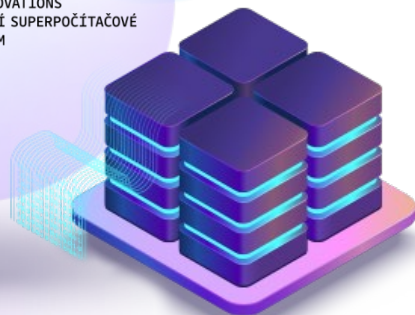




cerit
scientific cloud

cesnet
.....

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | IT4INNOVATIONS NÁRODNÍ SUPERPOČÍTAČOVÉ CENTRUM



Aktuální stav přípravy rozvoje národní datové infrastruktury ČR

Luděk Matyska

ÚVT MU

5. Národní den velkých výzkumných infrastruktur ČR



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

- IT podpora vědy se přesouvá od „jednoduchého“ počítání k péči o data
 - Reakce na nezanedbatelný počet rušených (retracted) publikací
 - Snaha minimalizovat riziko „garbage in garbage out“ syndromu
 - Reprodukovatelnost výsledků
 - Publikace neobsahují dostatek informací/podkladů pro opakování
 - Součástí nejsou programy a podrobná workflow, chybí konkrétně popsaná vstupní data
 - Neúplná nebo neexistující primární data
- **Dlouhodobá hodnota dat** zpravidla podceňována
 - Péče o data redukována na „pomocná data k publikací“
 - A ta vyžadovaná zpravidla nedostatečná

Open Science, Open a FAIR Data

- Na začátku radikální koncept **Otevřených dat**
 - Inspirace konceptem Open Access k publikacím
- Transformován do **FAIR dat**
 - **Findable** – data jsou nalezitelná
 - **Accessible** – data jsou přístupná
 - **Interoperable** – data jsou dostatečně precizně popsána
 - **Reusable** – data je možné využít opakovaně
- Samotné „otevření“ nestačí (navíc není vždy právně či jinak možné)
 - Data musí být kvalitní a spolehlivá
 - Musí být k dispozici dostatečně popisná metadata
 - A musí být kde data ukládat a dlouhodobě se zárukou držet

European Open Science Cloud (EOSC)



- Základem idea **federace (webu) FAIR** dat a souvisejících služeb
- Abstraktněji implementace **prostředí** (světa), které řádně spravuje **výzkumná data**, která jsou **kvalitní**, dobře **popsaná**, **dostupná** (za jasně definovaných podmínek) a mohou být **znovupoužita**
 - Tedy pokrytí všech aspektů (písmen) konceptu FAIR
 - Autoři (majitelé) dat si stále drží plnou kontrolu nad daty a tím, kdo je může (znovu)použít
 - Ostatní mohou data nalézt a ověřit si podmínky přístupu
 - Po udělení přístupu mohou s daty pracovat
- Na celoevropské úrovni podpořeno již několika desítkami souvisejících projektů
 - Zázemí, architektura služeb i infrastruktury, oborově-vědní klastrování, portál, ...

Implementace EOSC v ČR

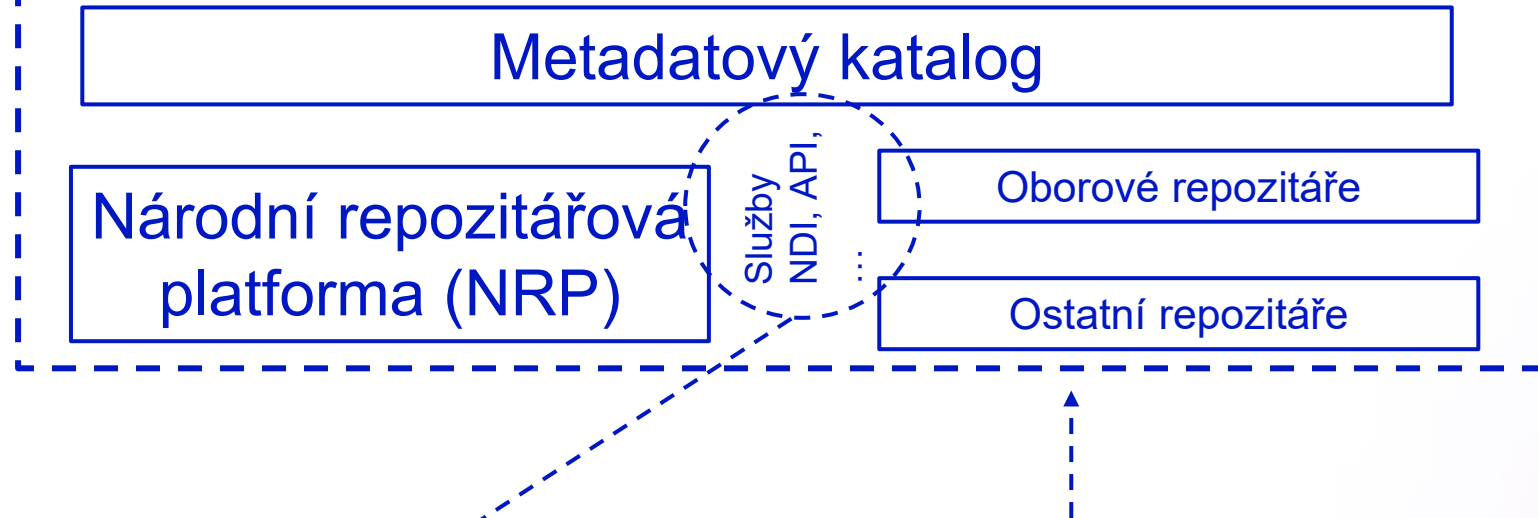
- Nultá fáze – politické rozhodnutí o podpoře EOSC v ČR (plán OP JAK)
- První fáze (od konce 2020) koordinována MŠMT
 - EOSC Platforma
 - Diskuse vhodného uchopení implementace EOSC v ČR
 - Výsledkem Architektura implementace EOSC v ČR (červen 2021)
 - Pracovní dokument, podklad pro další kroky
- Druhá fáze fakticky zahájena seminářem 23.9.2021
 - Ustavení pracovních skupin
 - Konkretizace konceptů z návrhu Architektury implementace EOSC v ČR
- Třetí fáze v současnosti
 - **Koordinační výbor pro implementaci iniciativy EOSC v ČR** při MŠMT
 - Konkretizace návrhu klíčových projektů
- Čtvrtá fáze – vypsání programu a příprava dalších projektů

Architektura implementace

- Prostředí pro ukládání, zpřístupnění a další práci s FAIR daty v ČR
 - Technické i procesní/operační standardy
 - Technická interoperabilita
- Nestaví se na zcela zelené louce
 - Existující repozitáře a prostředí pro ukládání dat
 - Plány pro institucionální repozitáře
 - Plány pro národní repozitář
- **Federace (mozaika, patchwork) konkrétních systémů**
 - Základní metadatový adresář
 - Národní repozitářová platforma
 - Repozitáře oborově-vědních clusterů
 - Služby a lidé
- **Národní datová infrastruktura** jako uchopitelné jádro EOSC v ČR

- **Národní repozitářová platforma**
 - Páteř, úložná kapacita, služby (AAI, PIDs)
 - Tvorba tenantů pro různé účely
- **Metadatový katalog**
 - Kde najít informace o tom, co v NDI je uloženo
- **Specifické repozitáře oborově-vědních klastrů**
 - Dlouhodobě budované, součástí i rozsáhlé znalosti metadatové, interoperability
- **Podpora lidských zdrojů**
 - **Datoví stewardi (kurátoři), datoví vědci**
 - Vzdělávání a lidé jako nezbytné „oživení“ technologie

Národní datová infrastruktura (NDI)



Služby NDI jsou dvojího typu:

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

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)
- Trainingového centra EOSC CZ zajišťující koordinaci tvorby školících materiálů pro NDI
- Rozvoj pozic datových kurátorů na akademických a výzkumných institucích v ČR

Pracovní skupiny

- Rozpracování aspektů navržené Architektury implementace EOSC
- Základem průřezové (aktuálně 4) pracovní skupiny
 - <https://www.e-infra.cz/eosc/pracovni-skupiny-pro-implementaci-eosc-v-cr>
- **Architektura**
 - David Antoš (CESNET)
- **Základní (Core) služby**
 - Michal Růžička (MU)
- **Metadata**
 - Petra Černošková (NTK)
- **Vzdělávání (a lidské zdroje)**
 - Radka Římanová (UK)

- Cíl: návrh architektury a implementace Národní datové infrastruktury
 - Implementace služeb NRP
 - Spolehlivé ukládání dat
 - Technické řešení systému řízení přístupu k metadatům i datům
 - Technické řešení práce s PID
 - Sběr a ukládání metadat přes celou NDI
 - Vyhledávání v metadatech
 - Vazba na služby e-INFRA CZ a podobné – zpracování dat
- Aktuálně 46 členů, 2 schůzky od podzimu 2021

- Cíl: definice základních služeb Národní repozitářové platformy
 - Práce s persistentními identifikátory (PID)
 - Služby řízení přístupu a autentizace a autorizace uživatelů
 - Služby pro podporu plánování správy výzkumných dat (kurátorství)
 - Služby pro FAIRifikaci výzkumných dat
 - ...
- Aktuálně 72 členů, 6 schůzek od podzimu 2021

- Cíl: Návrh architektury a implementace Metadatového adresáře NDI
 - Struktura a základní funkce Národního metadatového adresáře/katalogu
 - Struktura záznamu
 - Koordinace metadat o datových sadách a jejich interoperability
 - Spolupráce při definici základního modelu pro ukládaná data
 - Přebírání a definice doporučení pro správu metadat
- Aktuálně 53 členů, 3 schůzky od podzimu 2021

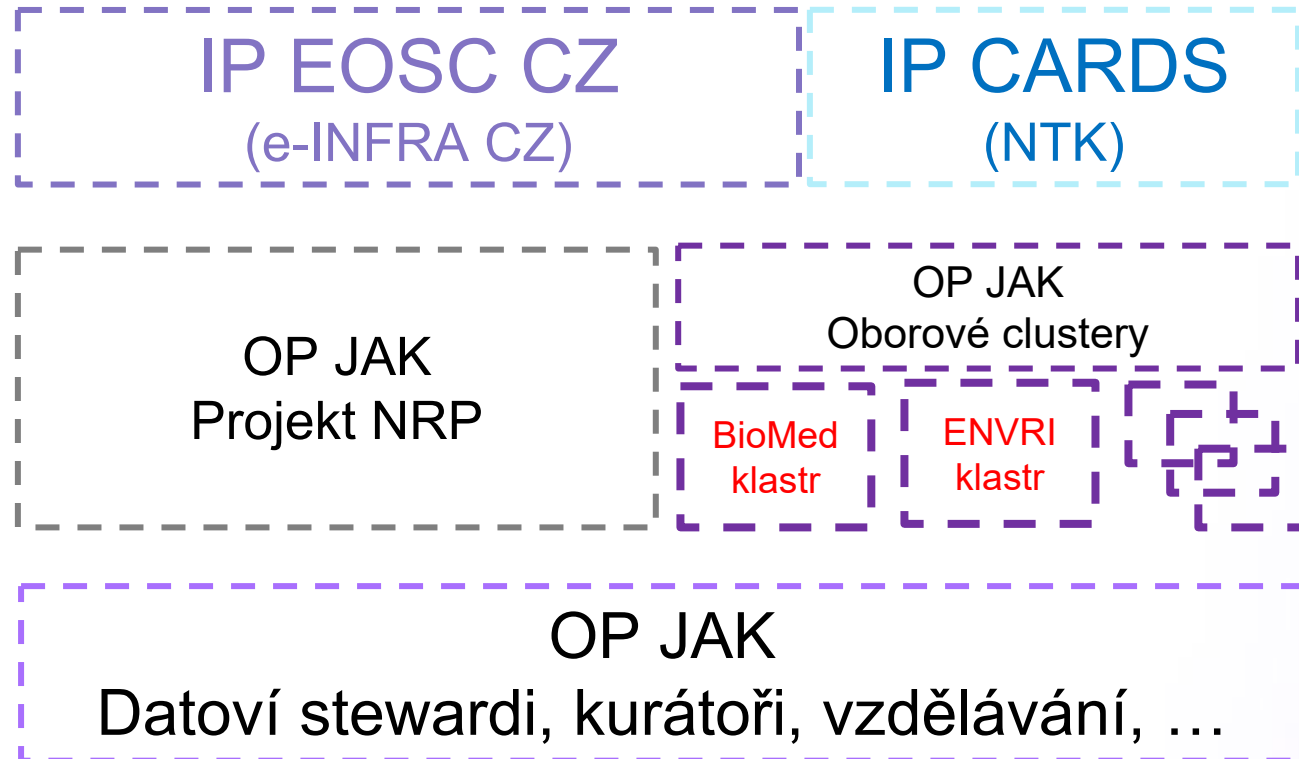
- Cíl: Podpora lidských zdrojů NDI
 - Definice potřebných kompetencí a související terminologie
 - Pozice a povolání, která jsou nutná pro dosažení cílů Open Science/EOSC (FAIR data) v ČR
 - Vzdělávání a podpora výzkumných pracovníků
 - Vzdělávání specifických pracovníků pro „data support“ – data steward, data curator
- Aktuálně 16 členů, první schůzka 2022

Pracovní struktura předpokládaných projektů



- **Integrované projekty systémové**
 - IP CARDS – NTK
 - IP EOSC CZ – e-INFRA CZ
- **Projekt Národní repozitářové platformy**
 - Institucionálně orientovaný
 - Implementace NRP a jejích služeb
 - Koordinace e-INFRA CZ
- **Projekty oborově-vědních klastrů**
 - Více oborově (infrastrukturně) orientovaných projektů
- **Podpora lidských zdrojů**

Předpokládaná struktura intervence OP JAK



IPs CARDS a EOSC CZ

- Vytvoření nezbytného zázemí pro implementaci EOSCu
 - Metadatový adresář a jeho služby
 - Včetně struktury metadat a napojení na oborové repozitáře
 - Kompletní implementace a provoz (garantovaná služba)
 - PID centrum, péče o PID
 - Služby kontroly přístupu (k metadatům i datům)
- EOSC sekretariát
 - Podpora pracovních skupin, zázemí celé implementace EOSC v ČR, napojení na mezinárodní aktivity, ...
 - Propojení EOSC a Open Data v ČR, monitoring postupu implementace, ...
- Podpora vzdělávacích aktivit
 - Koordinace, pořádání školení, ...

Projekt Národní repozitářové platformy

- Implementace výstupů pracovních skupin
- Kontinuální spolupráce s pracovními skupinami
 - Standardy pro manipulaci s daty, vklad nových požadavků a rozvoj služeb i samotné infrastruktury
 - Zapojení organizací mimo projekt, spolupráce s IPs CARDS a EOSC CZ, ...
- NRP vychází z principu cloudové platformy
 - Univerzální, ale nikoliv homogenní centrální prostředí (federalizace i na této úrovni)
 - Tenanty
 - Např. institucionální či oborově-specifické adresáře
 - Extenzivní podpora řízení přístupu
- Napojení na zázemí pro zpracování dat
 - Role institucí a zejména e-INFRA CZ

- Řada subjektů (zejména velké výzkumné infrastruktury) již dlouhá léta buduje vlastní systémy ukládání a zpřístupnění dat se specificky oborovým zaměřením
- Ty je nutné propojit do ekosystému Národní datové infrastruktury
 - Při zachování jejich specifických vlastností a potřeb
 - Přirozené rozšíření mozaiky národní repozitářové platformy
- Cílem EOSCu je překročení úzce oborově-specifických hranic
 - Provázání uvnitř i mezi klastry
- Zatím stále fluidní definice klastrů
 - Prvotní návrh v Architektuře implementace (FORDS), klastrování podle H2020 ([5 klastrů](#)) či Národní roadmapy velkých výzkumných infrastruktur, ...
 - Úkol pro příslušné oborově-vědní pracovní skupiny a jejich spolupráci
 - Svolání registrovaných zástupců, zahájení diskuse napříč obory

Projekty oborově-vědních clusterů

- Menší množství projektů, v souladu s dohodnutým členěním
- 2 primární cíle
 - Propojit oborově blízké instituce, infrastruktury, týmy v jejich práci s (meta)daty
 - Napojit na národní repozitářovou platformu a metadatový katalog
- Nezbytné aktivity v rámci oborově-vědních projektů EOSC v ČR
 - Podpora **zpřístupnění i mimo původní oblast zájmu**
 - Vnitro- i mezi-oborová spolupráce, podmínky přístupu
 - **Napojení** (prostřednictvím standardů a API) **na národní repozitářovou platformu**
 - Zapojení existujících repozitářů, koncept péče o mezioborová data
 - **Poskytnutí metadat do základního metadatového adresáře**
 - Spolupráce na definici struktury základních metadat, technické propojení, ...
 - Případné rozšíření kapacit a služeb

- Zahájena třetí fáze přípravných kroků implementace
 - Governance prostřednictvím Koordinačního výboru EOSC
 - Příprava Charty IPs CARDS a IPs EOSC CZ
 - Očekávané výstupy pracovních skupin
 - Metadatový adresář, zastřešující společné služby včetně vzdělávání nezbytných zastřešujících či průřezových služeb a související podpory pro IPs projekty
 - Preciznější vydefinování oborově-vědních klastrů
 - Příprava rámce dalších projektů, definice architektury a služeb NRP, ...
 - Zpřesnění harmonogramu výzvy/výzev Open Science v rámci OP JAK
- Časování umožňuje výborné propojení s aktuálními aktivitami EOSC Association
 - Jak přebírání trendů, tak i ovlivňování vývoje na celoevropské úrovni
 - Propojení Pracovních skupin na EOSC Association Task Forces



Děkuji za pozornost

e-infra.cz