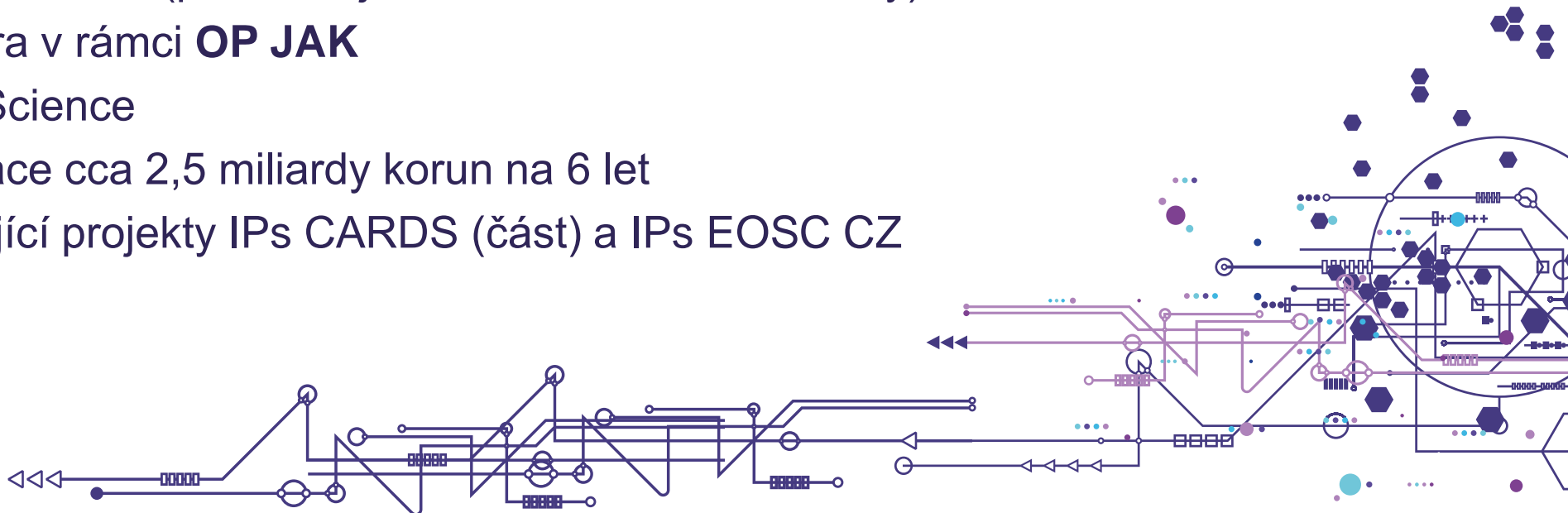
# European Open Science Cloud (EOSC)



- Technologické, programové i procesní zázemí, které podporuje uspořádanou práci s výzkumnými daty
  - sdružuje a poskytuje **kapacitu**, kde data lze dlouhodobě udržovat
  - poskytuje nástroje pro (snadnou) **správu** dat
  - poskytuje nástroje a prostředí pro **řízení přístupu** k datům
  - zpřístupňuje nástroje a prostředí pro zpracování (analýzu, ...) a propojuje je s daty
- **Federovaný princip**
  - umožňuje nezávislý vývoj, rozvoj i údržbu jednotlivých komponent
  - vyžaduje shodu na základních standardech a rozhraních
- Nevytváří jedno správné úložiště, ale propojuje komponenty do společného celku
- Evropská komise podporuje již téměř 10 let budování zázemí EOSC
  - Definice komponent, prostředí, politik, procesů, nezbytných vztahů, ...

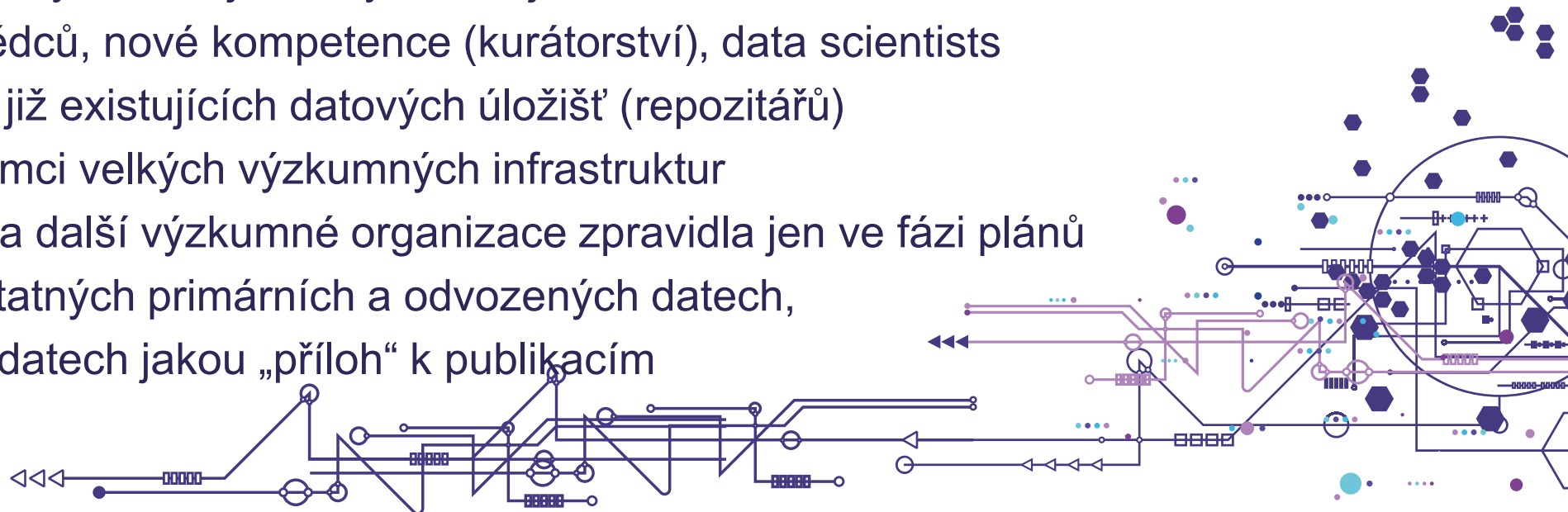


- **Architektura implementace EOSC v ČR**
  - Dokument vytvořen v rámci práce EOSC platformy při MŠMT v roce 2021
  - Definiuje základní komponenty a principy implementace EOSC v ČR
- **Řízení:** Koordinační výbor EOSC
  - Ustaven na přelomu roku 2021/2022
  - Poradní orgán MŠMT (předsedají nám. Wildová a Velčovský)
- Plánovaná podpora v rámci **OP JAK**
  - Výzvy Open Science
  - Celková alokace cca 2,5 miliardy korun na 6 let
  - Plus zastřešující projekty IPs CARDS (část) a IPs EOSC CZ



# Principy implementace EOSC v ČR

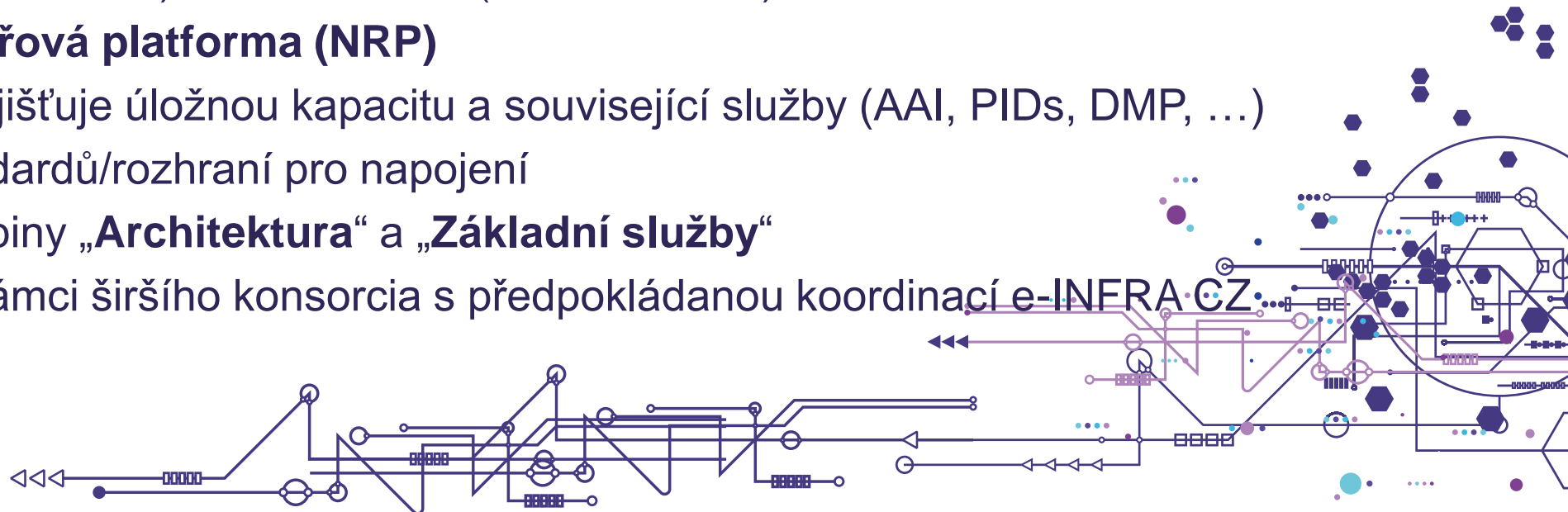
- EOSC CZ jako federovaný systém FAIR dat a služeb → **Národní datová infrastruktura**
- Komplexní systém, který se snaží zajistit
  - Základní **infrastrukturu a kapacitu pro ukládání** vědeckých dat v ČR
  - Otevřený soubor **klíčových a dalších služeb** pro práci s daty
  - Propojení na výkonné výpočetní systémy
- Za současné podpory a rozvoje lidských zdrojů
  - Vzdělávání vědců, nové kompetence (kurátorství), data scientists
- Počítá s existencí již existujících datových úložišť (repozitářů)
  - Zpravidla v rámci velkých výzkumných infrastruktur
  - Vysoké školy a další výzkumné organizace zpravidla jen ve fázi plánů
- Mluvíme o samostatných primárních a odvozených datech, ne pouze o datech jakou „příloh“ k publikacím





## Klíčové komponenty

- **Národní metadatový adresář (NMA)**
  - Klíčová komponenta, zajišťující vyhledatelnost a interoperabilitu dat
  - Pracovní skupina „**Metadata**“
  - Vazba na všechny ostatní aktivity v rámci NDI
  - Centrálně spravovaná komponenta
    - NTK (IPs CARDS) a e-INFRA CZ (IPs EOSC CZ)
- **Národní repozitářová platforma (NRP)**
  - Páteř NDI, zajišťuje úložnou kapacitu a související služby (AAI, PIDs, DMP, ...)
  - Definice standardů/rozhraní pro napojení
  - Pracovní skupiny „**Architektura**“ a „**Základní služby**“
  - Realizace v rámci širšího konsorcia s předpokládanou koordinací e-INFRA CZ (CESNET)



## Další součásti

- **Lidské zdroje**
  - Průřezová (horizontální) aktivita
  - Pracovní skupina „**Vzdělávání**“
  - Podpora v rámci prakticky všech zvažovaných projektů
- **Existující tematické (oborově-vědní) repozitáře**
  - Podpora existujících aktivit a repozitářů (zejména v infrastrukturách)
    - Zajištění kontinuity, podpora dlouhodobé udržitelnosti
  - Podpora nových oborově specifických potřeb
    - Ne nové (na zelené louce), ale s využitím kapacity NRP
  - Sjednocení s požadavky NMA a NRP
    - Zajištění plné interoperability na technické i logické (metadatové) úrovni
- Specifická **péče o citlivá data**



# Národní datová infrastruktura (NDI)

Národní metadatový katalog (NMA)

Národní repozitářová platforma (NRP)

Služby NDI

Tematické repozitáře

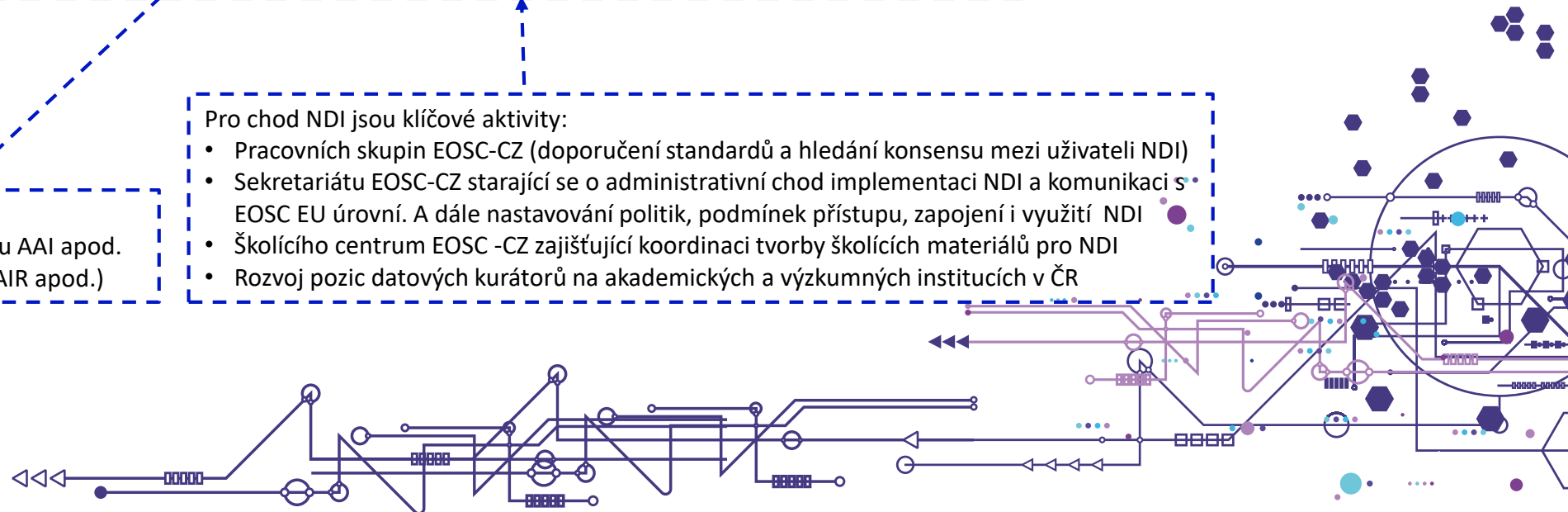
Ostatní repozitáře

Služby NDI jsou dvojího typu:

- Základní technické služby typu AAI apod.
- EOSC-CZ Core služby (DMP, FAIR apod.)

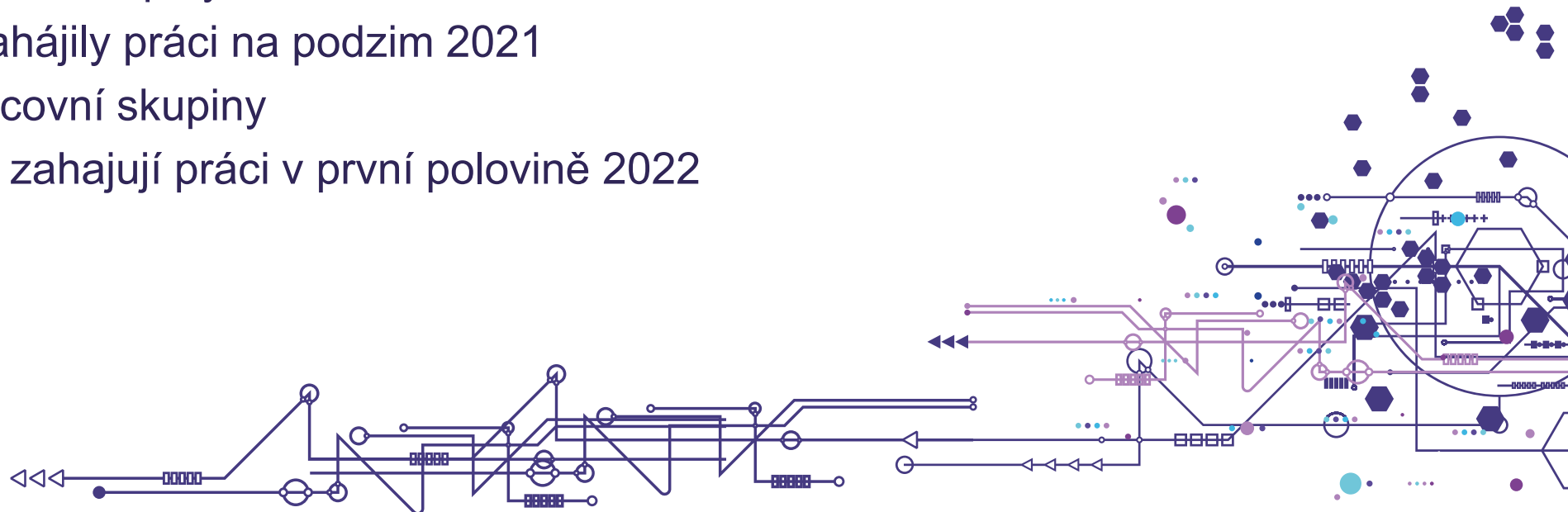
Pro chod NDI jsou klíčové aktivity:

- Pracovních skupin EOSC-CZ (doporučení standardů a hledání konsensu mezi uživateli NDI)
- Sekretariátu EOSC-CZ starající se o administrativní chod implementaci NDI a komunikaci s EOSC EU úrovní. A dále nastavování politik, podmínek přístupu, zapojení i využití NDI
- Školícího centrum EOSC -CZ zajišťující koordinaci tvorby školících materiálů pro NDI
- Rozvoj pozic datových kurátorů na akademických a výzkumných institucích v ČR



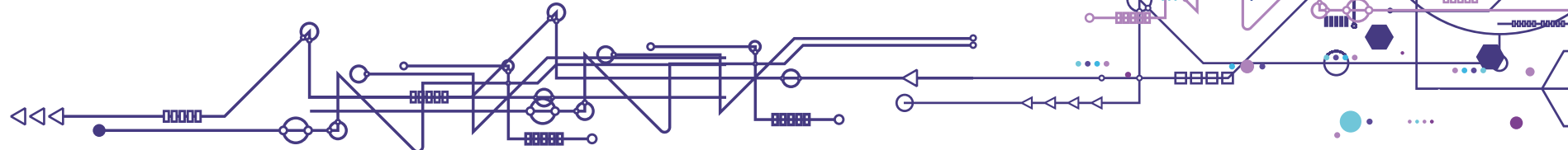
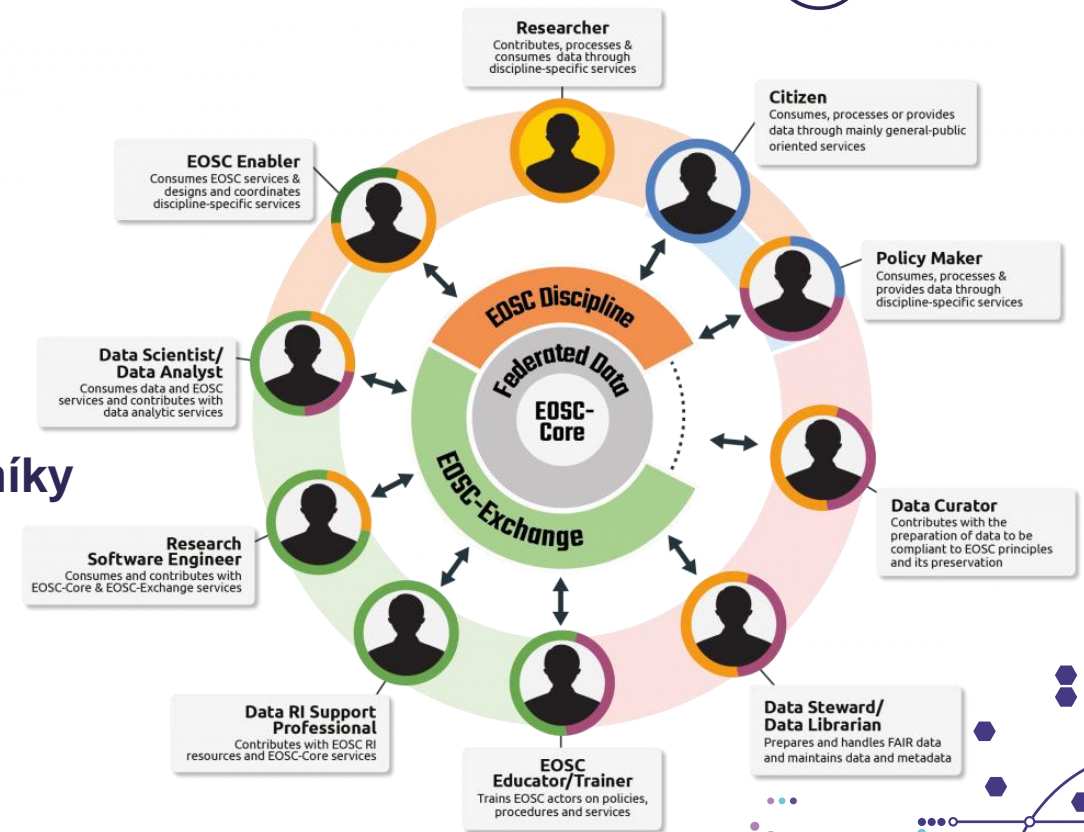
# Přípravné kroky implementace EOOSC v ČR

- Pracovní skupiny
  - Otevřená platforma
  - Diskuse základních principů (implementace) EOOSC
    - co vědecká komunita potřebuje a očekává
    - a jak tato očekávání a potřeby nejlépe naplnit
- 4 průřezové pracovní skupiny
  - Ustaveny a zahájily práci na podzim 2021
- 7+1 tematické pracovní skupiny
  - Ustavovány a zahajují práci v první polovině 2022



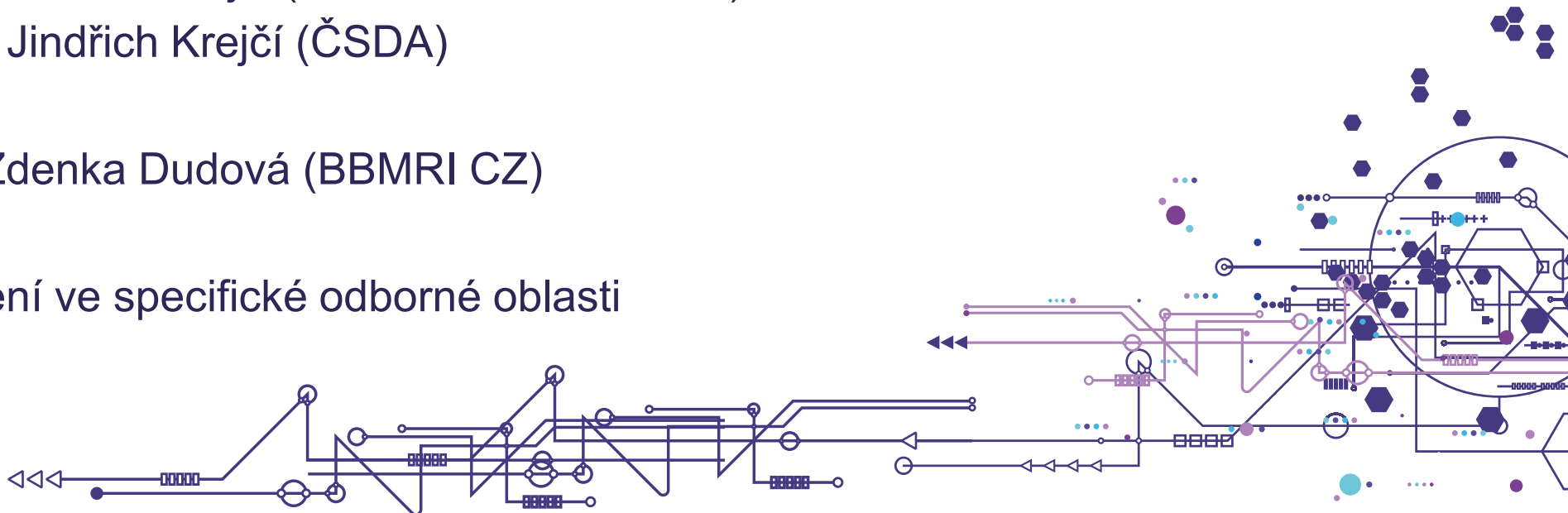
# Průřezové pracovní skupiny EOSC CZ

- Metadata – Petra Černohlávková (NTK)
- Architektura – David Antoš (CESNET)
- Základní služby – Michal Růžička (MU)
- Vzdělávání – Radka Římanová (UK)
  - vzdělávání a podpora pro **výzkumné pracovníky**
    - rozšiřování kompetencí v oblasti **správy a sdílení výzkumných dat**
  - vzdělávání pro “**data support**”
    - lidské zdroje
    - pracovní pozice
    - studijní programy
    - Kurzy
- Stále otevřená možnost zapojení



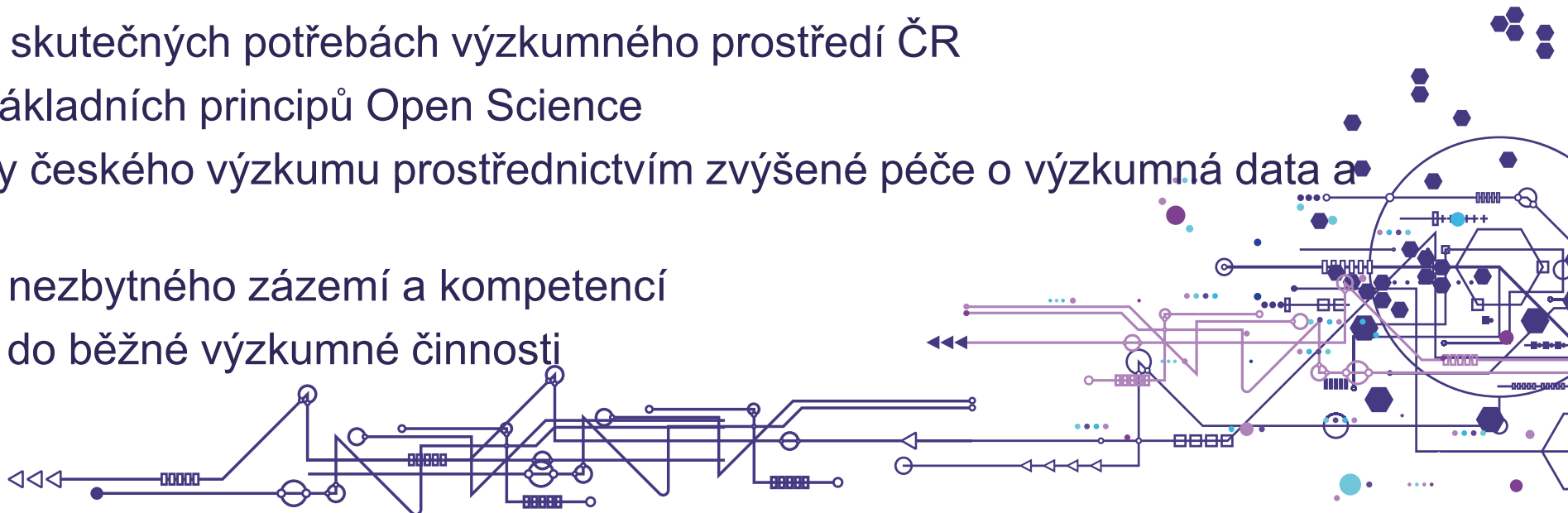
# Tematické (oborově-vědní) pracovní skupiny

- 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)
- a
- Sensitive Data – Zdenka Dudová (BBMRI CZ)
  - Prostor pro zapojení ve specifické odborné oblasti



# Další kroky

- Výzvy OP JAK – příprava projektových konsorcií a konkrétních projektů
  - Vlastní implementace EOSC v ČR
- Pracovní skupiny
  - Návrh obsahu
  - Postupná „krystalizace“ konsorcií
- Budování „národního EOSCu“ zdola
  - Postaveno na skutečných potřebách výzkumného prostředí ČR
  - Při naplnění základních principů Open Science
  - Zvýšení kvality českého výzkumu prostřednictvím zvýšené péče o výzkumná data a práci s nimi
    - Vytvoření nezbytného zázemí a kompetencí
    - Integrace do běžné výzkumné činnosti

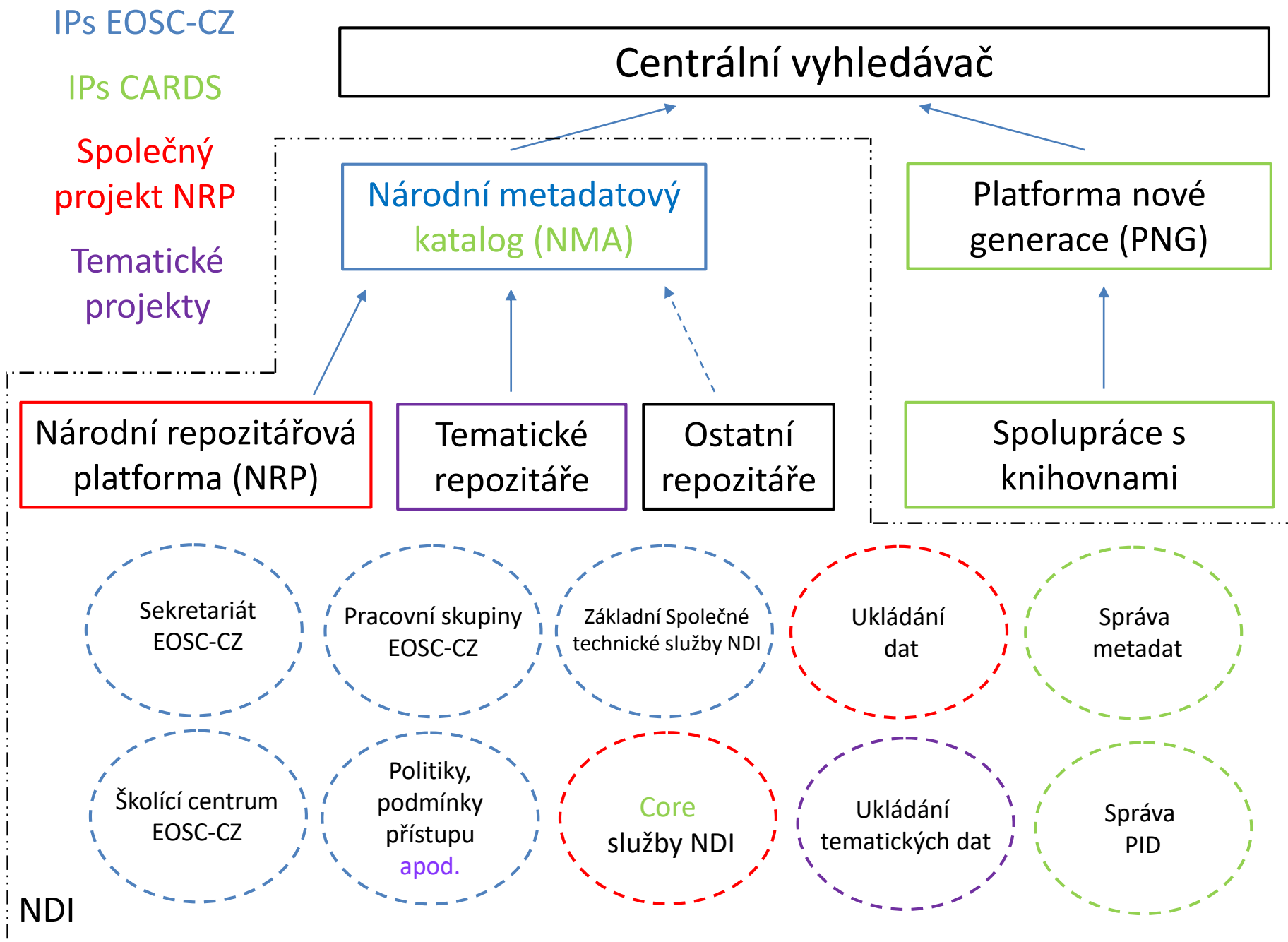


# Zvažovaná struktura projektů

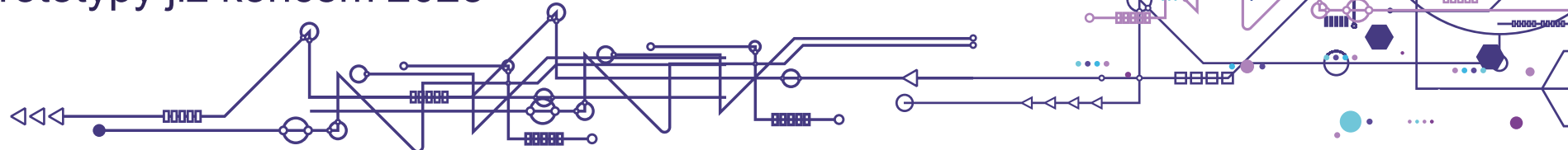
- Dvojice zastřešujících – individuální projekty systémové
  - IPs CARDS a IPs EOSC-CZ
    - Společné zázemí, portál, monitoring, ... (Sekretariát EOSC CZ)
    - Metadata a NMA
    - PIDs
    - Jádro AAI, vysokokapacitní datové přenosy
- Konsorciální projekt budování Národní repozitářové platformy
  - Infrastruktura, nepředpokládají se kompetitivní projekty
- Tematicky orientované projekty
  - V diskusi tematických pracovních skupin
  - Zatím otevřený počet a časování (první a druhá výzva)
- Podpora vzdělávání, nové kompetence





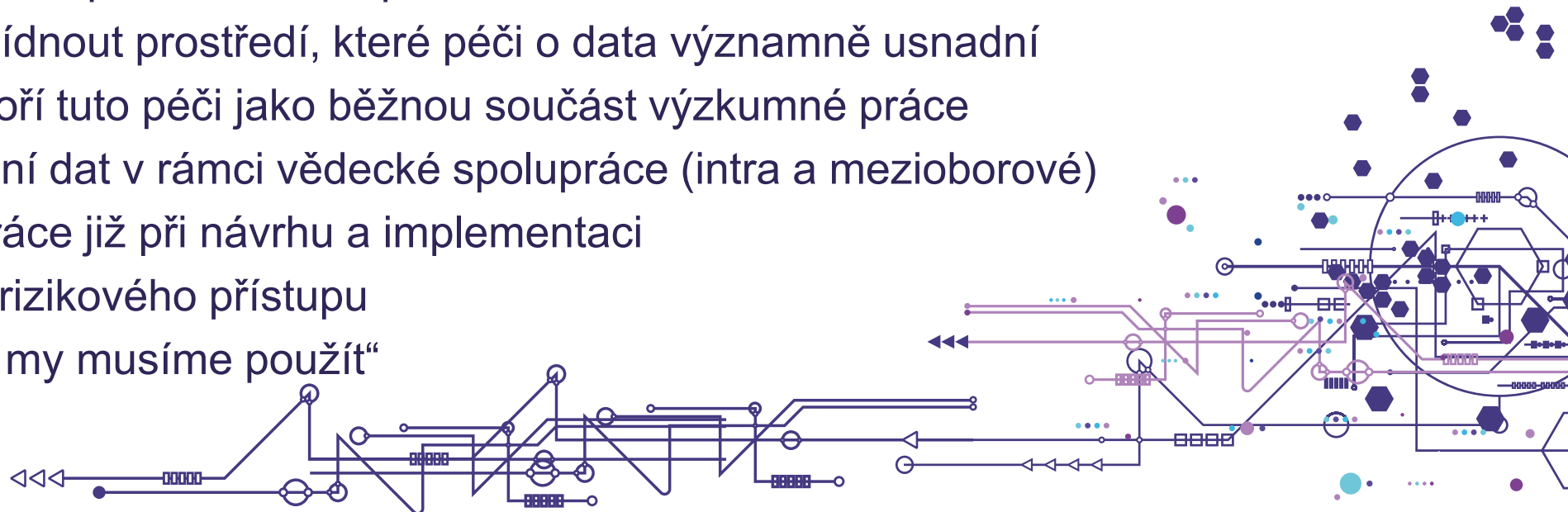


- Projekty slouží k **budování NDI**
  - Partneři konsorcií instituce/týmy se zkušeností z oblasti budování a provozu zázemí (technického i znalostního) pro práci s daty
  - Závazek udržitelnosti vytvořeného prostředí
    - Ideálně integrace do institucionálních či infrastrukturních plánů
    - Přirozená role e-INFRA CZ i dalších výzkumných infrastruktur
- **Užití výsledků není nijak vázáno na zapojení do projektů**
  - NDI jako součást zázemí Open Science v ČR, nikoliv „closed club“
  - Zpřístupnění vytvořeného prostředí jednou z povinností pro projektová konsorcia
  - Monitorováno KV EOSC CZ za podpory IPs EOSC-CZ
- První šířeji použitelné výsledky by měly být k dispozici v roce 2024
  - Pilotování a prototypy již koncem 2023



# Proč to děláme

- Principy Open Science představují základ strategie směřování výzkumu v EU
  - Korektně „opečovaná“ data – základ excelentního reprodukovatelného výzkumu
- EOSC vnímán jako kritický nástroj implementace těchto principů
  - Průřezové evropské partnerství
  - Významné finanční prostředky na celoevropské úrovni i v členských státech
- Využití této příležitosti pro kvalitativní posun i na národní úrovni
  - Vzdělat a nabídnout prostředí, které péči o data významně usnadní
  - Umožní/podpoří tuto péči jako běžnou součást výzkumné práce
  - Usnadní sdílení dat v rámci vědecké spolupráce (intra a mezioborové)
- Vzájemná spolupráce již při návrhu a implementaci
  - Minimalizace rizikového přístupu
  - „Oni udělají a my musíme použít“



# Co nám to všem přinese

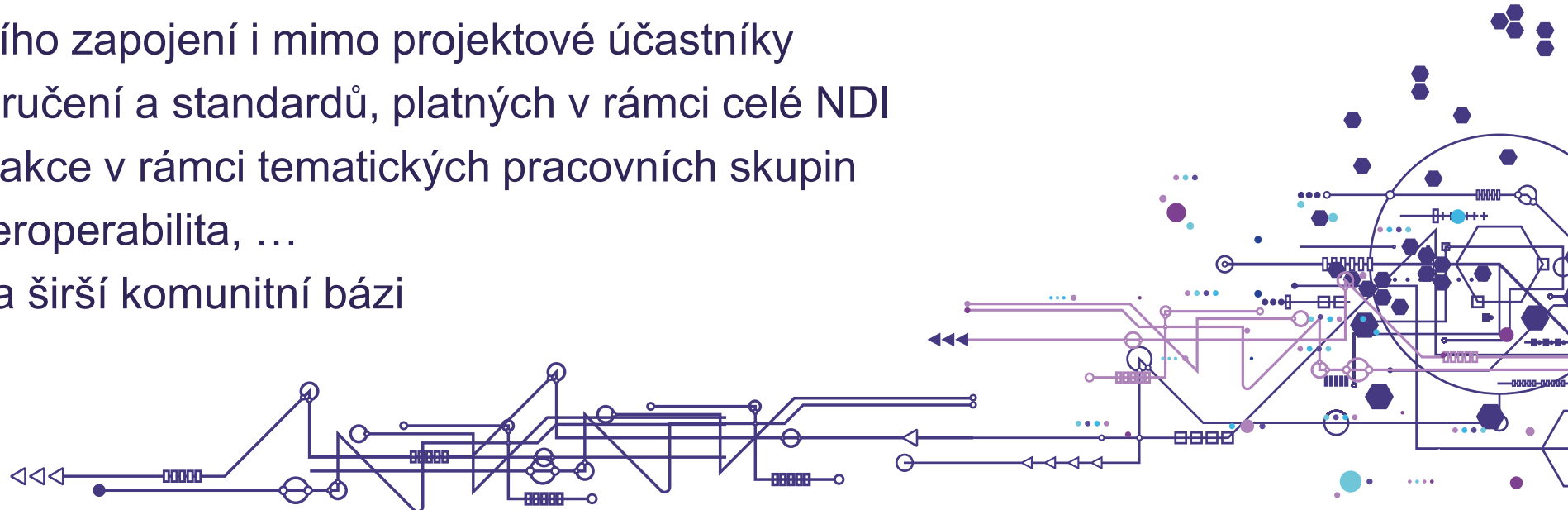
- **NDI** – komplexní prostředí pro strukturovanou práci s výzkumnými daty
  - Robustní infrastruktura pro ukládání a zpřístupnění dat, vazba na zpracování
    - Živá i archivní data
    - Kontrola přístupu k datům i metadatům v rukou autorů/majitelů dat
  - Podpůrné nástroje (DMP, FAIRifikace, ...)
- **Viditelné a vyhledatelné ontologie a metadata**
- **Propojení** našich repozitářů a služeb na **celoevropské prostředí**
- **Školení a vzdělávání**
- Aktivní zázemí pro sdílení a přenos informací o EOSCu
  - Sekretariát EOSC CZ, Portál EOSC CZ, konzistentní informace z EOSC Asociace
  - Podpora, konzultace, ...



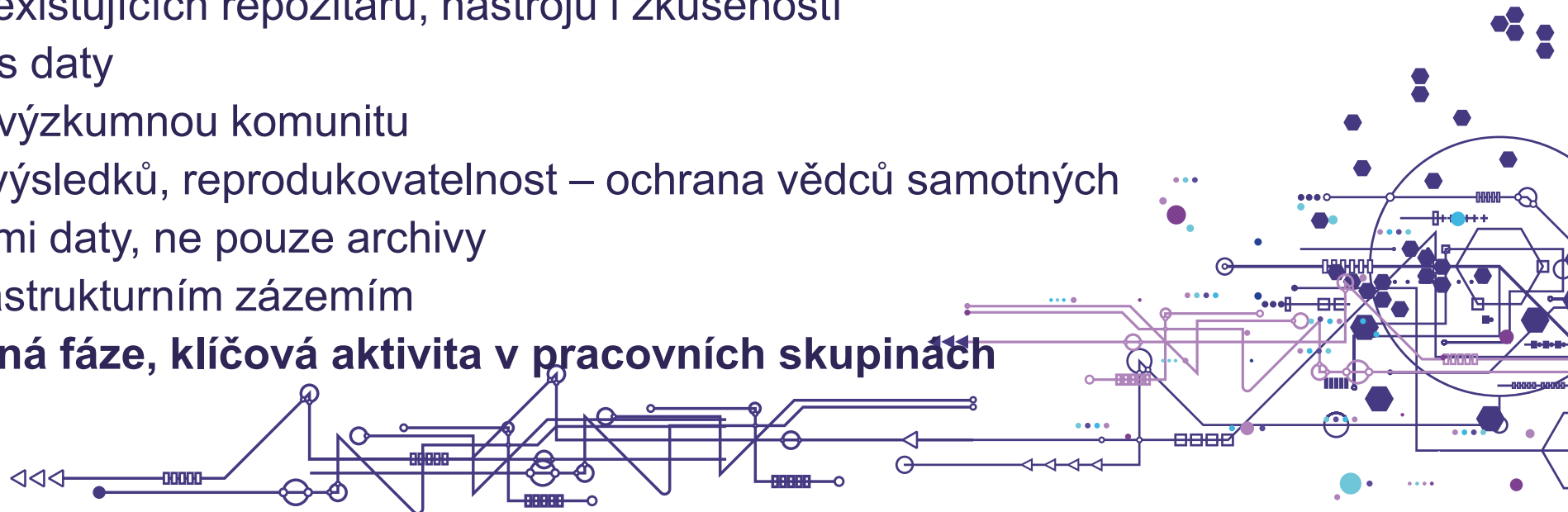
# Jak se zapojit

## A co to znamená

- Aktivní účastí v pracovních skupinách (průřezových i tematických)
  - Možnost „být přítom“
  - Ovlivňovat směřování přípravy realizace
  - Zapojit se do diskuse o obsahu, formě i zaměření budoucích projektů
- Pracovní skupiny budou klíčové i v procesu implementace
  - Budování komunit, předávání znalostí
  - Nástroj aktivního zapojení i mimo projektové účastníky
  - Definice doporučení a standardů, platných v rámci celé NDI
- Mezioborové interakce v rámci tematických pracovních skupin
  - Metadata, Interoperabilita, ...
  - Spolupráce na širší komunitní bázi



- EOSC jako příležitost k zavedení strukturovanější práce s výzkumnými daty
  - Technické i procesní zázemí, služby, nástroje, ...
  - Orientace na podporu znalostí (lidé, ne pouze technologie)
- Zapojení všech významných institucí
  - Budování Národní repozitářové platformy
  - Pokrytí oborových specifik, zapojení zejména velkých výzkumných infrastruktur
  - Začlenění již existujících repozitářů, nástrojů i zkušeností
- Usnadnění práce s daty
  - Orientace na výzkumnou komunitu
  - Vyšší kvalita výsledků, reprodukovatelnost – ochrana vědců samotných
  - Práce i s živými daty, ne pouze archivy
- Integrace s e-infrastrukturním zázemím
- **Aktuálně přípravná fáze, klíčová aktivita v pracovních skupinách**





**Děkuji za pozornost**

**[eosc-info@e-infra.cz](mailto:eosc-info@e-infra.cz)**

**Další informace na <https://www.e-infra.cz/eosc>**

A large, stylized logo for e-infra.cz. It consists of a large circle with a thick black outline, and a smaller circle inside it, also with a thick black outline. The text 'e-infra.cz' is centered within the larger circle.

**e-infra.cz**

- Formálně založena 29. 7. 2020 v Belgii, formálně potvrzena 11.9.2020 (Royal Decree)
  - EOSC AISBL (nezisková organizace)
  - V popisu má výzkum, financování výzkumu i poskytování služeb
  - President a 8 ředitelů voleni (každý rok obměna části ředitelů), nevolený sekretář
- Mise
  - Representovat širokou komunitu EOSC „stakeholderů“ a dát jí jeden hlas při komunikaci s EK
  - Spolupracovat s EK při definici a sjednocení plánů EK a potřeb EOSC
  - Potenciálně i poskytovat základní služby pro bezbariérovou práci s daty
- Aktuálně za ČR
  - 4 řádní členové – e-INFRA CZ (**CESNET**, MU, VŠB-TUO) a NTK
  - 1 pozorovatel – Ústav fyzikální chemie J. Heyrovského (AV ČR)

