

EOSC CZ

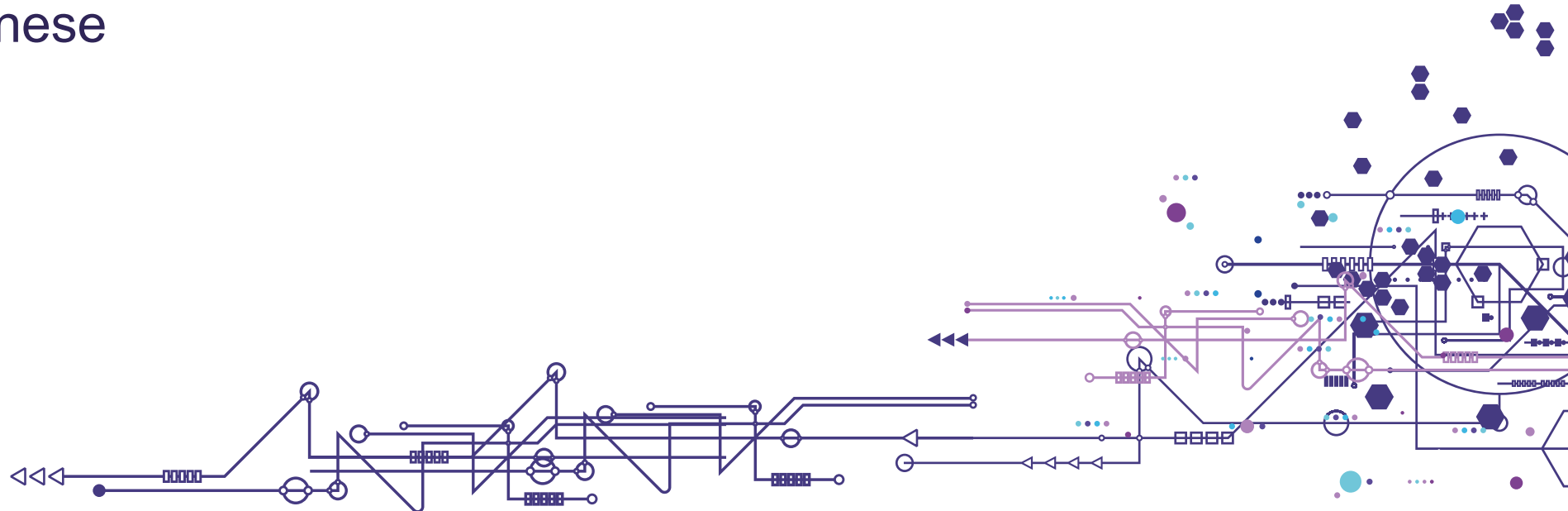
Stav přípravy národní datové infrastruktury

Luděk Matyska

CERIT-SC/MU (e-INFRA CZ) a ELIXIR CZ

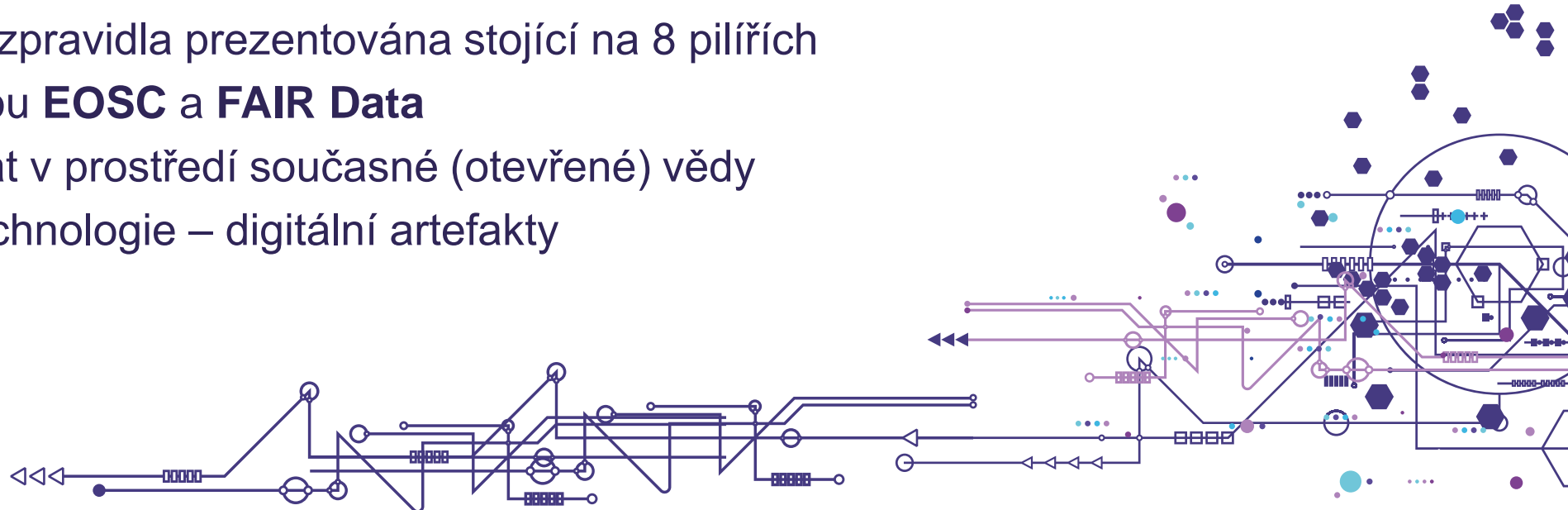
Obsah přednášky

- 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 a EOSC

- 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)
 - Spolupráce, sdílení, ...
- Otevřená věda je zpravidla prezentována stojící na 8 pilířích
 - Dva z nich jsou **EOSC** a **FAIR Data**
- Odráží význam dat v prostředí současné (otevřené) vědy
 - Informační technologie – digitální artefakty



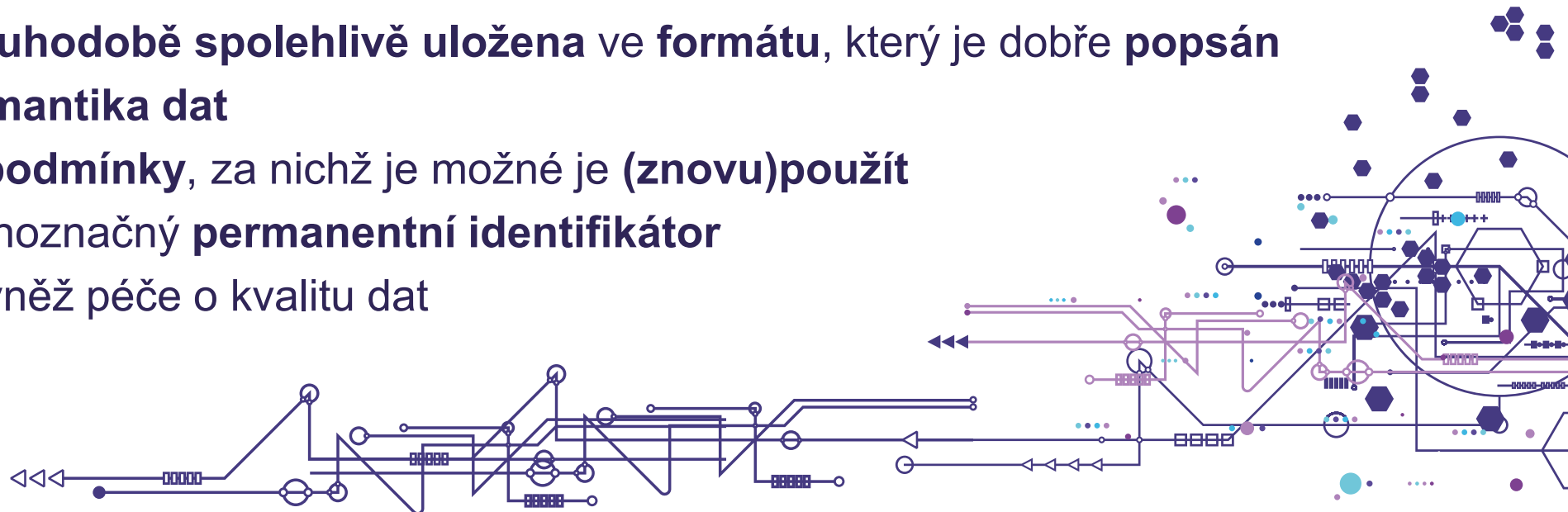
European Open Science Cloud (EOSC)



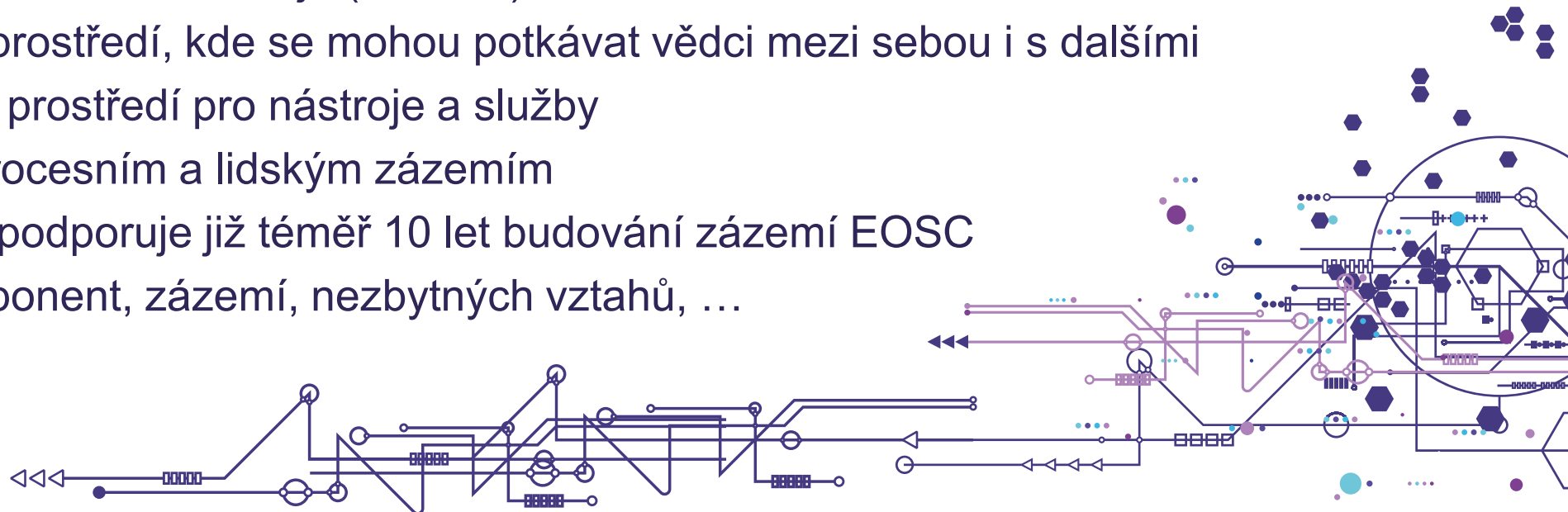
- Technologické, programové i procesní zázemí, které podporuje uspořádanou práci s výzkumnými daty
 - poskytuje kapacitu, kde data lze dlouhodobě udržovat
 - poskytuje nástroje pro (snadnou) správu dat
 - poskytuje nástroje a prostředí pro řízení přístupu k datům
 - zpřístupňuje nástroje a prostředí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



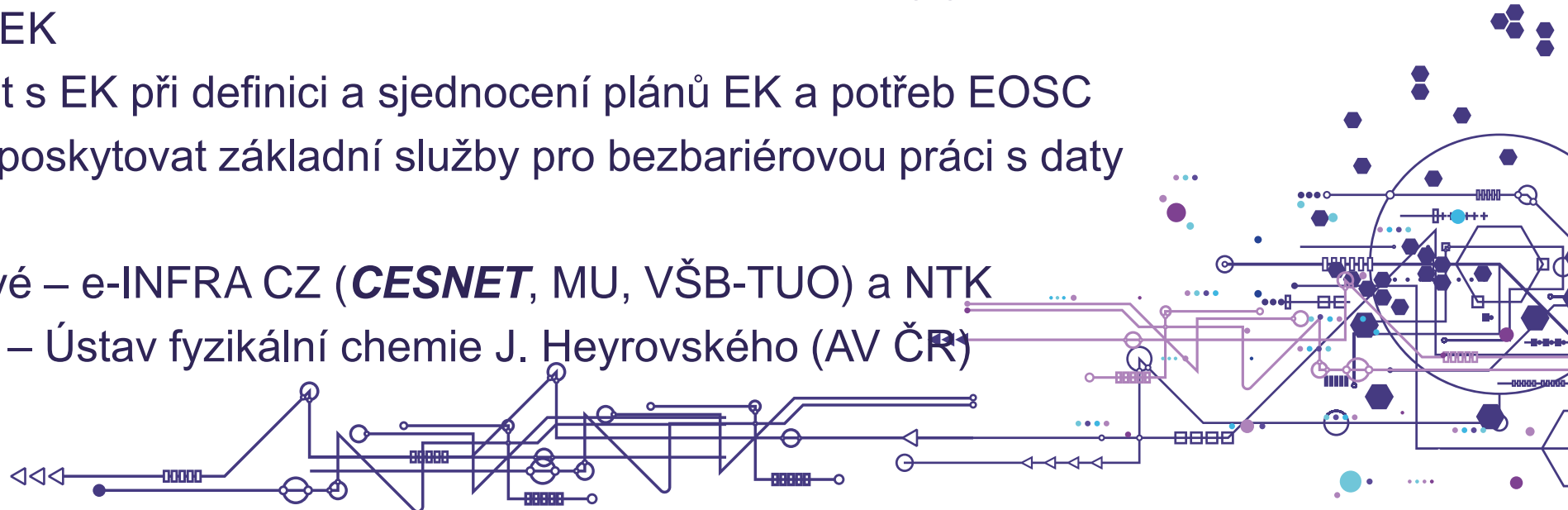
- Původní radikální koncept „Otevřených dat“ transformován do principu FAIR data
 - **F**indable – nalezitelná
 - **A**ccessible – dostupná
 - **I**nteroperable – interoperabilní (metadata)
 - **R**eusable – znovupoužitelná data
- Tedy
 - Data jsou **dlouhodobě spolehlivě uložena** ve **formátu**, který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



- EOSC má zajistit prostředí, v němž je možné FAIR data vytvářet, kultivovat a dále využívat
- EOSC = Federace FAIR dat a souvisejících služeb
- Federace rovněž na úrovni států
 - Základem prostředí pro péči o (FAIR) výzkumná data
 - Současné i budoucí zdroje (úložiště) dat
 - Virtuální prostředí, kde se mohou potkávat vědci mezi sebou i s dalšími
 - Otevřené prostředí pro nástroje a služby
 - Společně s procesním a lidským zázemím
- Evropská komise podporuje již téměř 10 let budování zázemí EOSC
 - Definice komponent, zázemí, nezbytných vztahů, ...



- 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)

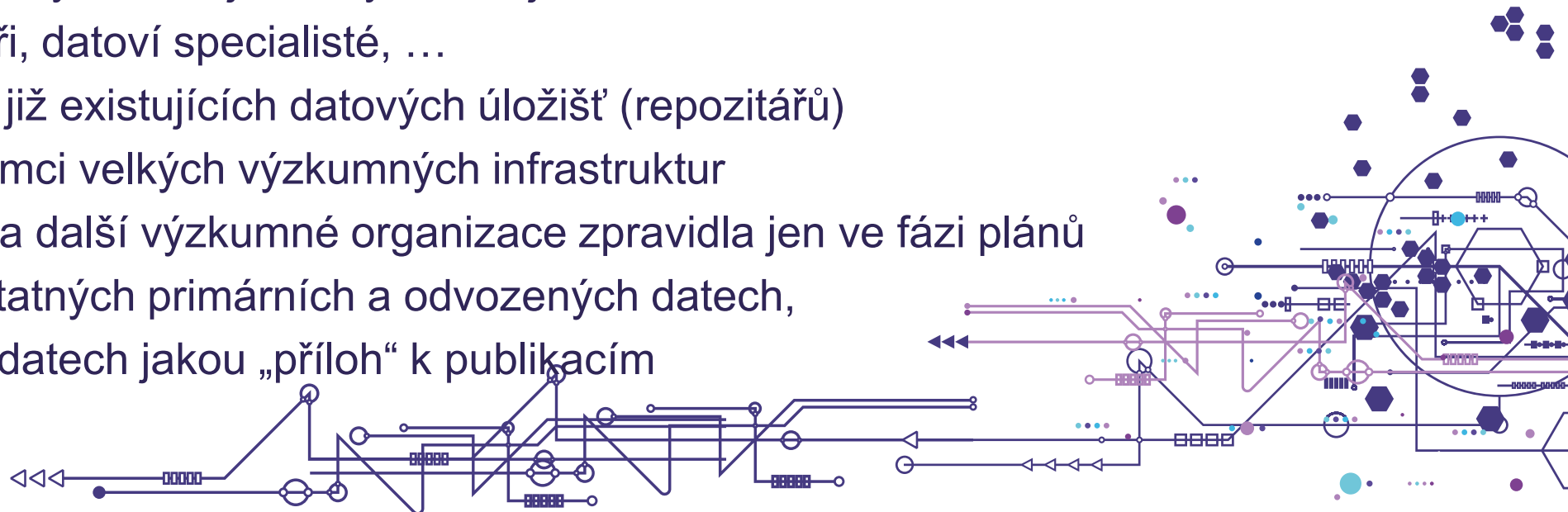


- Architektura implementace EOSC v ČR
 - Dokument vytvořen v rámci práce EOSC platformy v roce 2021
 - Definuje základní principy implementace EOSC v ČR
 - Národní datová infrastruktura, její komponenty a lidské zdroje
- Koordinační výbor EOSC
 - Ustaven na přelomu roku 2021/2022
 - Poradní orgán MŠMT (předsedají nám. Wildová a Velčovský)
- Plánována 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



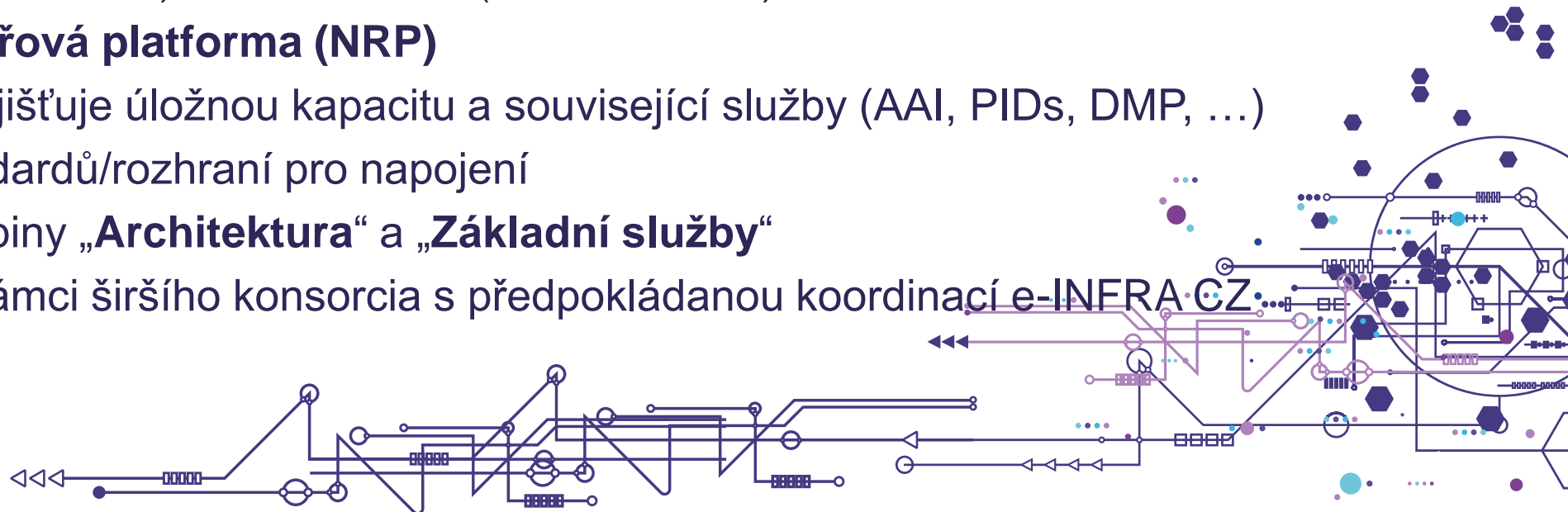
Principy implementace EOSC v ČR

- EOSC 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
 - Datoví kurátoři, datoví specialisté, ...
- 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



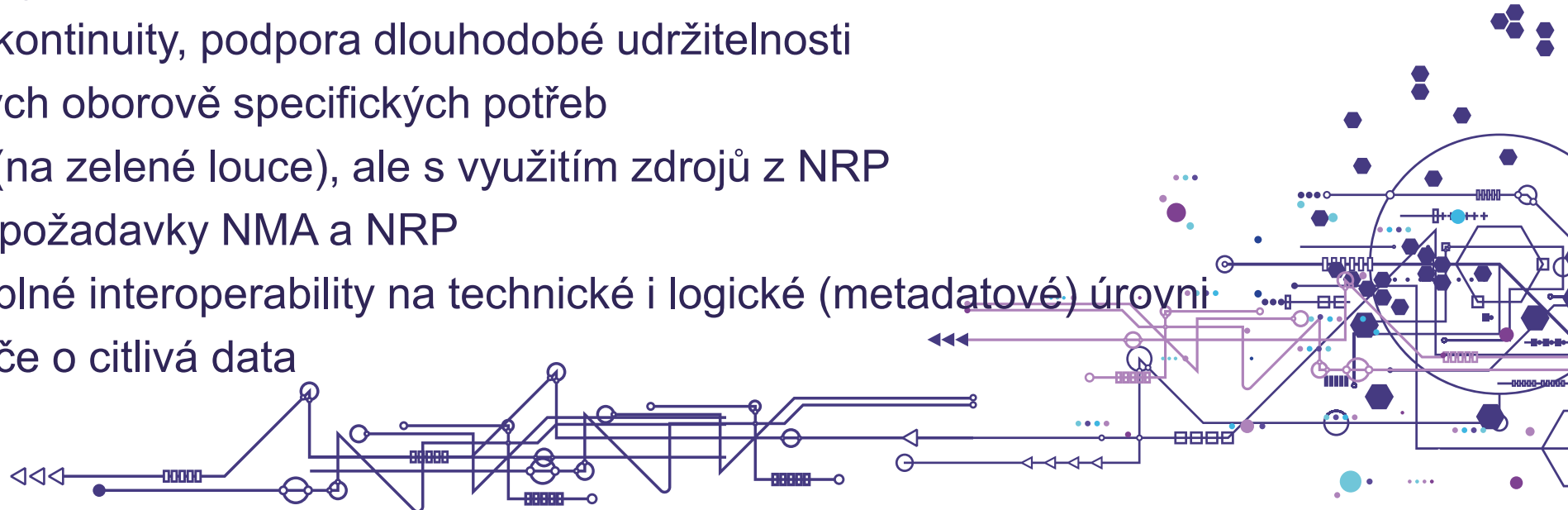
Klíčové komponenty

- **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)

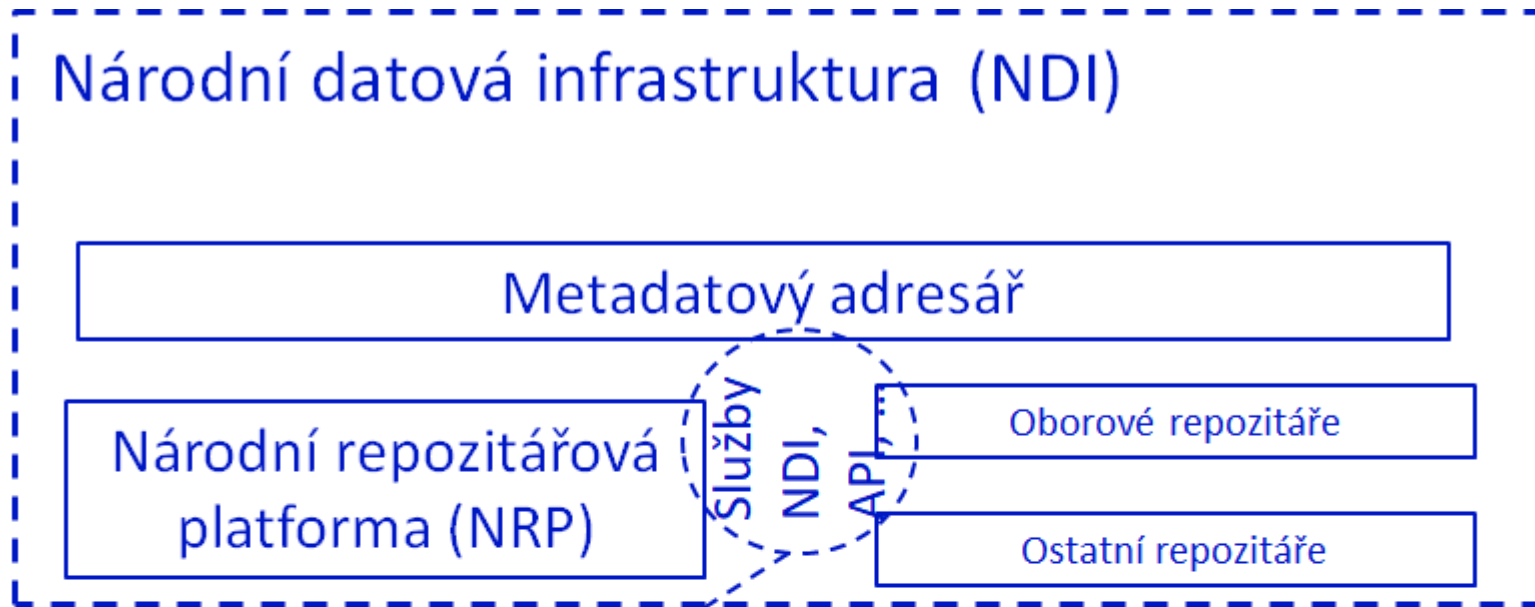


Další součásti

- **Lidské zdroje**
 - Průřezová (horizontální) aktivita
 - Pracovní skupina „**Vzdělávání**“
 - Podpora v rámci prakticky všech zvažovaných projektů
- **Existující oborově-vědní repozitáře**
 - Podpora existujících aktivit a repozitářů
 - Zajištění kontinuity, podpora dlouhodobé udržitelnosti
 - Podpora nových oborově specifických potřeb
 - Ne nové (na zelené louce), ale s využitím zdrojů z 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)

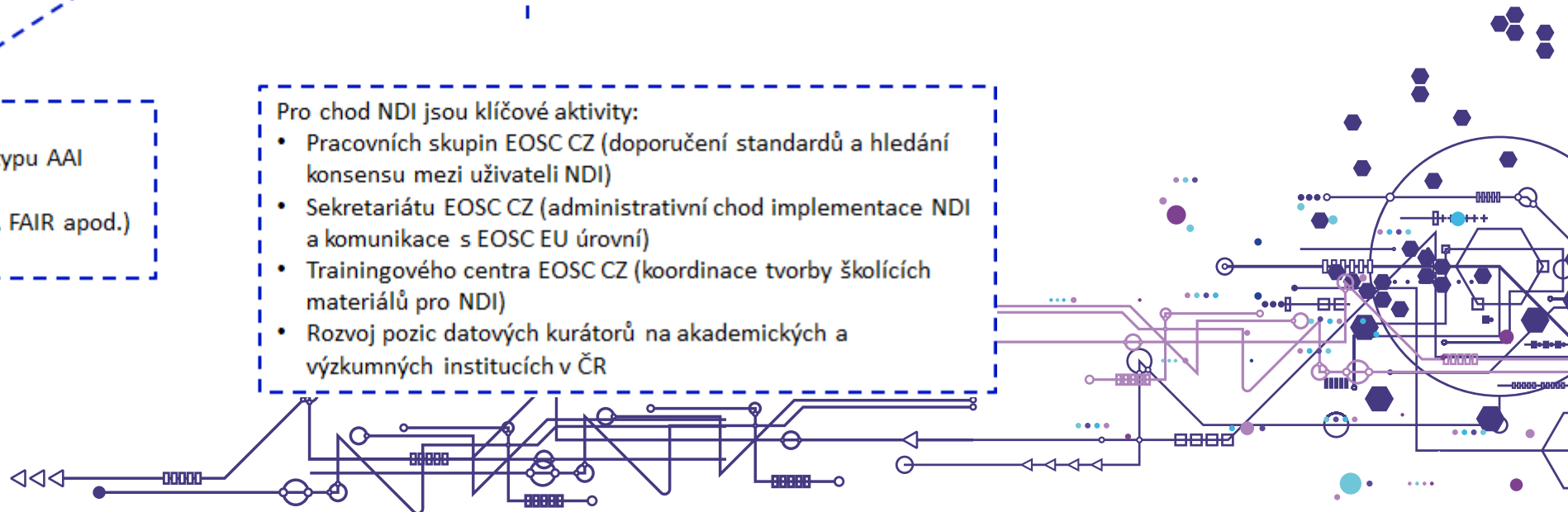


Služby NDI jsou dvojího typu:

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

Pro chod NDI jsou klíčové aktivity:

- Pracovních skupin EOSC CZ (doporučení standardů a hledání konsensu mezi uživateli NDI)
- Sekretariátu EOSC CZ (administrativní chod implementace NDI a komunikace s EOSC EU úrovní)
- Trainingového centra EOSC CZ (koordinace tvorby školících materiálů pro NDI)
- Rozvoj pozic datových kurátorů na akademických a výzkumných institucích v ČR

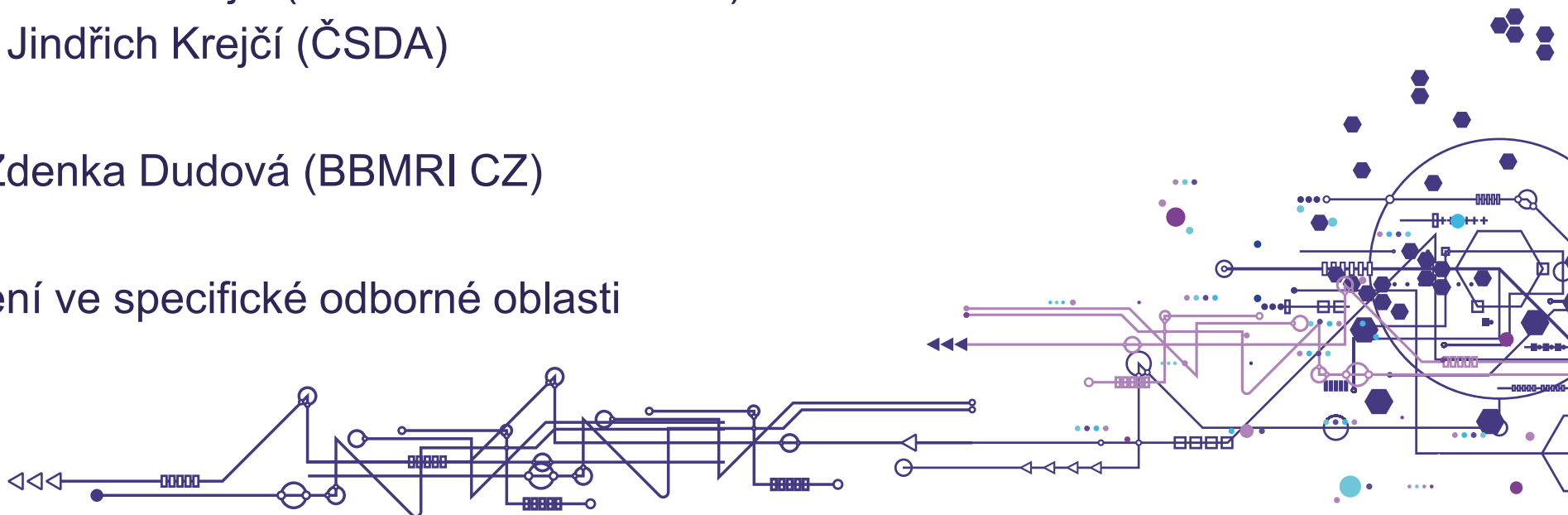


Oborově-vědní pracovní skupiny

Aktuálně 8 ustavovaných

- 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)

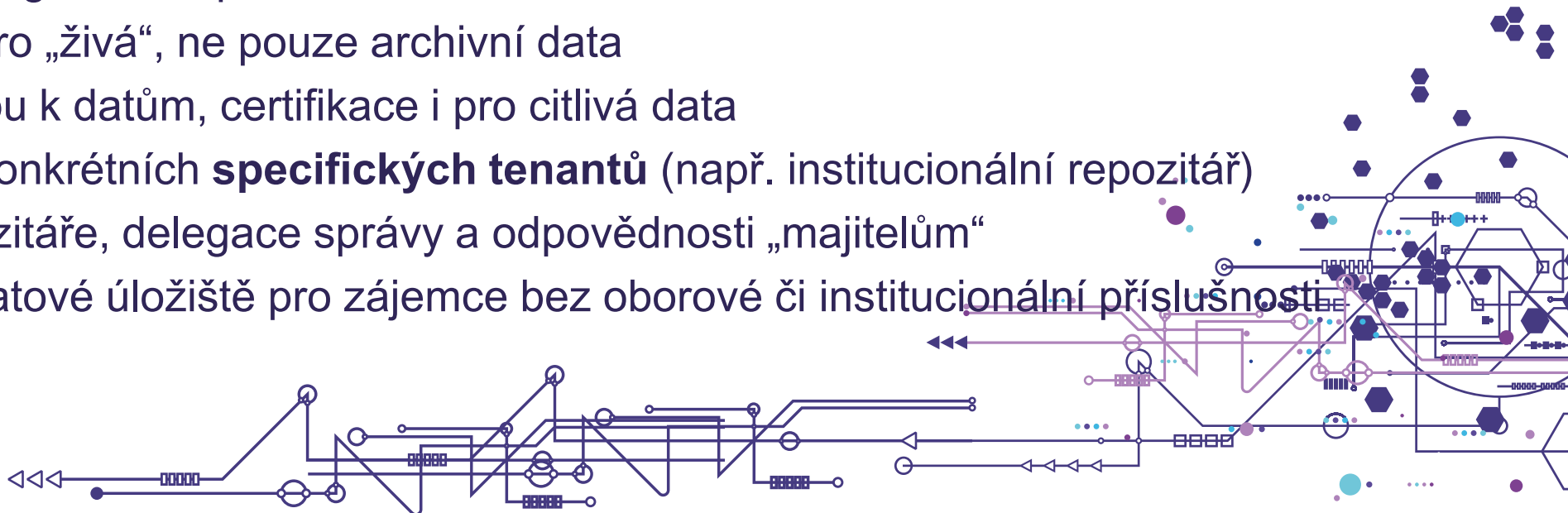
- Prostor pro zapojení ve specifické odborné oblasti



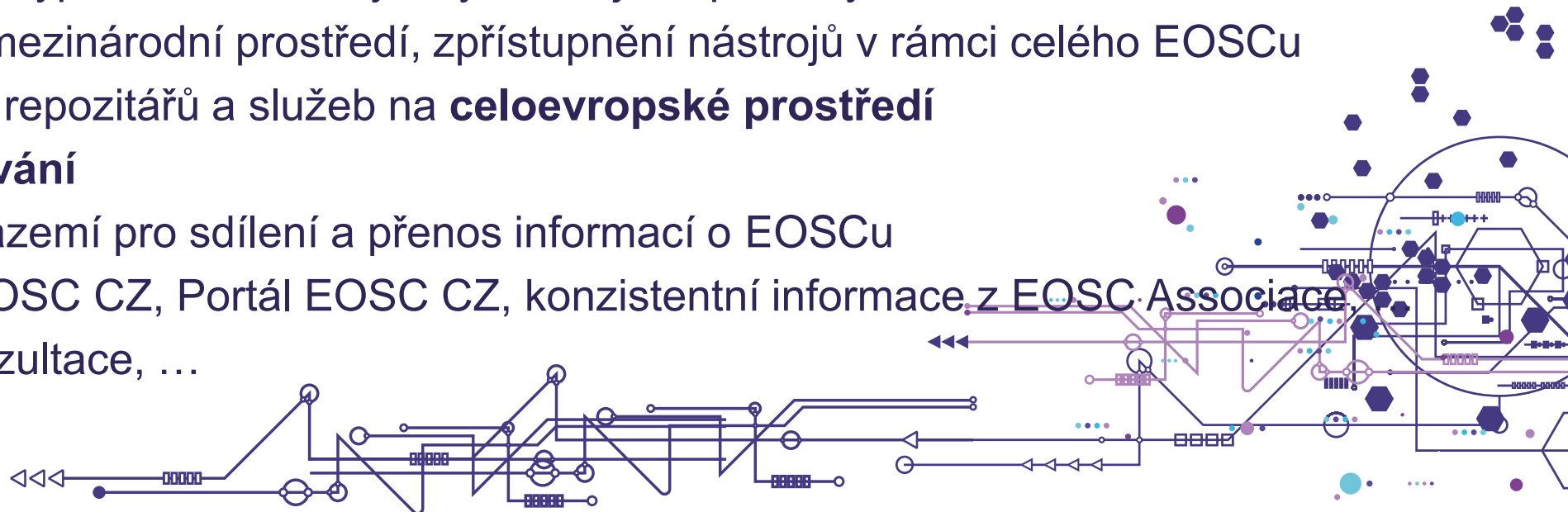
- NDI – komplexní prostředí pro strukturovanou práci s výzkumnými daty
 - Podpora tvorby **Data Management Plan**
 - Vykonatelný v prostředí NRP/NDI
 - PID zázemí
 - **Základní struktura Metadat**
 - Definice v DMP, vazba na NMA a NRP (případně kompatibilní oborově specifické repozitáře)
 - Jednotná **autentizace**, silné nástroje pro **řízení přístupu**
 - I na úrovni Metadat
 - Nástroje pro FAIRifikaci (kontrola a podpora „otevřenosti“ dat)



- **Robustní Národní metadatový adresář**
 - Distribuované, vysoce spolehlivé prostředí
 - Vysoká efektivita, rychlost vyhledávání (přístupu)
 - Řízení přístupu k metadatům
- **Robustní infrastruktura pro ukládání a zpřístupnění dat**
 - Distribuovaná, garance spolehlivého dlouhodobého udržení dat
 - Použitelná i pro „živá“, ne pouze archivní data
 - Řízení přístupu k datům, certifikace i pro citlivá data
- Podpora **tvorby** konkrétních **specifických tenantů** (např. institucionální repozitář)
 - Virtuální repozitáře, delegace správy a odpovědnosti „majitelům“
 - Univerzální datové úložiště pro zájemce bez oborové či institucionální příslušnosti



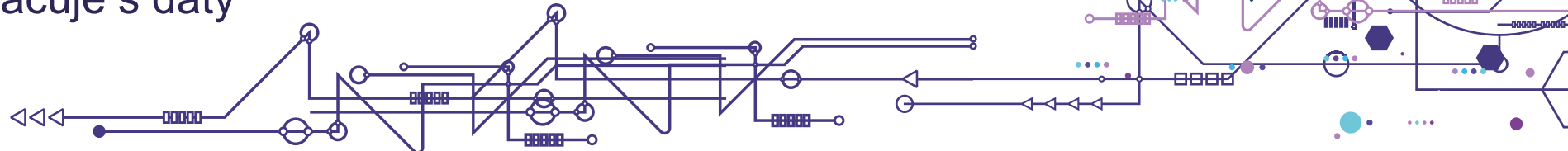
- Podpora **interoperability**
 - Metadatové struktury, ontologie
 - Hierarchický přístup, základní data v NMA, další ve specifických adresářích
 - Umožní vyhnout se problémům s „nejmenším společným jmenovatelem“
- Řada dalších nástrojů pro práci s daty
 - Přenosy dat k výpočetním a analytickým zdrojům; přenosy k uživatelům
 - Napojení na mezinárodní prostředí, zpřístupnění nástrojů v rámci celého EO SCu
- **Propojení** našich repozitářů a služeb na **celoevropské prostředí**
- **Školení a vzdělávání**
- Celkové aktivní zázemí pro sdílení a přenos informací o EO SCu
 - Sekretariát EO SC CZ, Portál EO SC CZ, konzistentní informace z EO SC Asociace
 - Podpora, konzultace, ...



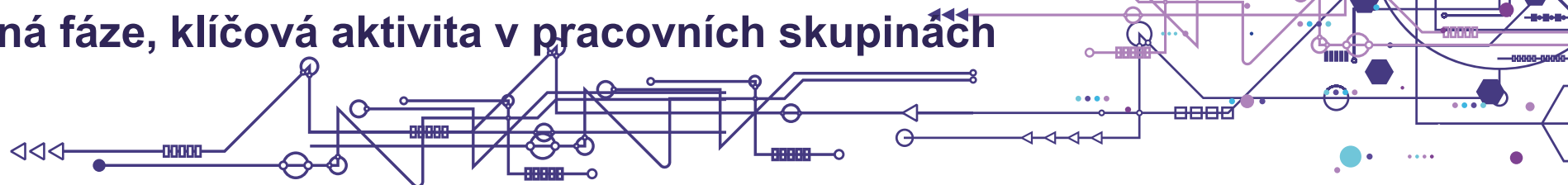
Jak se zapojit

A co to znamená

- Aktuálně – (aktivní) účast v pracovních skupinách
 - Možnost „být přítom“
 - Ovlivňovat směřování přípravy realizace
 - Zapojit se do diskuse o formě a zaměření budoucích projektů
- Využití finanční podpory (=zapojení do budoucích projektů)
 - Aktivní účast na budování NDI
 - Není přímo určeno pro uživatele NDI (ale pomoc s lidskými zdroji ano)
 - (Spolu)odpovědnost za příslušné komponenty, služby, procesy, ...
 - (Spolu)odpovědnost za budoucí udržitelnost infrastruktury
 - Technické i softwarové řešení
- Využití výsledků implementace EOOSC v ČR
 - Každý, kdo pracuje s daty



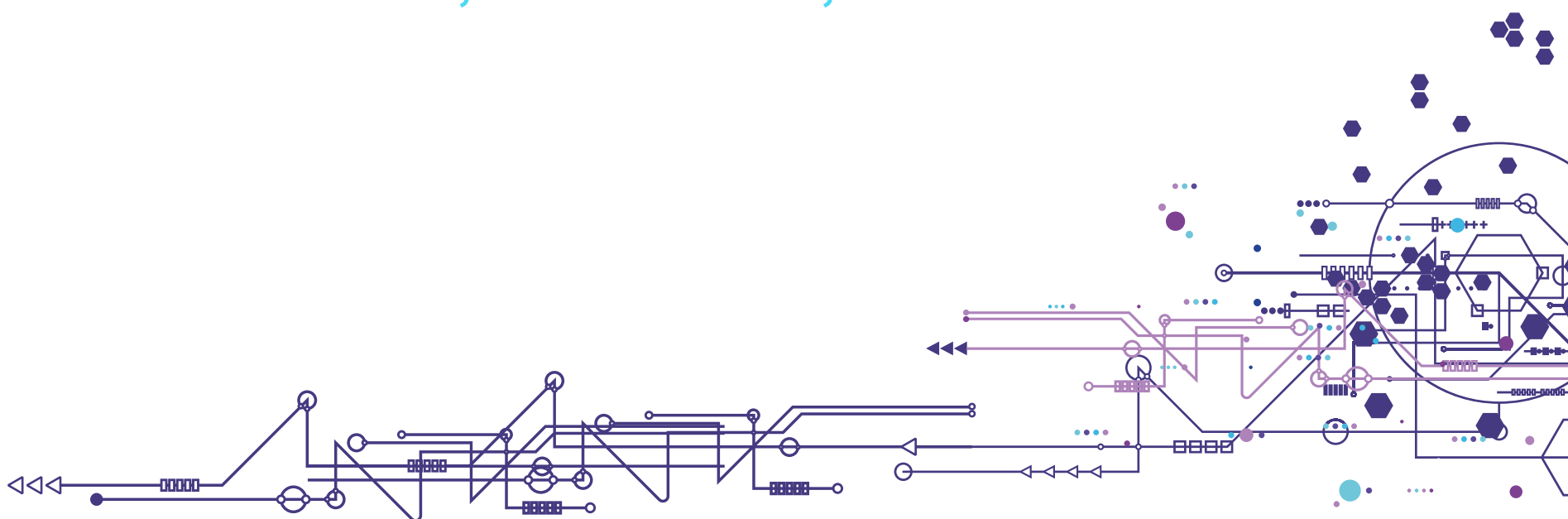
- EOSC jako příležitost k strukturovanější práci s výzkumnými daty
 - Technické i procesní zázemí, služby, nástroje, ...
 - Orientace na podporu znalostí (lidé, ne pouze technologie)
- Zapojení všech významných institucí
 - Budování Národní repozitářové platformy
 - Pokrytí oborových specifik, zapojení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





Připravujeme EOSC CZ Roadshow

1. 6. v Praze, 3. 6. v Brně, 17. 6. v Ostravě





Napište nám

info@e-infra.cz

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

A large, thin-lined circle containing the text 'e-infra.cz' in a bold, sans-serif font. The circle is centered on the page and is surrounded by decorative circuit-like patterns and colorful dots.

e-infra.cz