

Datové repozitáře a Národní repozitářová platforma

Open Science Week 2024

Marika Hrubá, Jaroslav Juráček

22. 10. 2024



Spolufinancováno
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

MUNI
ICS

Co nás čeká?

1. Repozitář pohledem poskytovatele podmínek (důvěryhodný repozitář)
2. Ukládání výzkumných dat v ČR - aktuální situace
prostředí EOSC / budování Národní datové infrastruktury
3. Diskuse

1. Repozitáře - projektové podmínky a současný stav



Repozitáře

„Informační systém určený k digitální archivaci, tj. zajišťující uložení, ochranu, integritu, autenticitu a zpřístupnění digitálních dokumentů v dlouhodobém horizontu“

- Co je repozitář?

repozitář je úložiště dat

- terminologie kolísá
- repozitář u nás: **úložiště dat s metadaty**
- s řízeným přístupem
- s odpovědností za uložená data (garance)

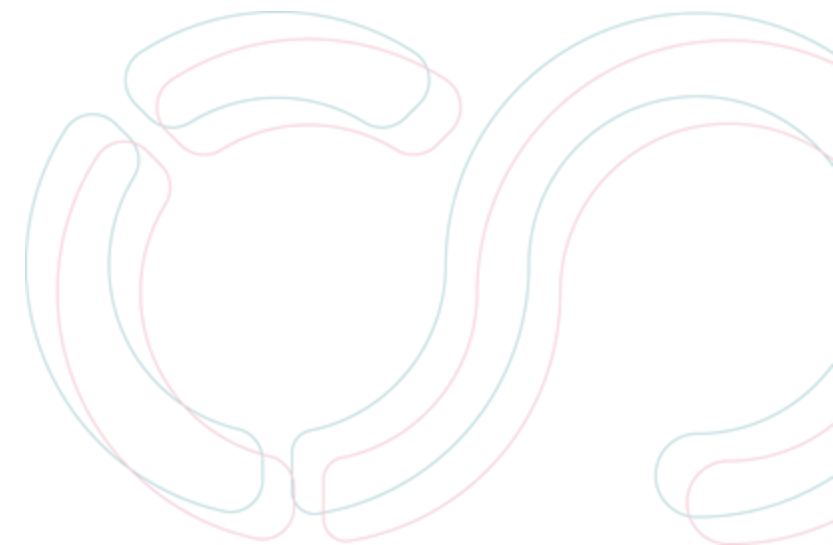
vyhledávání

Typy repositářů

- Oborové, národní, institucionální, catch-all, ...
- Běžný postup práce s výzkumnými daty je dělení po oborech, nikoliv po státech či institucích.
- Primárně nezohledňujeme národní/institucionální/... specifika, ale oborová specifika.
- Pokud řešíme projektový/institucionální/národní standard, musíme být kompatibilní s mezinárodními oborovými zvyklostmi.

Vlastnosti repositářů / Proč ukládat?

- Většinou zadarmo a vysoce bezpečné
- Respektují licence (ale podpora myšlenky Open Data)
- Data jsou široce dostupná a indexovaná
- Data lze snadno citovat
- Pro některá data dodatečné služby (fulltextové vyhledávání, ověření FAIRifikace)
- **Důvěryhodnost/zabezpečení**



Důvěryhodný repozitář - požadavek poskytovatele

Důvěryhodné repozitáře (Trusted repositories) :

- Certifikované repozitáře (např. CoreTrustSeal, Nestor Seal DIN31644, ISO16363) nebo oborové a doménové repozitáře běžně používané a schválené výzkumnými komunitami. Takové repozitáře by měly být mezinárodně uznávané. Není jich mnoho.
- Repozitáře pro všeobecné použití nebo institucionální repozitáře, které vykazují základní charakteristiky důvěryhodných repozitářů.

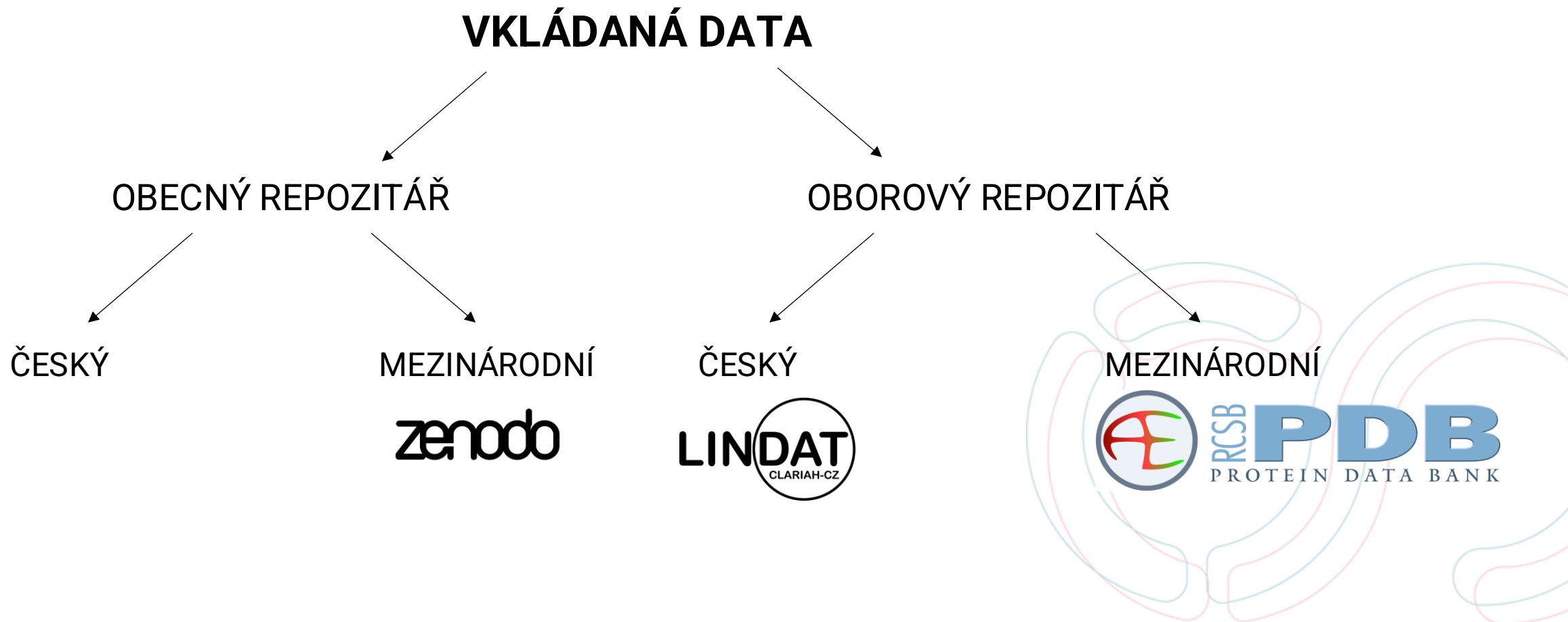
Důvěryhodný repozitář -vlastnosti

- **podporuje všechny platné licence určující přístup k datům a možnosti jejich používání**
- monitoruje dodržování pokynů
- má strategický plán (politiku) pro zajištění dlouhodobého provozu a uchování jeho obsahu
- **přebírá odpovědnost za dlouhodobé uchování, jehož zajištění má podchycené v dokumentaci a plánech**
- v maximální možné míře zajišťuje, že data jsou vytvářena, zpracovávána, zpřístupňována
- a používána v souladu s oborovými a etickými standardy;
- zaručuje integritu a autenticitu dat
- přijímá data a metadata na základě definovaných kritérií, aby byla zajištěna relevance
- metadata a srozumitelnost pro uživatele dat
- má tým s odpovídajícími odbornými znalostmi pro řešení kvality technických dat a metadat
- zajišťuje, aby koncoví uživatelé měli k dispozici dostatek informací, na jejichž základě mohou provádět hodnocení související s kvalitou
- **umožňuje uživatelům vyhledávat data a trvale na ně odkazovat prostřednictvím patřičné citace nebo trvalého odkazu**
- umožňuje opakované použití dat v průběhu času a zajišťuje, že jsou k dispozici vhodná

Posouzení důvěryhodného repozitáře

- Certifikace: Pokud repozitář prošel certifikací (např. CoreTrustSeal, Nestor Seal nebo ISO16363)
- Technické parametry (Příručka otevřené vědy pro OP JAK - vychází z CORE Trust Seal):
 - Podporuje všechny platné licence určující přístup k datům a možnosti jejich používání a monitoruje dodržování pokynů
 - Má strategický plán (politiku) pro zajištění dlouhodobého provozu a uchování jeho obsahu
 - Přebírá odpovědnost za dlouhodobé uchování, jehož zajištění má podchycené v dokumentaci a plánech
 - V maximální možné míře zajišťuje, že data jsou vytvářena, zpracovávána, zpřístupňována a používána v souladu s oborovými a etickými standardy
 - Zaručuje integritu a autenticitu dat

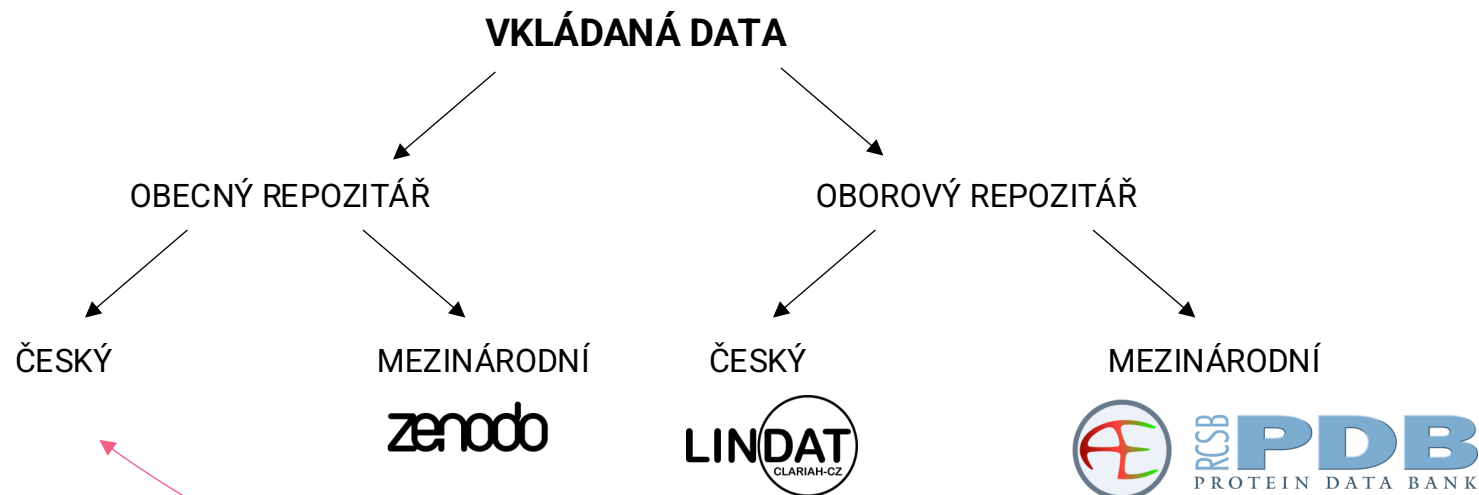
Doporučované repozitáře



2. Ukládání výzkumných dat v ČR - aktuální situace



Doporučované repozitáře

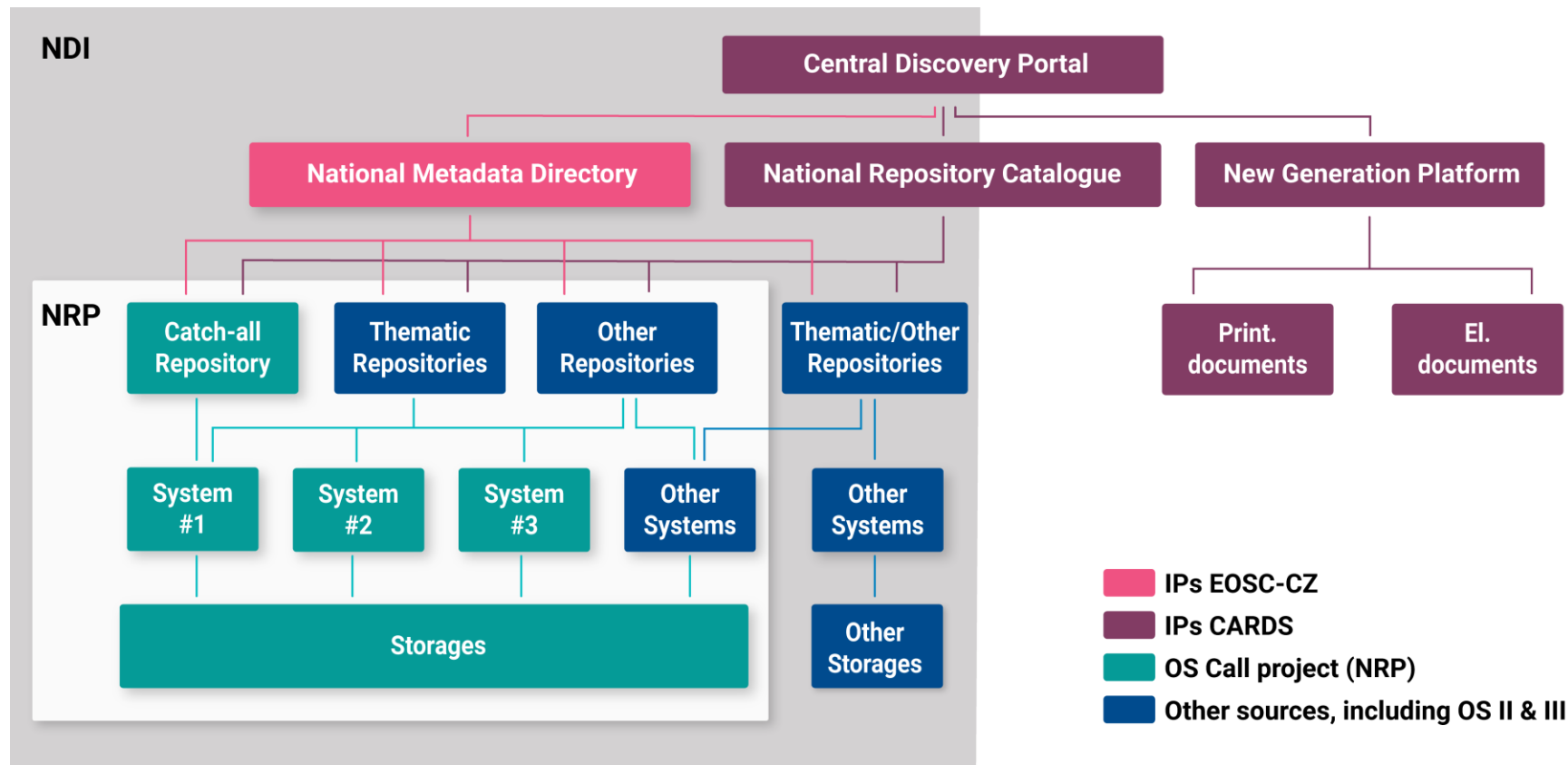


Národní datová infrastruktura

- Nové možnosti pro správu dat v ČR
- Společná platforma pro ukládání, sdílení, správu a přístup k datům a výpočetním zdrojům pro výzkumné účely
- Prostředí pro FAIR data
- Repozitáře, služby, nástroje
- Kompatibilita s mezinárodními technickými i datovými standardy



Národní datová infrastruktura



Co s daty?

- Kde uložit?
- Jak vyhledávat?
- Na čem počítat?
- Jak znovu použít?

Co Národní datová infrastruktura přinese z pohledu ukládání dat?

- Národní repozitářová platforma (distribuovaný systém pro vytváření instancí repozitářů)

- Český národní repozitář ← Data pro které není k dispozici vhodnější tematický či jiný specializovaný repozitář
- Čtyři pilotní oborové repozitáře (Repozitář pro biologická zobrazovací data, repozitář Národní úložiště pro biodiverzitní data, repozitář Databáze molekulární biofyziky, repozitář HUME Lab)

- Projekt Open Science II

- Rozvoj tematických/oborových repozitářů
 - Budování nových
 - Integrace již provozovaných/etablovaných
 - Rozšiřování kapacit stávajících

Pracovní skupiny EOOSC CZ



Architektura národní datové infrastruktury (NDI)



Metadata



Základní služby



Vzdělávání a lidské zdroje



Bio/zdraví/potraviny



Materiálové vědy a technologie



Data management pro umělou inteligenci a strojové učení



Sociální vědy



Fyzikální vědy



Humanitní vědy a umění



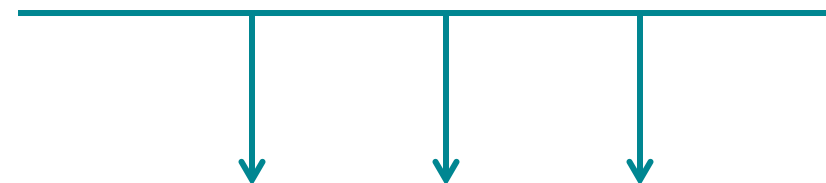
Environmentální vědy



Citlivá data

Prostředí budované vědci pro vědce

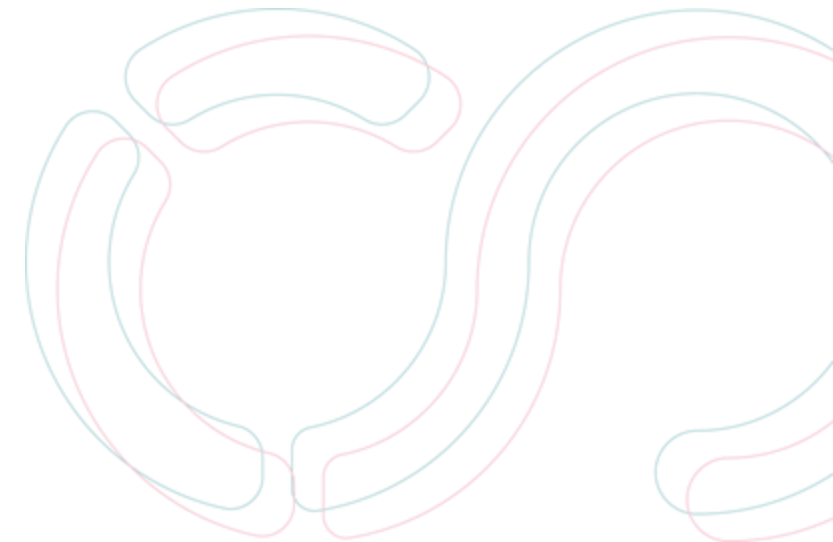
- Prostor, kde se vědci setkávají
- 4 průřezové
- 8 oborových
- otevřené zájemcům (kdykoliv přijímá nové členy)
- fungování po celou dobu trvání iniciativy



- Podílí se na tvorbě NRP a souvisejících služeb, nastavují **obsah ekosystému EOOSC v ČR**
- Definují potřebu budování nových repositářů a rozvoj služeb, prostředků a nástrojů zajišťujících zacházení s daty dle FAIR principů (projekt Open Science II)

Národní datová infrastruktura v širším kontextu

- Infrastruktura nechce duplikovat již zavedené repozitáře
- Převážně pro ty, kteří nemají k dispozici odpovídající řešení pro ukládání dat
- Spolu s ukládáním řeší také další fáze práce s výzkumnými daty
 - Nástroje pro DMP
 - Licence udělování přístupů
 - PID pro datové sady
 - Napojení na výpočetní zdroje e-INFRA CZ



Děkujeme za pozornost

www.eosc.cz



Spolufinancováno
Evropskou unií



MUNI
ICS

cesnet
.....

VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA

IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ
CENTRUM