

National Czech
Programme



Implementace iniciativy EOSC v ČR

Luděk Matyska

Open Science Workshop MU 2023
14.6.2023



Spolufinancováno
Evropskou unií



Open Science

- 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
 - Řešení složitých společenských problémů
- Důraz na spolupráci a znovupoužití dosažených výsledků obecně
 - Nové způsoby hodnocení kvality vědy i vědců samotných

Open Science

- Osm pilířů Otevřené vědy:
 1. **FAIR Data**
 2. Integrita výzkumu (Research Integrity)
 3. Metriky nové generace (Next Generation Metrics)
 4. Budoucnost vědecké komunikace (Future of Scholarly Communication)
 5. Věda s občany a pro občany (Citizen Science)
 6. Vzdělávání a schopnosti (Education and Skills)
 7. Odměny a stimuly (Rewards and Incentives)
 8. **EOSC (European Open Science Cloud)**

Snaha změnit způsob realizace vědecké práce, zvýšit kvalitu, reprodukovatelnost a dostupnost výsledků

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
 - Ale i nové formy zpracování (AI/DNN, ...)
 - Potenciál pro bezbariérový přístup
- Open Science → Open Access, Open Data, Open ...
 - Postupně se ukazuje, že může jít o příliš radikální přístup

FAIR Data

- Původní radikální koncept „Otevřených dat“ transformován do principu FAIR dat
 - **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 – o data pečujeme**

Co pro to potřebujeme/můžeme udělat

- Uvědomit si, že data se nebudou spravovat sama
 - **Problém způsobený nekvalitně zpracovanými/připravenými daty již existuje**
- **Data Management Plan**
 - Ne úplně šťastný pojem, nejde o jednorázovou aktivitu, ale dlouhodobý proces
 - Data je třeba sledovat (pečovat o ně) po celý jejich životní cyklus
- **Vyhledávání a interoperabilita**
 - Klíčová **metadata** a **ontologie**
 - Velmi rozdílná situace v různých vědeckých oblastech
 - Opět jde o trvalý proces, ne nějakou jednorázovou aktivitu
- **Znovupoužití**
 - Jsou definovány a dostupné podmínky, za nichž je možné je použít
 - Data někde musí být spolehlivě uložena

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, ...

Co je (a není) EOSC

- EOSC je především iniciativa, jak pomoci konkrétní složky vytvořit zázemí Open Science
- EOSC není „jedno řešení“ (jeden repozitář), ale, stejně jako Otevřená věda sama, způsob nahlížení a práce s výzkumnými daty a dalšími digitálními artefakty
- Implementace EOSCu má různou podobu v různých zemích i na Evropské úrovni
 - „Web prostředí a služeb souvisejících s daty“ je možné realizovat mnoha způsoby
- V srdci jsou data – celkem přirozeně se implementace točí kolem nějaké formy datové infrastruktury
 - Někde jen procesy – co a jak se má s daty tvořit
 - Jinde nástroje – podpora při manipulaci a posílení existujících systémů
 - Někde důraz na konkrétní vědecké disciplíny a jejich propojení
 - Jinde snaha o komplexnější zázemí
- **Všechny přístupy k implementaci EOSCu spojuje zájem o data**

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 Asociace

- 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
 - Reprezentovat 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)
 - ELI-ERIC

EOSC Partnerství

- Jediné skutečně horizontální partnerství v rámci HEU
- Příslib 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 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**
 - **Czech National Tripartite Event proběhl 1.6.2023**

Principy implementace EOSC v ČR

- EOSC jako federovaný systém FAIR dat a služeb → **Národní datová infrastruktura**
 - Národní repozitářová platforma, Národní metadatový adresář, tematické repozitáře, ...
 - Řízení přístupu, spolehlivé přesuny dat, ...
 - Procesní a lidské zázemí
- Cílem jsou samostatná primární a odvozená data
 - ne pouze data/„přílohy“ k publikacím
- „Stavitelé“ a „uživatelé“
 - Jeden pilíř je vytvoření tohoto federovaného systému na národní úrovni a propojení na mezinárodní úroveň
 - Kapacita, služby, ...
 - Další pilíř je podpora výzkumných týmů i vědců, využití vybudovaného EOSCu
 - Školení, vzdělávání, ...

Národní datová infrastruktura (NDI)

Klíčové komponenty

- **Národní metadatový adresář (NMA)**
 - Klíčová komponenta, zajišťující vyhledatelnost a interoperabilitu dat
 - Vazba na všechny ostatní aktivity v rámci NDI
 - Centrálně spravovaná komponenta
- **Národní repozitářová platforma (NRP)**
 - Páteř NDI, zajišťuje úložnou kapacitu a související služby (AAI, PIDs, DMP, ...)
 - Konkrétní repozitáře jsou fakticky “virtuální” tenanty
 - Institucionálně i komunitně budovatelné repozitáře
 - Řízení přístupu v rukou autorů/správce
 - Definice standardů/rozhraní pro napojení
 - Distribuovaná mozaika (patchwork) fyzických úložišť
 - Realizace v rámci širšího konsorciálního projektu

Národní metadatový adresář (NMA)

- Speciální repozitář“, který drží metadata všech dat, vytvořených v ČR
 - V principu databáze metadat
- Spolehlivost a výkon
 - Musí být neustále dostupný
 - Metadata se nesmí ztratit
 - Lze očekávat velký počet současných přístupů
 - Zápis i čtení (čtení by mělo převažovat)
 - Podpora automatického nahrávání metadat
- Plánujeme vysoce distribuovanou architekturu
 - 5 lokalit (3-5 měst)
 - Synchronizace
 - Propustnost, rychlý failover, paralelní přístup, ...
- Řízení přístupu k metadatům

Národní datová infrastruktura (NDI)

Další součásti

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

Očekávané služby EOSC-CZ

Výstup pracovní skupiny Základní služby

- Důvěryhodné repozitáře
- Služby pro podporu práce s persistentními identifikátory (PID)
- Služby pro řízení přístupu k datům, autentizaci a autorizaci uživatelů (AAI) a podporu jejich spolupráce
- Služby pro podporu a plánování správy výzkumných dat
- Machine actionability
- Služby pro FAIRifikaci výzkumných dat
- Služby transferu dat
- Kvalita služeb
- Kompatibilita s existujícími platformami
- Statistiky a přehledy
- Národní příspěvek k budování EOSC

Řízení přístupu

- Data uložená v NDI jsou pod plnou kontrolou jejich vlastníků
 - Zpřístupnění dalším osobám se děje na pokyn jejich vlastníka
- Už samotná informace o existenci dat či metadata může být citlivá informace
 - Řízení přístupu až na úrovni řízení viditelnosti existence samotných metadat či jejich částí apod.
- Federované AAI služby
 - Autentizace uživatelů přes federované služby přihlašování (eduID.cz apod.),
 - tj. s využitím existující identity z instituce, bez nutnosti vytváření dalšího účtu, hesla atd.

Repozitář jako tenant RNP

- Tenant NRP
 - Standardní repozitář
 - určený specifickou konfigurací
 - URL
 - vizuální podoba (skin)
 - popis v něm používaných metadatových modelů
 - podpora pro specifická metadata (např. vizualizace)
 - s určeným řízením přístupu
 - navázaný na NMA a registr repozitářů
 - zejména exportuje metadata, je registrován, ...
 - nepoužití dostupné komponenty má být řádně zdůvodněno
- Tenant typicky oborový repozitář
 - pravidla jejich vytváření

Přípravné kroky 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
 - 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



Tématické (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** – Adam Svobodník, Věra Franková (BBMRI CZ, CZECRIN, EATRIS CZ)

Koordinační výbor EOSC v ČR

- **Koordinační výbor EOSC pro implementaci iniciativy EOSC v ČR**
 - Odborný poradní orgán ministra školství, mládeže a tělovýchovy
 - Koordinace stakeholderů výzkumu, vývoje a inovací (dále jen „VaVaI“) ČR při rozvoji datové infrastruktury ČR
 - Podílí se na přípravě architektury provádění iniciativy EOSC v ČR a na přípravě koncepce podpory rozvoje datové infrastruktury ČR
 - Plní funkci diskuzní a konzultativní platformy, která formou svých usnesení přijímá doporučení, která MŠMT bere v potaz jako vstup do přípravy investičních intervencí OP JAK a zohlední je při přípravě výzev OP JAK cílených na podporu rozvoje datové infrastruktury ČR.

Projektová podpora 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, IPs CARDS schválen, IPs EOSC v hodnocení

2. Národní repozitářová platforma a její služby

- Výzva vypsána, realizace od Q4 2023/Q1 2024
- Společný konsorciální projekt NRP v přípravě

3. Tematické a další specifické repozitáře a jejich služby

- Výzva připravována 2023, otevřena začátkem 2024, realizace od Q3/Q4 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

Projektová zapojení

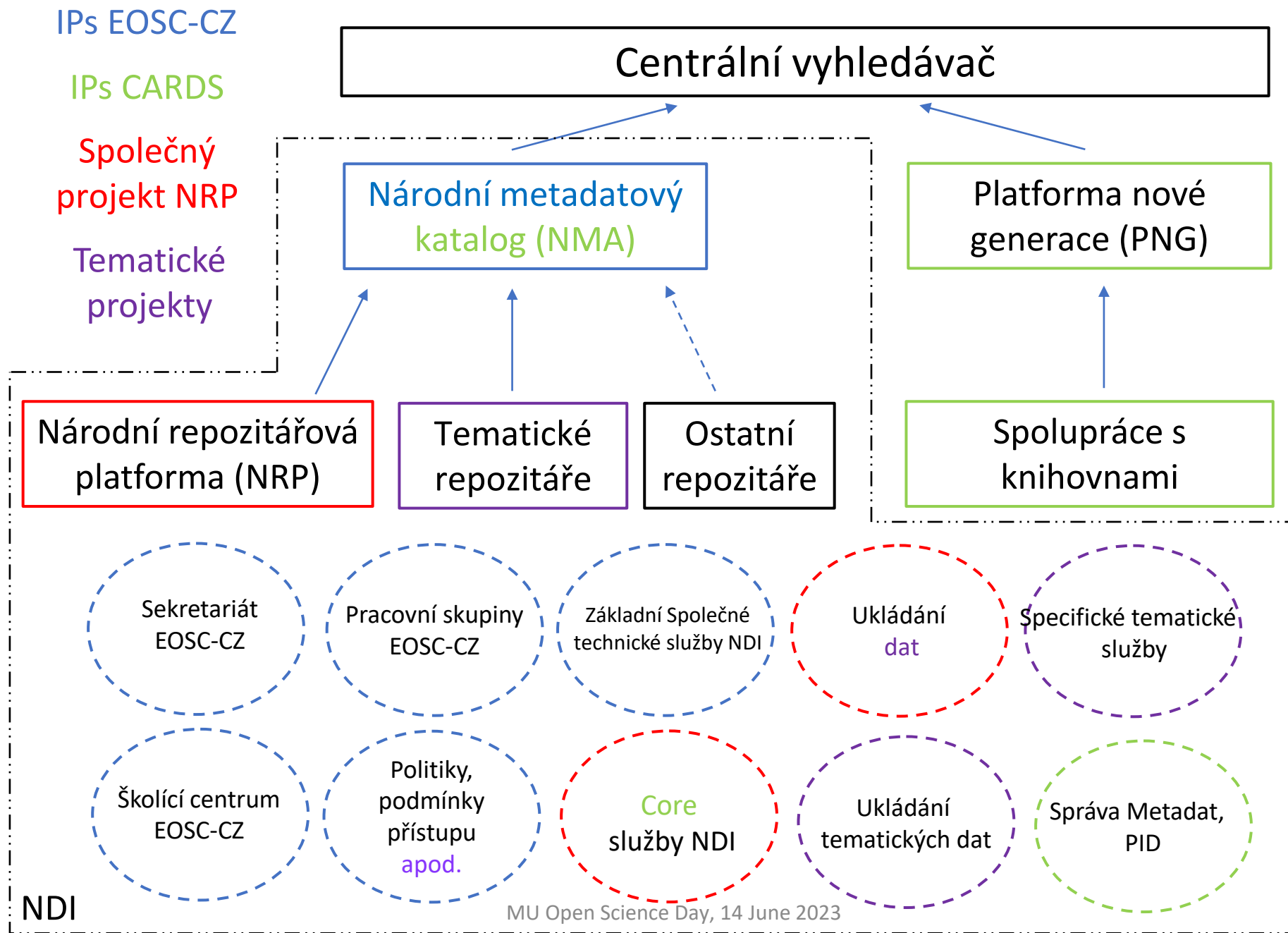
- Projekty slouží k **budování NDI**
 - Partneři konsorcií instituce/týmy se zkušeností z oblasti budování a provozu zázemí (technického i znalostního) pro práci s daty
 - Závazek udržitelnosti vytvořeného prostředí
 - Ideálně integrace do institucionálních či infrastrukturních plánů
 - Přirozená role 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

IPs CARDS (NTK) a EOSC-CZ (MU/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

Projekt NRP

- Konsorciální projekt, až 15 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%



Co přinese implementace EOSC v ČR

- **NDI** – Snadno použitelné komplexní prostředí pro strukturovanou práci s výzkumnými daty
 - **NRP** – Robustní infrastruktura pro ukládání a zpřístupnění dat, vazba na zpracování
 - Živá i archivní data
 - Kontrola přístupu k datům i metadatům v rukou autorů/majitelů/správce dat
 - **Tematické repozitáře**: otevřenost NDI
 - Podpůrné nástroje (DMP, FAIRifikace, ...)
- **NMA**: Viditelné a vyhledatelné ontologie a metadata
 - Vyhledatelná a interoperabilní data
- **Školení a vzdělávání**
 - Nejen použití NDI, ale i širší zázemí Open Science
- **Propojení** našich repozitářů a služeb na **celoevropské prostředí**
- Aktivní zázemí pro sdílení a přenos informací o EOSCu a FAIR data principech
 - Sekretariát EOSC CZ, Portál EOSC CZ, konzistentní informace z EOSC Asociace, ...
 - Podpora, konzultace, ...

EOSC, FAIR data a MU – Strategie

- FD1: Strategie pro správu a zpřístupňování výzkumných dat MU
- FD2: Evidence výzkumných dat
- FD3: Datový repozitář
- FD4: Podpora pro správu a zpřístupňování výzkumných dat
- FD5: Infrastruktura pro výzkumná data

FD1: Strategie pro správu a zpřístupňování výzkumných dat MU

- Výsledky implementace EOSC v ČR nabízí zázemí a nástroje
- Vazba na „best practices“ v rámci ČR i EU
 - úzká spolupráce s výzkumnými komunitami (např. prostřednictvím velkých výzkumných infrastruktur)
- Rovnováha mezi „přáními“ a ekonomickou možností
 - Rozsah i typ uchovávaných dat
- Významná role Data Management Plans

FD2: Evidence výzkumných dat

- Klíčová role Národního metadatového adresáře
- Agregace informací o datech, vzniklých na nebo prostřednictvím MU
- Přehled nezávisle na místě fyzického uložení vlastních dat

FD3: Datový repozitář

- V rámci FD1 zhodnotit smysl a přínos institucionálního repozitáře
 - Kam vědci budou skutečně ukládat data
- Vazba na oborové, infrastrukturní, národní i mezinárodní repozitáře
- Význam vlastní kopie dat na MU
 - A také ekonomické dopady
- Vědci musí mít možnost data někam ukládat, ale nemusí to být fyzicky na MU
 - Role univerzitní strategie
 - Role Open Science manažera
 - Role a zapojení vlastních vědeckých týmu
- Přehled drží NMA

FD4: Podpora pro správu a zpřístupňování výzkumných dat

- Kompromis mezi národním (i EU) zázemím a vlastní podporou na MU
- Open Science/FAIR data manažeři, kurátoři a vědci
 - Role centrálních pozic versus vlastní výzkumné týmy
- FAIR data jako standardní přístup práce s daty
 - Především práce vlastních výzkumných týmů
- Přidaná hodnota versus zbytečná další byrokracie
 - Péče o data musí přispívat k lepší výzkumné práci
- Role sekundárních (tedy ne-vlastních) datových zdrojů
- Kvalita a důvěryhodnost dat
 - Zcela klíčové pro jejich hodnotu a znovupoužití

FD5: Infrastruktura pro výzkumná data

- NRP a oborové repozitáře
 - Kvalita služeb, podpora rozvoje i využívání
- Nástroje a služby
 - Actionable DMP – automatizace procesů
 - Data Stewardship Wizard – podpora FAIRifikace
- Vlastní role/příspěvek MU
 - Ukazování směru, definice a podpora „best practices“
 - Kurátorství dat, doba retence dat, ekonomické důsledky, ...
- Repozitáře s daty versus živá virtuální výzkumná prostředí
 - Role dalších projektů
 - Zapojení do mezinárodní spolupráce

EOSC a Data ve výzkumu

- Data jsou nedílnou a kritickou součástí vědecké práce
- Data vyžadují péči, ta by měla být součástí standardní výzkumné činnosti
 - EOSC vytváří zázemí a podporu, aby se toto stalo
- Implementace EOSC představuje příležitost významného „kulturního“ posunu
 1. Zlepšení práce s vlastními daty
 - Metadata, provenance, kvalita
 2. Kontrolované zpřístupnění dat
 3. Využití „ne-vlastních“ datových sad
 4. Integrace do výzkumu
 - Role datových kurátorů a datových vědců
- Role institucionálních strategií a aktivit

Shrnutí

- Implementace EOSC nám poskytne zázemí a nástroje pro strukturovanější práci s výzkumnými daty
 - Technické i procesní zázemí, služby, nástroje, ...
 - Orientace na rozvoj znalostí (lidé, ne pouze technologie)
- Umožní vědcům a institucím se soustředit na vlastní péči o data
 - Současně ale za ně péči o data neudělá
 - Klíčová role vlastních strategií a cílů
- Význam pro MU
 - Nemusíme si vše budovat „jen pro sebe“
 - Máme příležitost ovlivnit vývoj na národní i mezinárodní úrovni
- **Můžeme demonstrovat, že péče o data vede k lepšímu výzkumu se širším dopadem**

National Czech
Programme



Děkuji za pozornost

Kontakt a více info:

info@eosc.cz

www.eosc.cz



**Spolufinancováno
Evropskou unií**

MU Open Science Day, 14 June 2023


MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

www.eosc.cz

Národní portál implementace iniciativy EOSC v ČR

