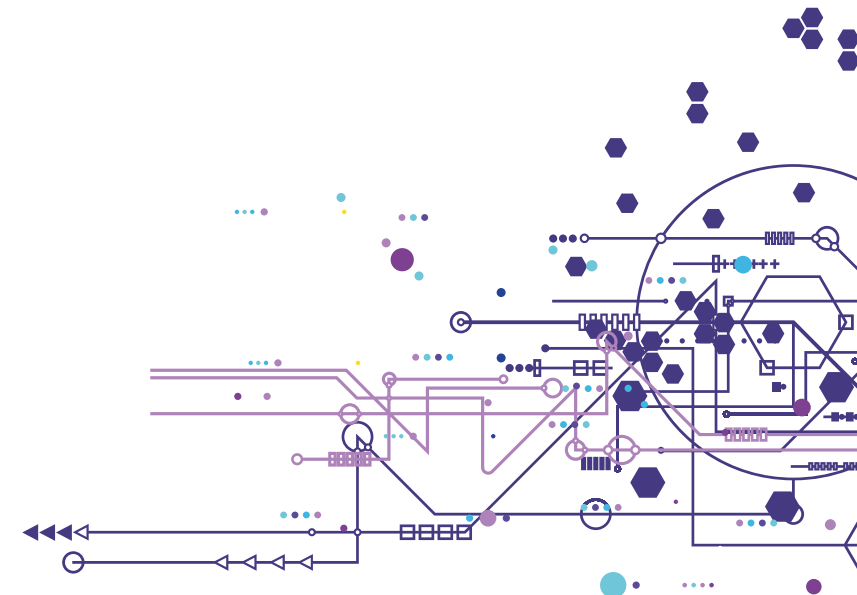
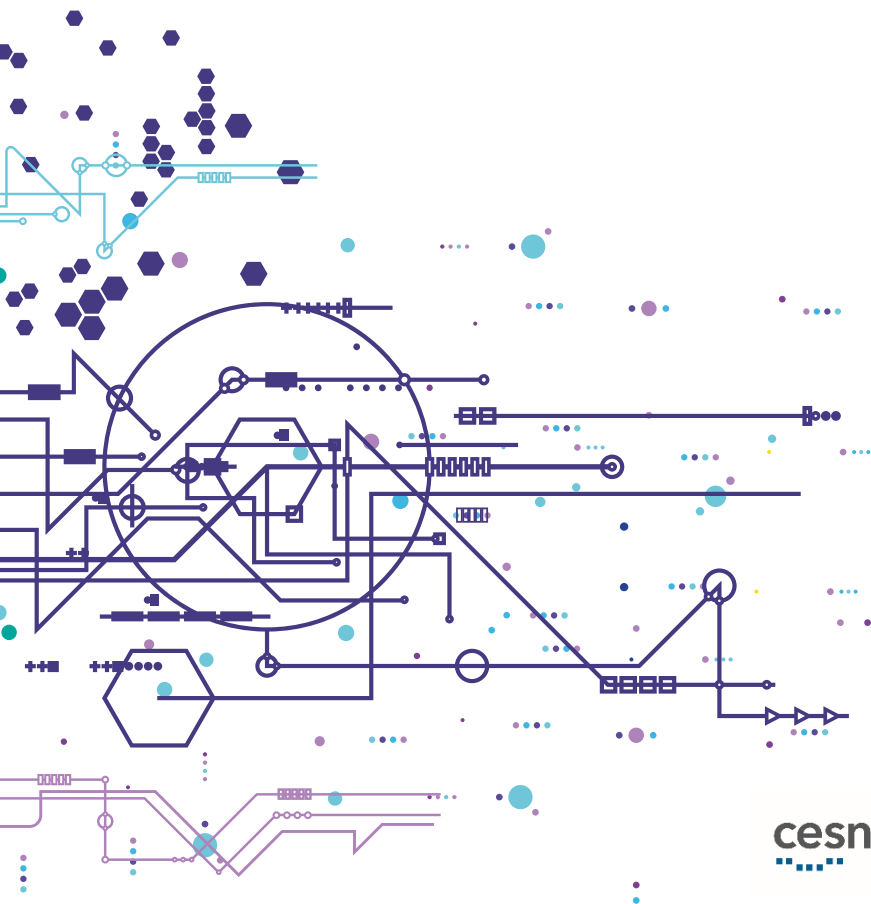


E-INFRA CZ a EOSC

Luděk Matyska

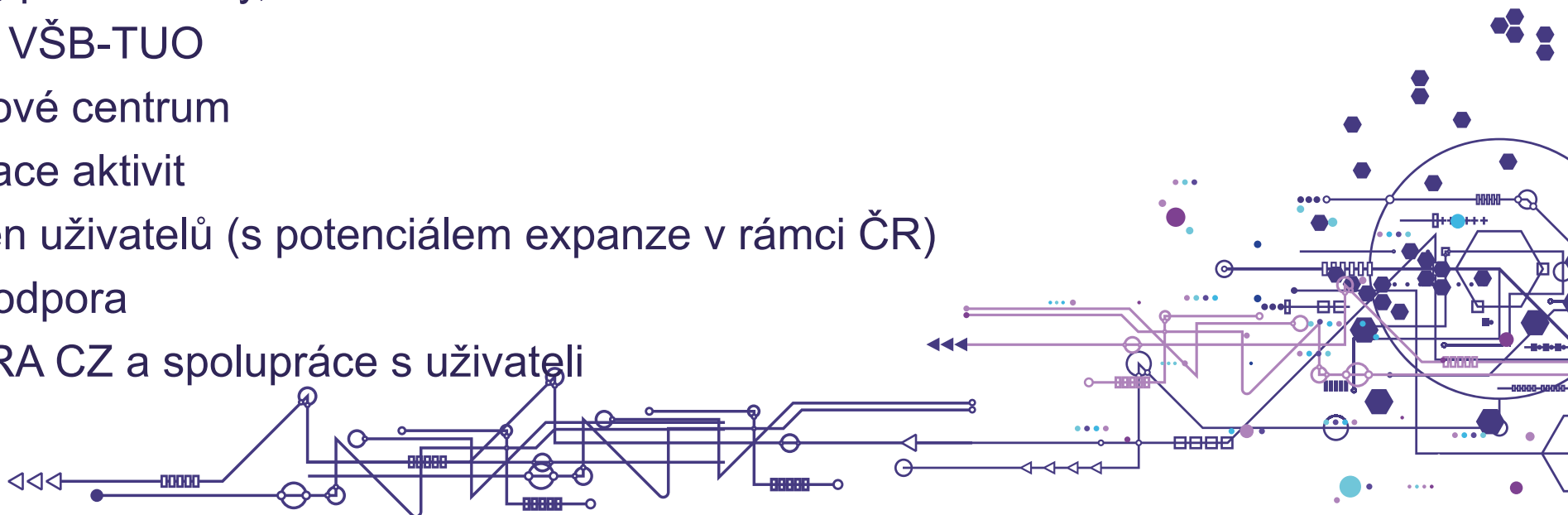
e-INFRA CZ



e-INFRA CZ – Stručné představení



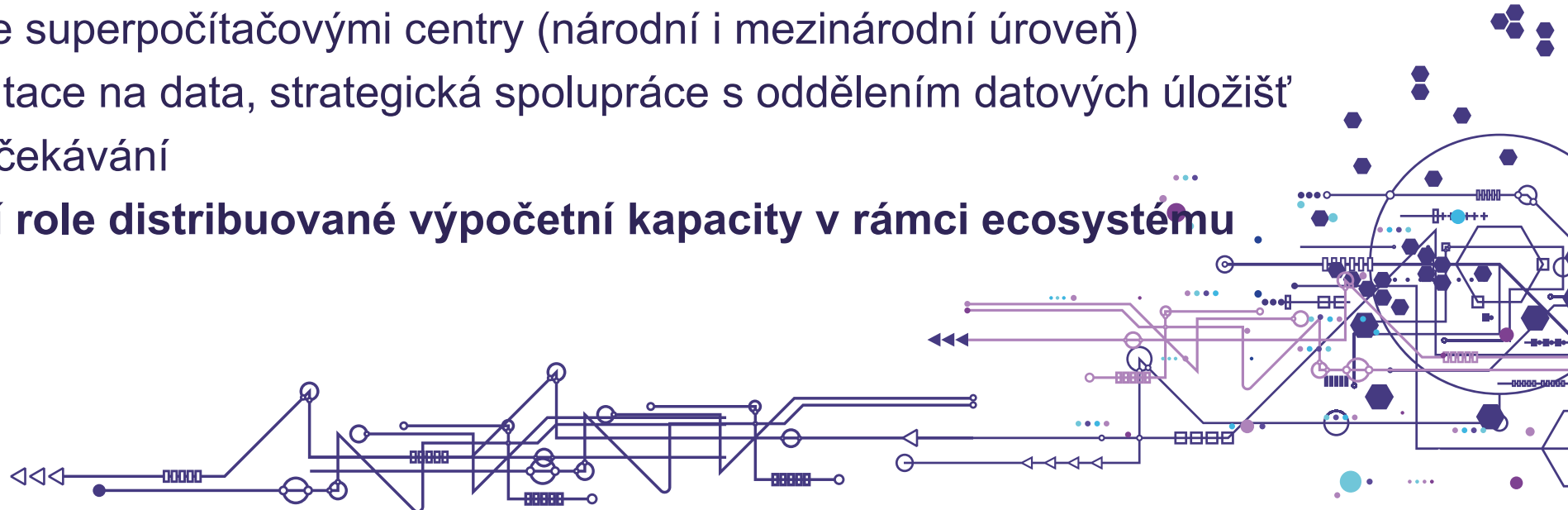
- Jediná národní velká výzkumná e-infrastruktura
- Konsorcium tří původně samostatných velkých výzkumných infrastruktur
- CESNET, koordinátor
 - Síť, Metacentrum, datová úložiště, multimediální přenosy, ...
- CERIT-SC na ÚVT MU
 - MetaCentrum, práce s daty, Life Science
- IT4Innovations na VŠB-TUO
 - Superpočítačové centrum
- Postupná koordinace aktivit
 - Jednotný kmen uživatelů (s potenciálem expanze v rámci ČR)
 - Uživatelská podpora
 - Služby e-INFRA CZ a spolupráce s uživateli



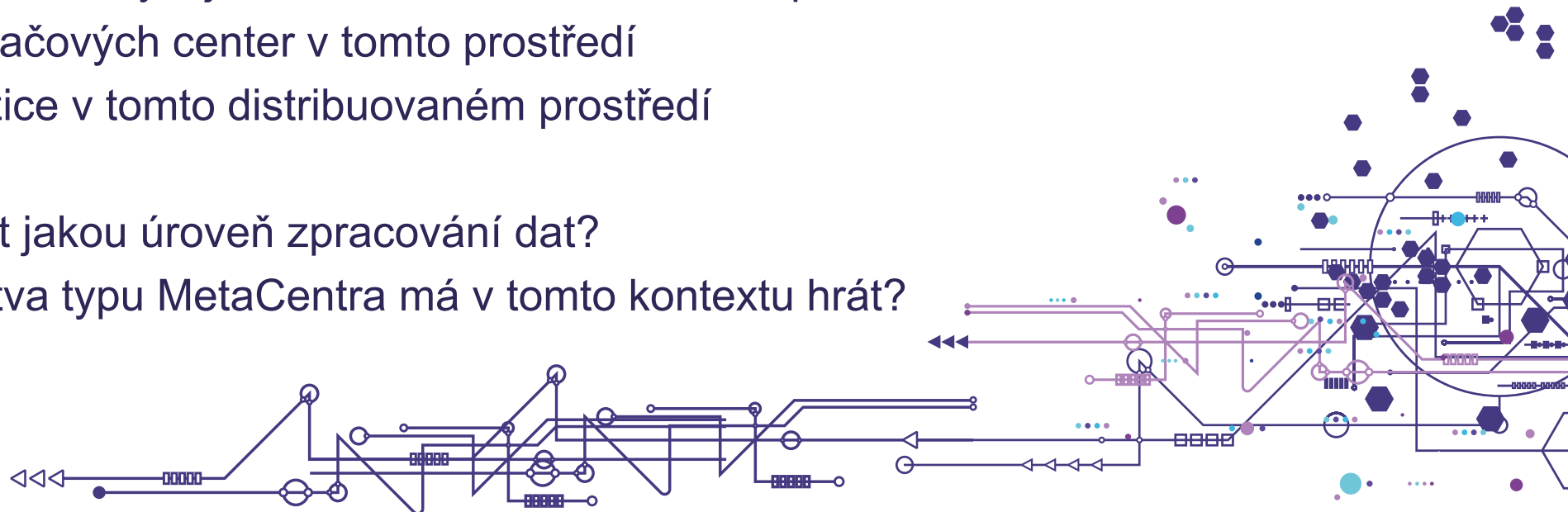
MetaCentrum v rámci e-INFRA CZ



- Distribuovaná výpočetní infrastruktura
- Úložné kapacity především pro „horká“ data
- Nabídka širokého portfolia především generických služeb
- V souvislosti s novým programovým obdobím (2023-2028) diskuse o budoucím strategickém směřování
 - Krácení finančních prostředků
 - Spolupráce se superpočítačovými centry (národní i mezinárodní úroveň)
 - EOSC a orientace na data, strategická spolupráce s oddělením datových úložišť
 - Uživatelská očekávání
- **Jaká je optimální role distribuované výpočetní kapacity v rámci ekosystému e-infrastruktur?**



- Silně se prosazující nová paradigmatata Otevřené vědy, EOSCu, FAIR DAT
- Výrazná orientace ze samotného „compute“ na „data“ a „data processing“
- Současně jasná federovaná distribuovaná infrastruktura
 - Data nebudou nikdy uložena jen na jednom jediném místě
- Zkušenosti z budování a provozu MetaCentra extrémně cenné v tomto kontextu
 - Současně ale další vývoj musí reflektovat změnu úhlu pohledu
- Pozice superpočítačových center v tomto prostředí
 - Optimální pozice v tomto distribuovaném prostředí
- Otevřené otázky
 - Kde realizovat jakou úroveň zpracování dat?
 - Jakou roli vrstva typu MetaCentra má v tomto kontextu hrát?



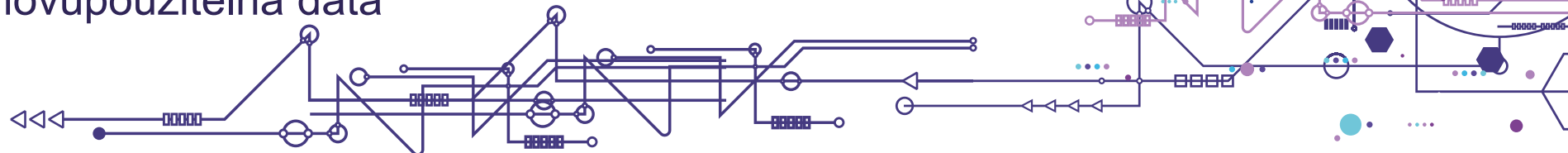
Open Science, EOSC a FAIR Data

- 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

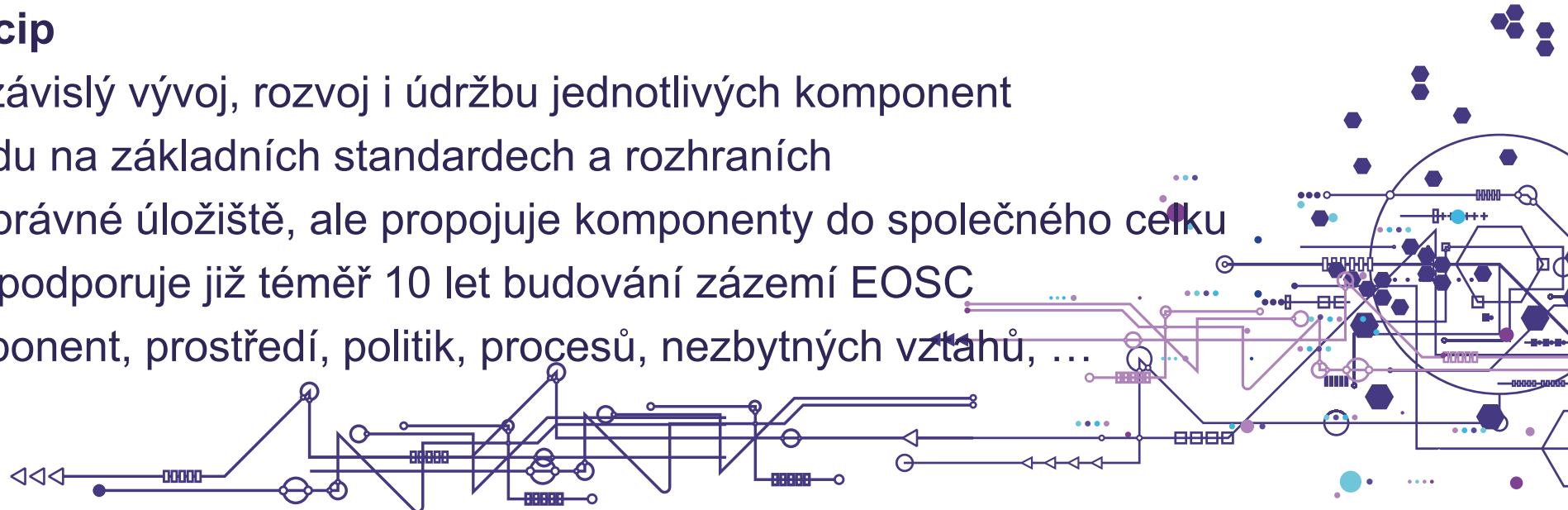
- Data a informace tvoří klíčový prvek konceptu Open Science
 - Data v nejširší možné interpretaci
 - Publikace, výzkumná data, software, workflows, ...
- Důraz na bezbariérový přístup
 - Vyžaduje i schopnost správná data vyhledat a znovupoužít
- IT jako „enabler“
 - Sdílení v uvažovaném rozsahu nereálné bez digitalizace
- Open Science → Open Access, Open Data, Open ... → FAIR Data
 - **F**indable – nalezitelná
 - **A**ccessible – dostupná
 - **I**nteroperable – interoperabilní (metadata)
 - **R**eusable – znovupoužitelná data



European Open Science Cloud (EOSC)

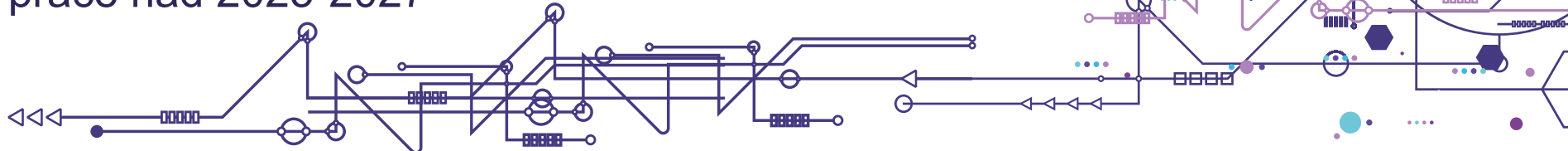


- Technologické, programové i procesní zázemí, které podporuje uspořádanou práci s výzkumnými daty
 - sdružuje a poskytuje **kapacitu**, kde data lze dlouhodobě udržovat
 - poskytuje nástroje pro (snadnou) **správu** dat
 - poskytuje nástroje a prostředí pro **řízení přístupu** k datům
 - zpřístupňuje 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, ...



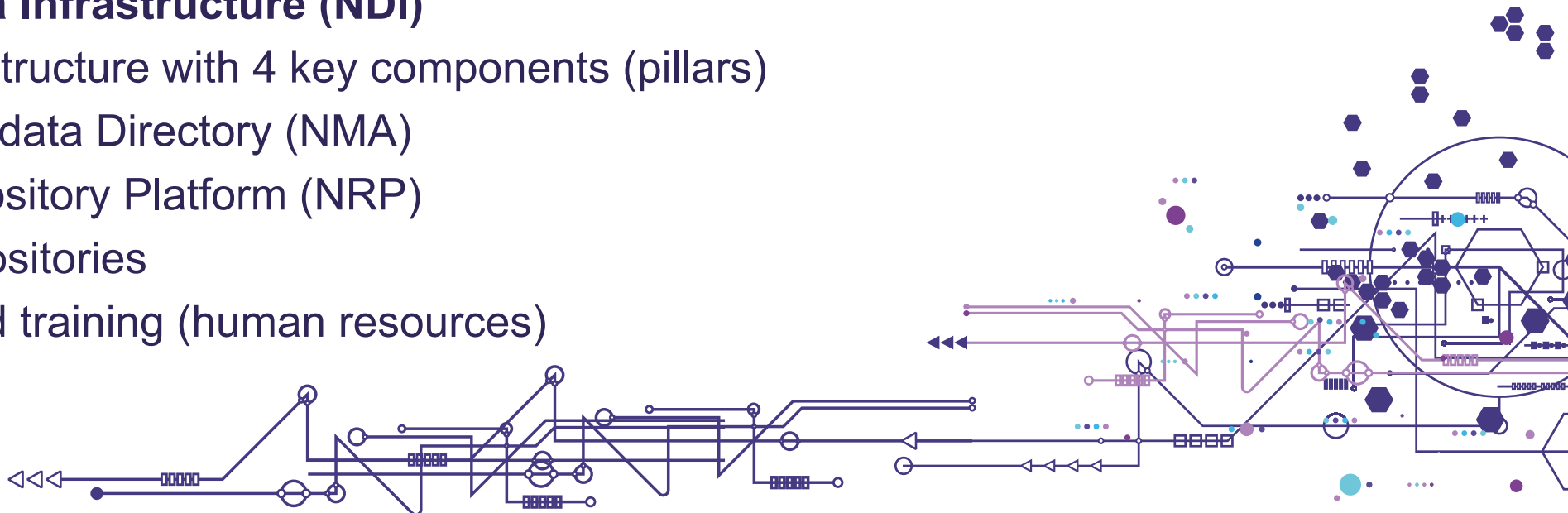
Implementace EOSC v EU

- Dlouhodobý proces, zahájený EK již v roce 2015
- První rozsáhlá fáze 2018-2020
 - Programová podpora H2020 (cca 250 Milionů Euro)
 - Vývoj specifickým komponent, strategické studie
 - Clusterové projekty – zapojení velkých výzkumných infrastruktur (ESFRI)
 - Určitým završením bylo ustavení EOSC Asociace
- 2021 – EOSC Strategic Partnership
 - Jediná horizontální aktivita HEU
- SRIA – Strategic Research and Innovation Agenda
 - Aktuálně verze 1.1, připravuje se verze 2
- MAR – Multi-annual Roadmap (sekce 8 SRIA)
 - Aktuálně se platí 2022-2024
 - Dokončuje se práce nad 2025-2027



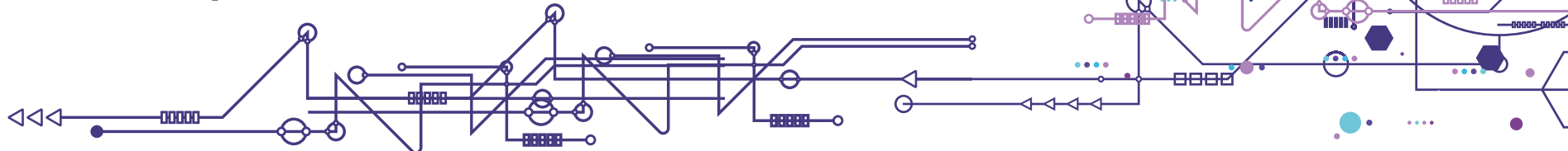
EOSC CZ Principles

- 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)



EOSC Partnerství

- Jediné skutečně horizontální partnerství v rámci HEU
- Příklad EK explicitně podpořit budování EOSC
 - 500 milionů Euro ze strany Komise
- Současně příslib členských států aktivně na implementaci EOSC přispět
 - Dalších 500 milionů Euro v podobě národních příspěvků (ale možné i ERDF)
- Cílem je spojit institucionální, národní i mezinárodní aktivity s cílem výrazně zlepšit situaci v EU v vztahu k ukládání, sdílení, kombinaci a znovuvyužití výzkumných dat napříč vědeckými disciplínami i kontinentem
- Partnerství uzavřela oficiálně **Evropská komise s EOSC Asociací**
 - EOSC Asociace zastupuje instituce
- **EOSC Steering Board** jako řídicí struktura (Partnership Board), tvořená zástupci států
- Tyto tři tělesa tvoří **EOSC Tripartitu**

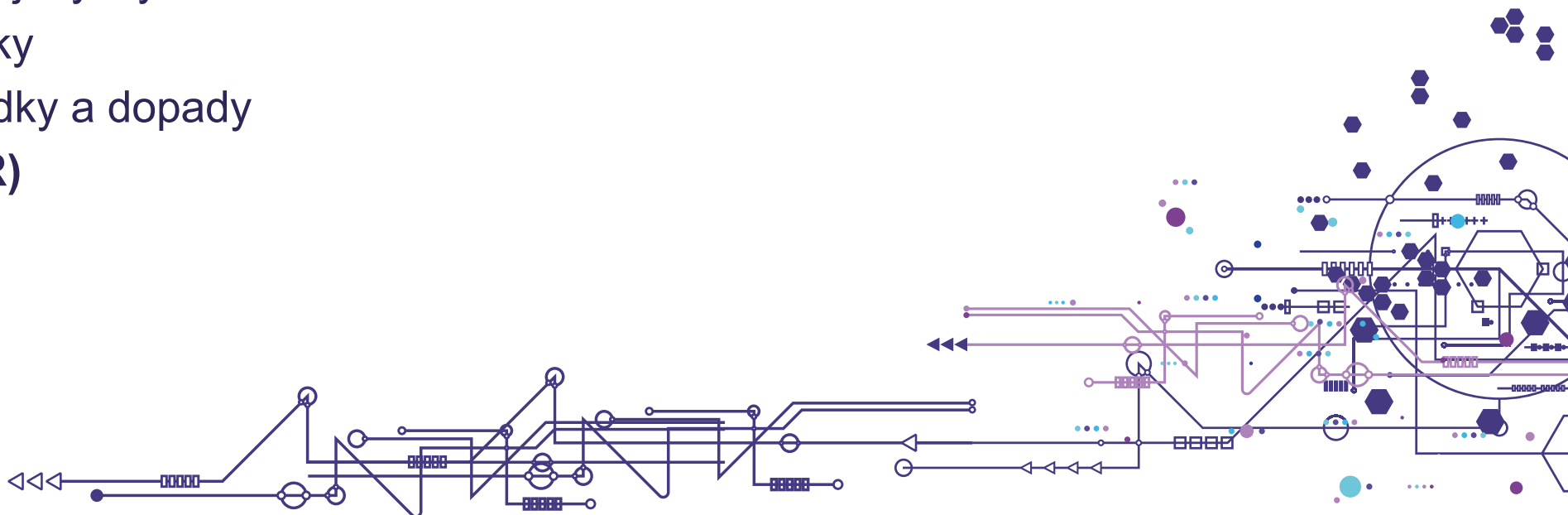


Strategic Research and Innovation Agenda (SRIA)



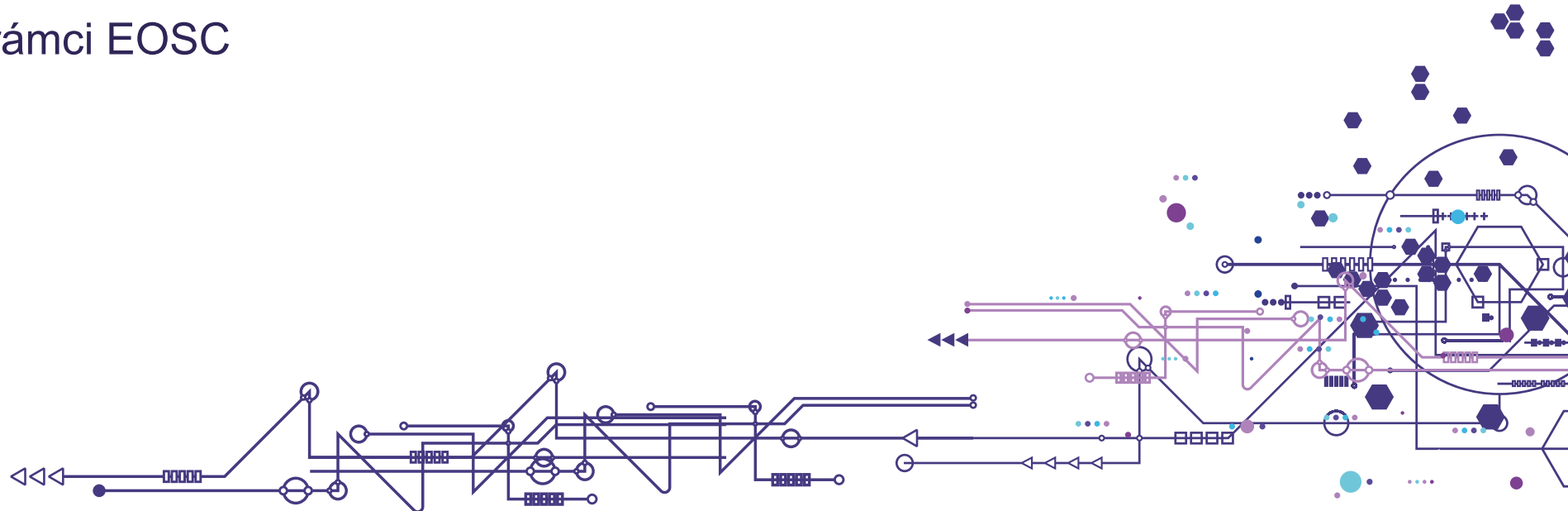
1. Úvod – Otevřená věd, nový způsob rozvoje vědy
2. EOSC a FAIR data jako pilíře Otevřené vědy
3. Strategické cíle EOSC
4. Základní principy
 - Vztah k e-infrastrukturám
5. Implementace a její výzvy
6. Okrajové podmínky
7. Očekávané výsledky a dopady
8. **Roadmapa (MAR)**
9. Závěry

Viz [SRIA 1.1](#)



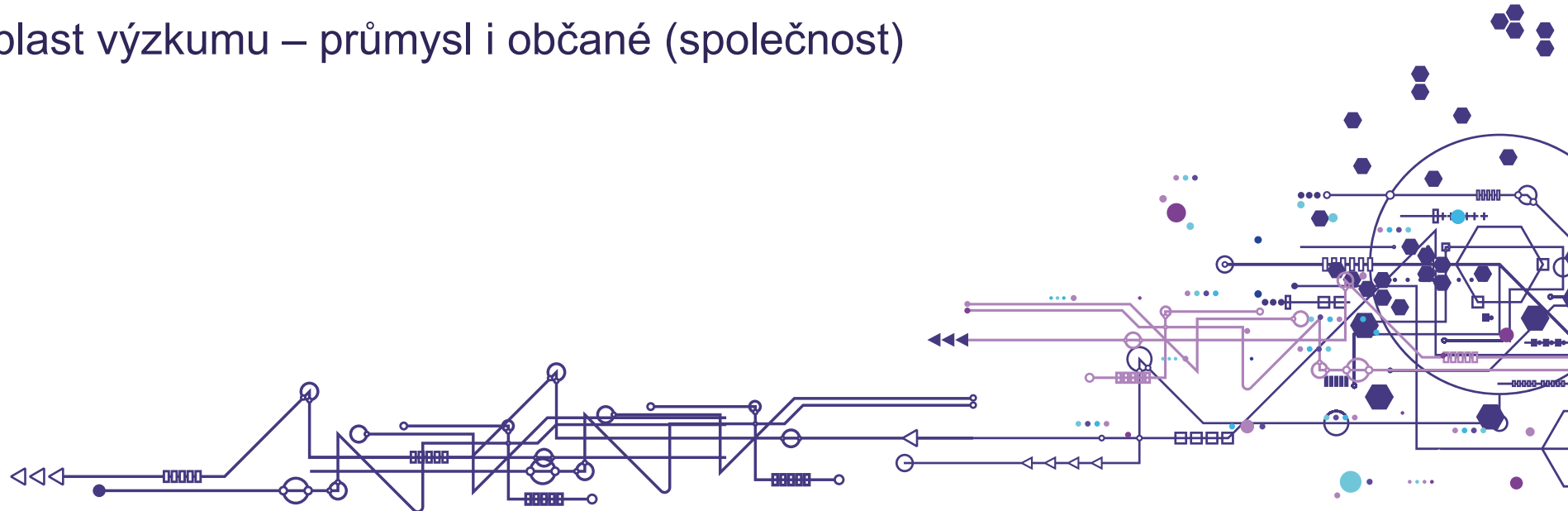
SRIA 5. Implementace EOSC a její výzvy

- Permanentní identifikátory
- Metadata a ontologie
- FAIR metriky a certifikace
- Autentizační a autorizační infrastruktura
- Uživatelské prostředí
- Zdroje (kapacity) a způsob jejich poskytování
- Interoperabilita v rámci EOSC



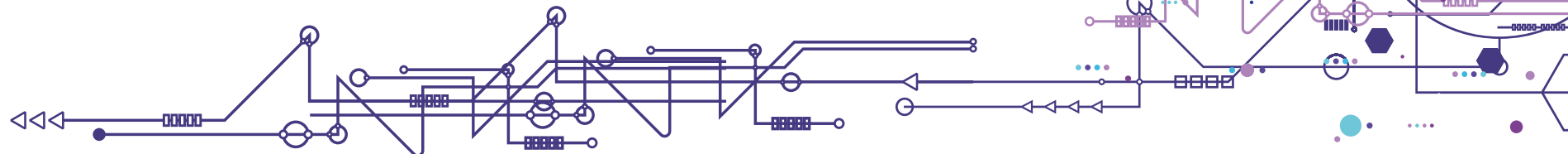
SRIA 6. Okrajové podmínky

- Pravidla účasti
- Landscaping a monitoring
- Byznys modely
- Skily a školení
- Ocenění a hodnocení
- Komunikace/publicita
- Rozšíření mimo oblast výzkumu – průmysl i občané (společnost)



Pracovní skupiny EOSC Asociace

- Odpovídají příspěvkům pro SRIA (klíčové oblasti)
- 1. Metadata a kvalita data
 - FAIR metriky a kvalita dat
 - Semantická interoperabilita
 - PID: Politiky a implementace
- 2. Kariéra ve vědě a kurikula
 - Datové kurátorství, kurikula a kariérní cesty
 - Kariéry ve výzkumu, uznání a kredity (za OS)
 - Podpora (upskilling) zemí při zapojení do rozvoje EOSCů
 - Zapojení vědců, přijetí principů
- 3. Technické výzvy
 - Architektura Autentizační a autorizační infrastruktury
 - Infrastruktura pro kvalitní výzkumný software
 - Technická interoperabilita dat a služeb
 - Dlouhodobé ukládání dat
- 4. Dlouhodobá udržitelnost EOSC
 - Finanční udržitelnost
 - Monitoring dodržování pravidel přístupu



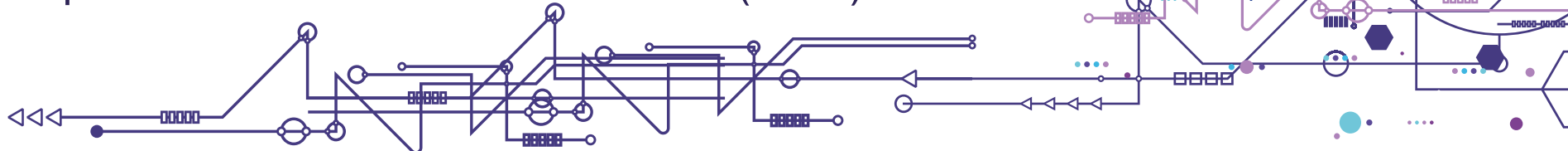
Principy implementace EOSC v ČR

- 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
- Základní dokument: **Architektura implementace EOSC v ČR**



Součásti implementace EOSC v ČR

- Pracovní skupiny
 - Otevřená platforma
 - Diskuse základních principů (implementace) EOSC
 - co vědecká komunita potřebuje a očekává
 - a jak tato očekávání a potřeby nejlépe naplnit
- 4 průřezové pracovní skupiny
 - Architektura, Základní služby, Metadata, vzdělávání a lidské zdroje
- 7+1 tematické pracovní skupiny
 - Bio/Med/Food, ENVRI, Physics, Material Sciences, AI, Humanities, Social Sciences
 - Sensitive Data
- Pracovní skupiny EOSC CZ představují primární prostředí pro aktivní (i pasivní) zapojení do implementace EOSC v ČR nezávisle na zapojení do konkrétního projektu
- Existence minimálně po celou dobu intervence OP JAK (2028)



Národní datová infrastruktura (NDI)

Národní metadatový adresář (NMA)

Národní repozitářová platforma (NRP)

Služby NDI

Tematické repozitáře

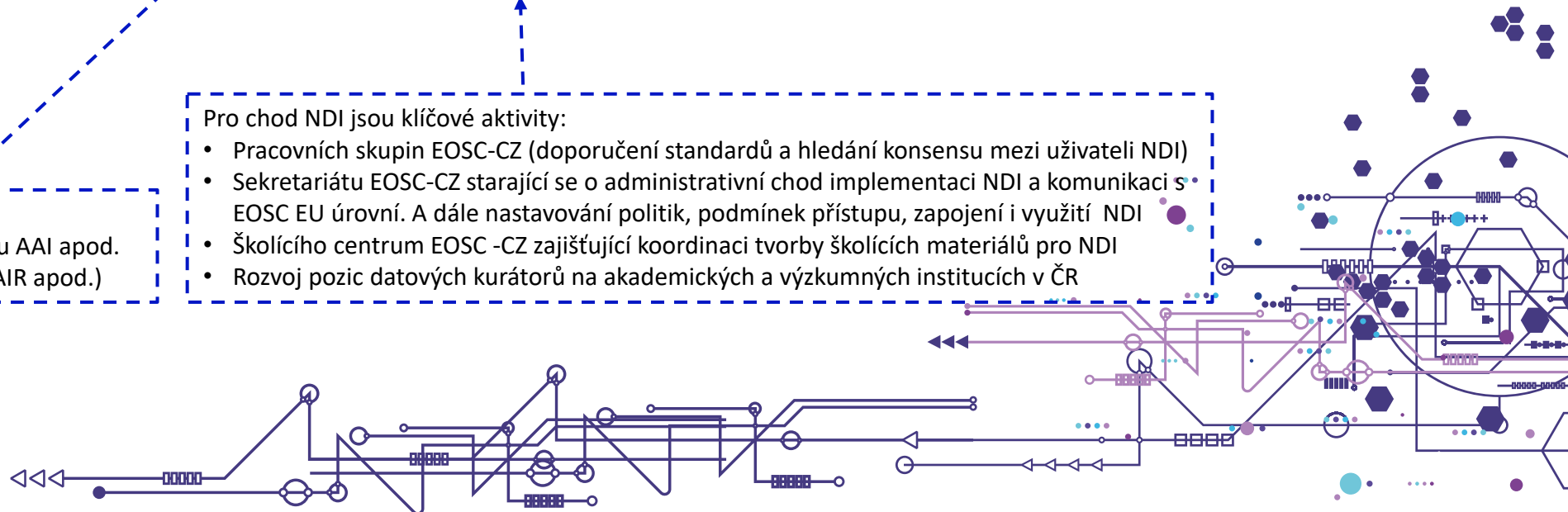
Ostatní repozitáře

Služby NDI jsou dvojího typu:

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

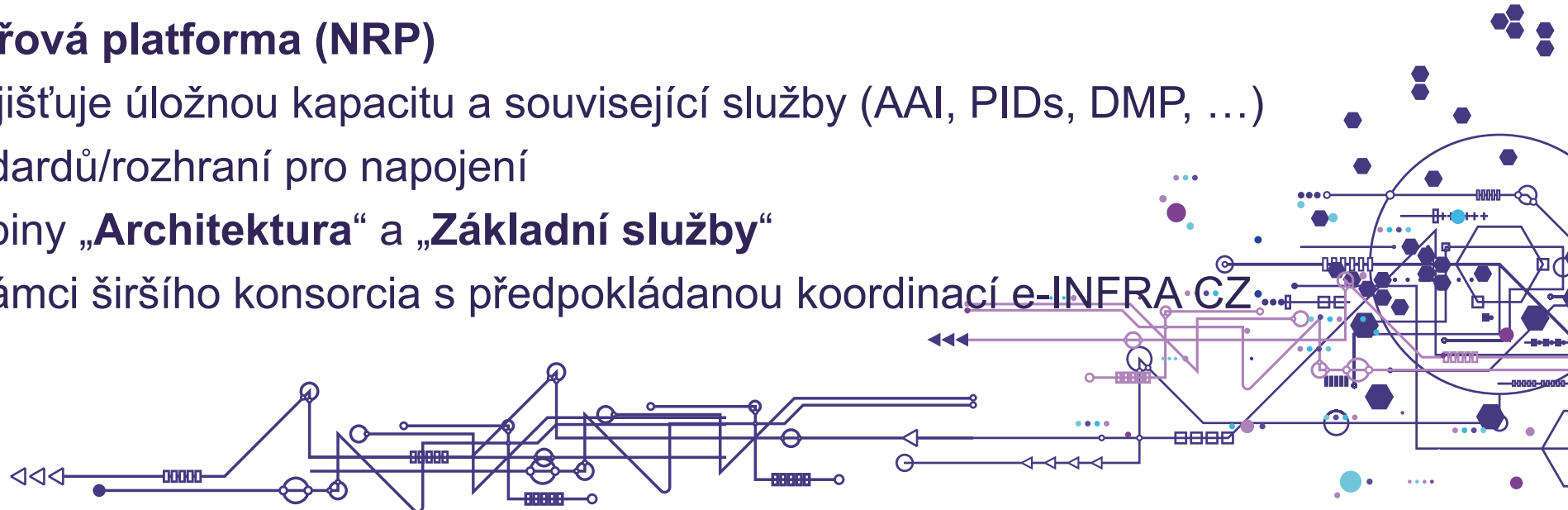
Pro chod NDI jsou klíčové aktivity:

- 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



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)



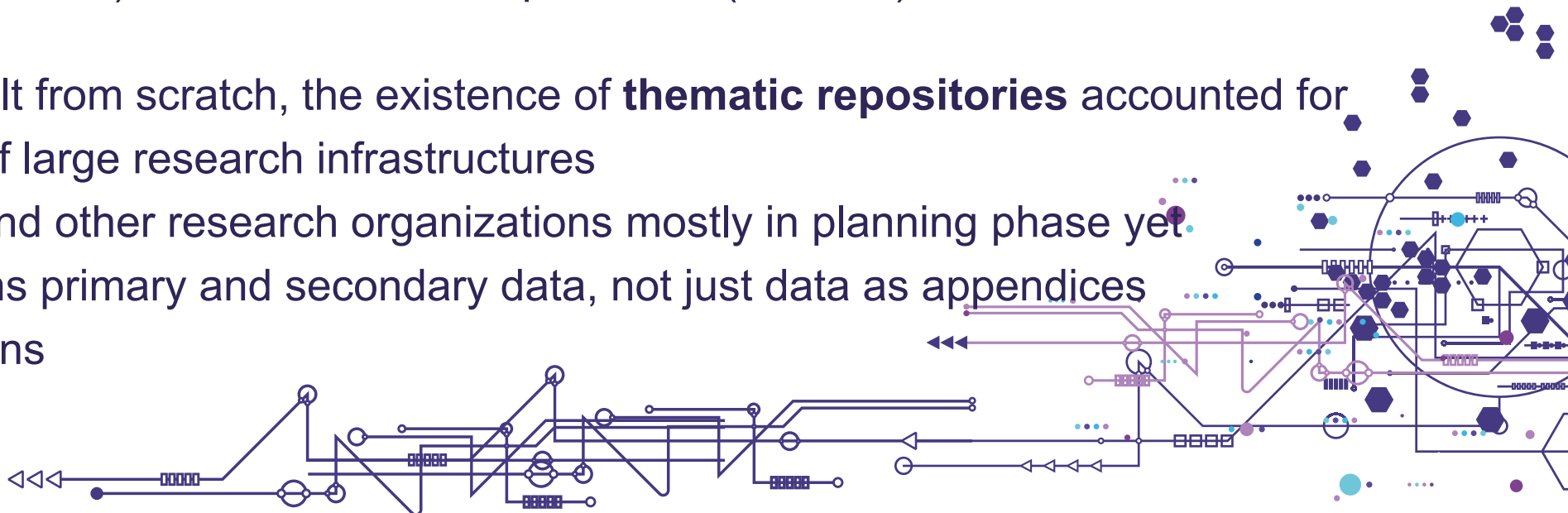
Další součásti

- **Lidské zdroje**
 - Průřezová (horizontální) aktivita
 - Pracovní skupina „**Vzdělávání a lidské zdroje**“
 - Podpora v rámci prakticky všech zvažovaných projektů
- **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**



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



OP JAK a implementace EOSC v ČR

Výzvy Open Science

1. Zázemí realizace

- Sekretariát EOSC CZ, práce s metadata a NMA, PIDs, ...
- IPs EOSC CZ a CARDS; výzva vypsána

2. Národní repozitářová platforma a její služby

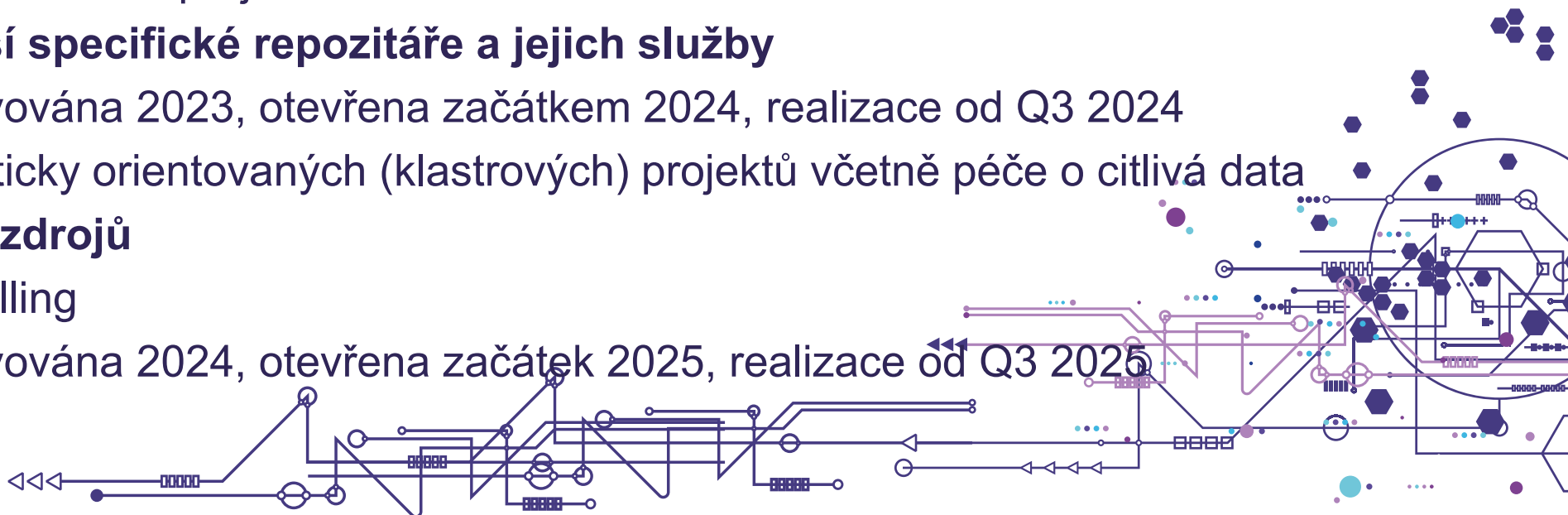
- Výzva v přípravě, otevření začátek 2023, realizace od Q3 2023
- Společný konsorciální projekt NRP

3. Tematické a další specifické repozitáře a jejich služby

- Výzva připravována 2023, otevřena začátkem 2024, realizace od Q3 2024
- Několik tematicky orientovaných (klastrových) projektů včetně péče o citlivá data

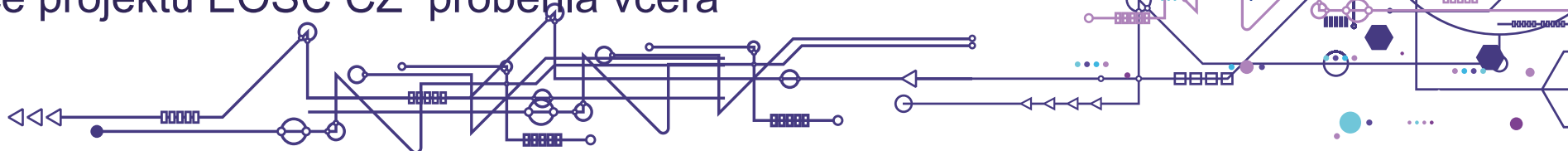
4. Rozvoj lidských zdrojů

- Uptake/Upskilling
- Výzva připravována 2024, otevřena začátek 2025, realizace od Q3 2025



Projekty IPs EOSC CZ (MU/e-INFRA CZ) a CARDS (NTK/e-INFRA CZ)

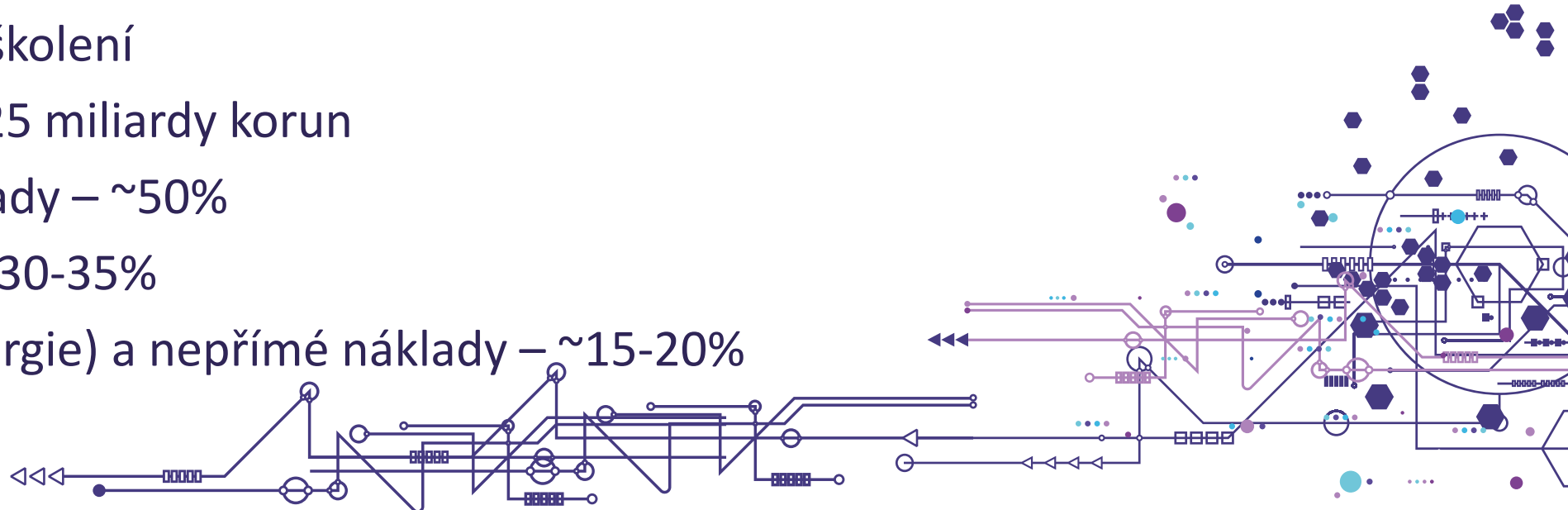
- Odpovědnost za zázemí implementace EOSC v ČR
- Organizační a metodické zázemí (primárně MU a VŠB)
 - Sekretariát EOSC CZ, vazba na Open Science at large, ... (MU)
 - Portál a informační hub pro EOSC
 - Organizace školení a vzdělávání (VŠB)
 - Metodická podpora pro metadata (ontologie) a PIDs (CARDS/NTK)
- Technické zázemí (primárně CESNET)
 - Národní metadatový adresář
 - AAI, datové přenosy, ...
- CARDS širší záběr (knihovní systém)
 - Metodická péče o Metadata a PIDs (národní centrum), OS zastřešení
- Veřejná prezentace projektu EOSC CZ proběhla včera



Společný projekt Národní repozitářové platformy

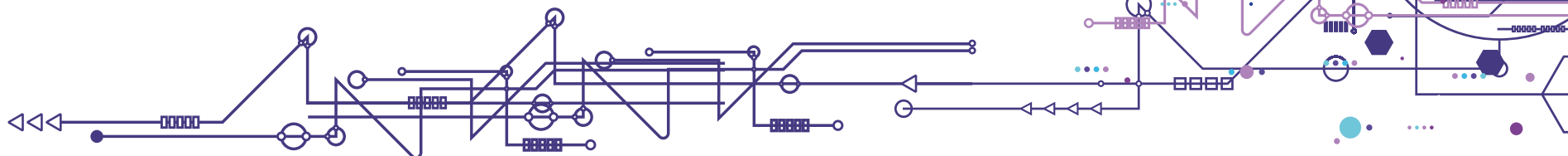
Základy přípravy projektu

- Konsorciální projekt, cca 10+ partnerů
 - Koordinace Pověřenou organizací EOSC Asociace (CESNET)
- Povinná spolupráce s IPs EOSC CZ a CARDS
 - Napojení na NMA
 - Dodržování standardů
 - Koordinace školení
- Rozpočet cca 1,25 miliardy korun
 - Osobní náklady – ~50%
 - Investice – ~30-35%
 - Služby (+energie) a nepřímé náklady – ~15-20%



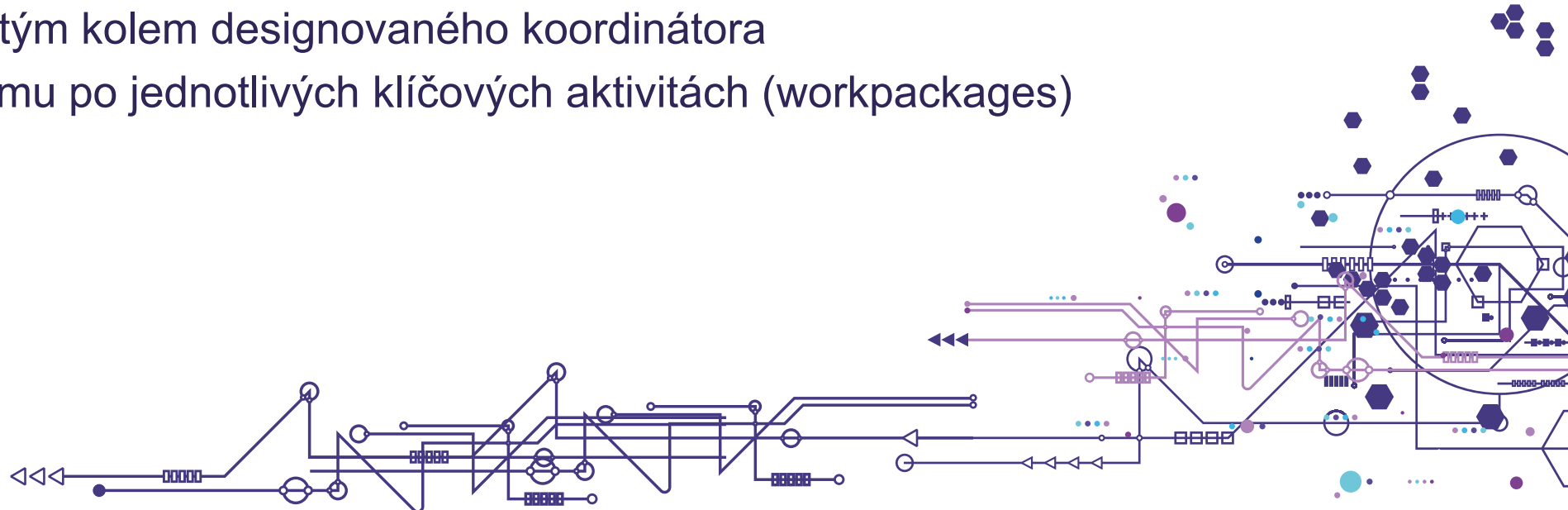
Projekt NRP – Očekávané výstupy

1. **Architektura Národní repozitářové platformy (NRP)** jako podklad pro její budování.
2. Kontinuálně rozvíjený **soubor služeb** s roadmapou jejich postupné implementace a zavádění do provozu.
3. **Národní repozitářová platforma** se základními i pokročilými službami, včetně odpovídající distribuované úložné kapacity a systému řízení přístupu k datům v NRP.
4. **Integrace klíčových aktérů datově-orientovaných akademických institucí a výzkumných týmů pro FAIR data** při budování jádra Národní datové Infrastruktury (NDI).
5. **Dlouhodobý provoz NRP a jejích služeb, včetně systematické podpory vědecké komunity v oblasti vzdělávání** o best-practices práce s výzkumnými daty a využití služeb NRP pro vědecká data.
6. **Školící materiály a školení pro práci s NRP a jejími službami** ve spolupráci s IPs EOSC CZ při realizaci školení.
7. **Tvorba konkrétních tenantů/repozitářů nad NRP**, jako jsou např. institucionální repozitáře, tam, kde si je instituce budou chtít vytvářet; **vytváření vzorů (templates) a postupů** pro vytváření specifických repozitářů podle potřeby výzkumných skupin a institucí.



Příprava projektu NRP

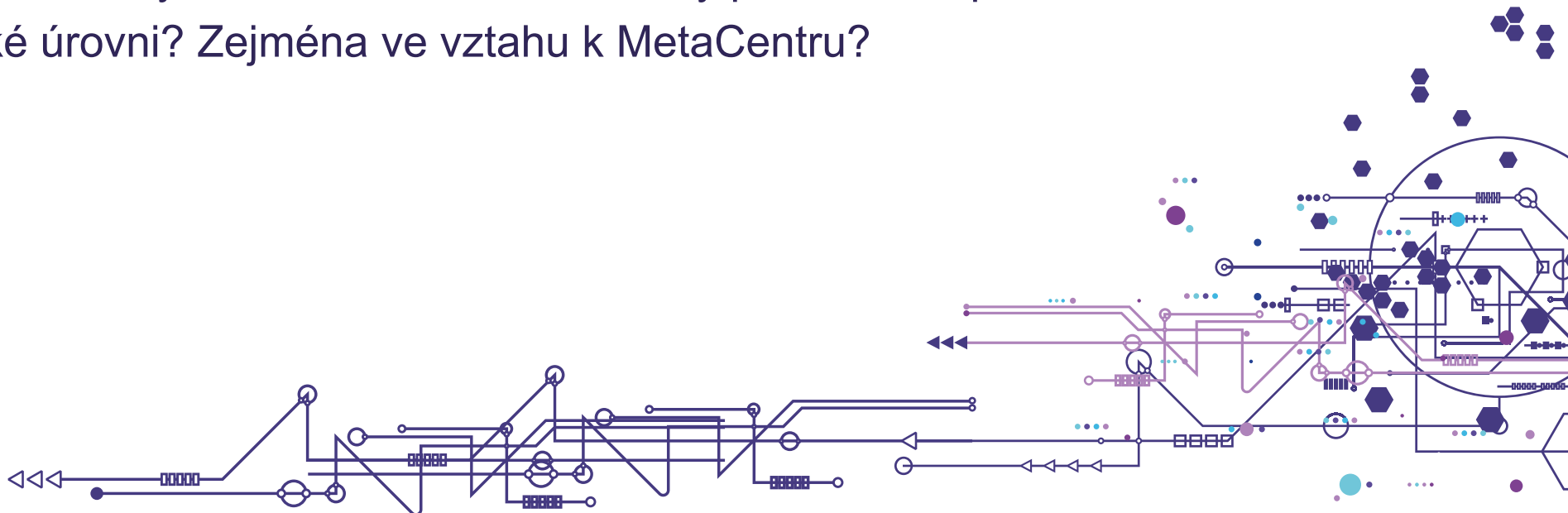
- Projekt aktuálně v přípravě
 - Výzva OP JAK Open Science I, uzávěrka konec roku 2023
- Přípravný výbor
 - Oficiální zástupci celkem 16 institucí (Asociace výzkumných univerzit, e-INFRA CZ, ZČU, AV ČR zastoupená 6 organizacemi, NTK)
- Ediční týmu
 - Hlavní ediční tým kolem designovaného koordinátora
 - Dílčí ediční týmu po jednotlivých klíčových aktivitách (workpackages)



EOSC a MetaCentrum/e-INFRA CZ



- Národní leader v oblasti implementace EOSC CZ
 - Koordinace obou klíčových projektů EOSC CZ (MU) a NRP (CESNET)
 - Spolupráce při definici strategických záměrů (ekvivalent SRIA MAR na národní úrovni)
 - Mezinárodní spolupráce, koordinace zapojení do aktivit EOSC Asociace
 - Koordinace klíčových průřezových pracovních skupin EOSC CZ
 - Primární zázemí, analytické a monitorovací aktivity pro celou implementaci EOSC CZ
- Ale co na technické úrovni? Zejména ve vztahu k MetaCentru?



EOSC a MetaCentrum/e-INFRA CZ

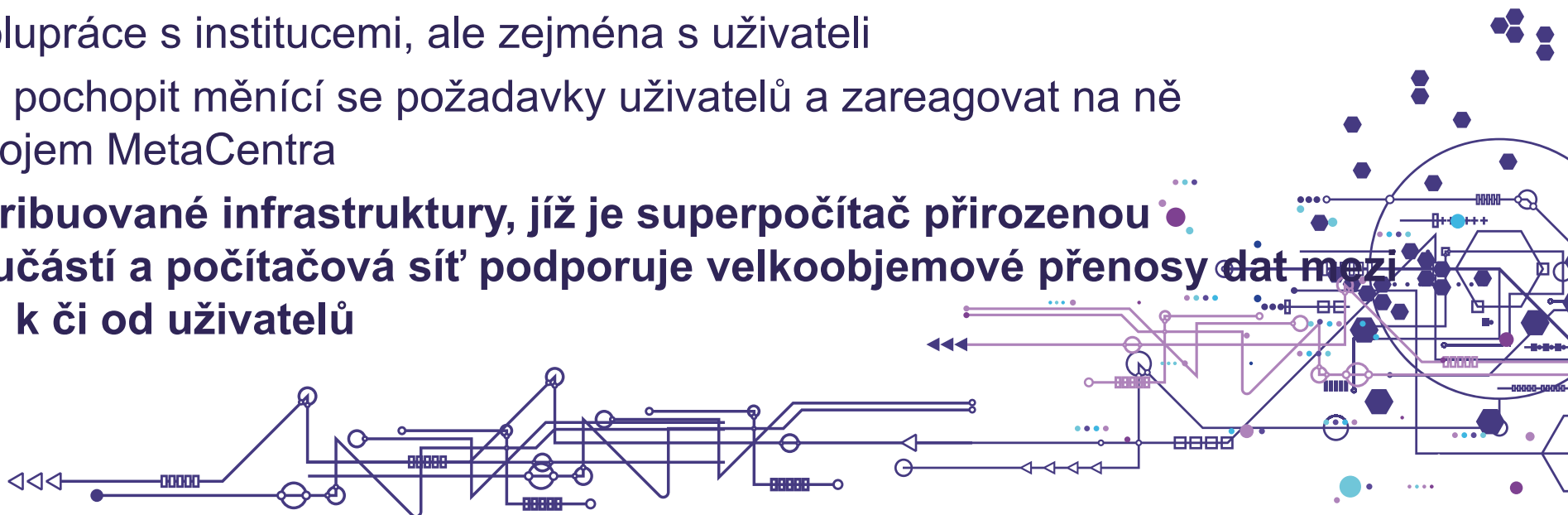


- NRP bude konstruován jako distribuovaná federovaná infrastruktura
 - Společné e-INFRA CZ AAI
 - Help desk
- Data vyžadují zpracovatelskou aktivitu
 - Synergický související rozvoj MetaCentra
 - Začlenění superpočítače
 - Velkoobjemové přenosy dat
- Dávkové versus interaktivní zpracování
- Spolupráce s datovými úložišti
 - Přímá interakce/odpovědnost za implementaci EOSCu
- EOSC předpokládá vývoj/nasazení specifických služeb
 - Často ale vyžadují výpočetní kapacitu, která je blízko dat



Summary

- Implementace EOSCu představuje příležitost rozvoje celé e-INFRA CZ a specificky MetaCentra
- Měníme úhel pohledu – od primárně výpočetního přístupu do datově orientovaného
 - EOSC CZ vytvoří infrastrukturu pro velká FAIR data
 - MetaCentrum e-INFRA CZ musí dodat odpovídající prostředí pro zpracování dat
- Na druhé straně e-INFRA CZ nemůže vybudovat EOSC CZ sama
 - Klíčová je spolupráce s institucemi, ale zejména s uživateli
- Nezbytné správně pochopit měnící se požadavky uživatelů a zareagovat na ně odpovídajícím vývojem MetaCentra
- **Posílení role distribuované infrastruktury, již je superpočítač přirozenou integrovanou součástí a počítačová síť podporuje velkoobjemové přenosy dat mezi úložišti i směrem k či od uživatelů**





Thanks and questions?

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

More information on

<https://www.e-infra.cz>

<https://www.eosc.cz>

e-infra.cz

