

Rozvoj Datového repozitáře AV ČR a podpora Open Science v AV ČR

Příprava projektu do OP JAK

24. 5. 2022

Současná situace v AV ČR

- Ústavy AV ČR, řeší své potřeby z hlediska ukládání a správy dat různými způsoby, většinou neexistuje jasná evidence o všech vyprodukovaných datových souborech ani na úrovni ústavů. Stejně tak zpravidla neexistují politiky, jak s daty nakládat. V řadě případů jsou využívány oborové repozitáře spravované třetími stranami (Zenodo,...)
- Pro většinu ústavů nejsou k dispozici spolehlivá řešení pro dlouhodobé uchování těchto dat, jejich správu a zpřístupnění.
- Chybí potřebné know-how v této oblasti.

Současná situace v AV ČR

KNAV provozuje Datový repozitář AV ČR ASEP

- omezená kapacita pro jednotlivé datové sety (max 20 GB)
- interoperabilita je zajištěna směrem k OpenAIRE
- metadatová struktura repozitáře vychází z průzkumu na mezinárodní úrovni (v ČR zatím neexistuje standard vůči kterému by bylo možné ji porovnat)
- podpora perzistentních identifikátorů handle a ORCID
- propojení na výsledky výzkumu (články, knihy, aplikované výsledky...)

Současná situace v AV ČR

- V současnosti je Datový repozitář AV ČR ASEP využíván pouze několika ústavy.
- Chybí větší propojení na národní úrovni - zatím neexistuje centrální autorita, která by informace o datech shromažďovala (Úřad vlády - RIV, NTK,...?).

Cíle nového projektu

Cílem projektu je vytvoření spolehlivého centrálního datového repozitáře AV ČR, zohledňujícího potřeby pracovišť AV ČR, kompatibilního s mezinárodními standardy a specializovaného týmu, který bude zajišťovat potřebnou podporu ústavům AV ČR a jejich vědeckým pracovníkům.

Cíle nového projektu

- 1) Rozvoj současného datového repozitáře po stránce standardů, kapacity a dlouhodobé udržitelnosti v souladu s Open Science a FAIR principy
- 2) Zapojení do EOSC
- 3) Zajištění správy vědeckých dat a podpory tvorby data management plánů v souladu s Open Science a FAIR principy
- 4) Vytvoření specializovaného podpůrného týmu pro oblast Open Science/FAIR Data - Data Stewards pro vědecká pracoviště AV ČR

Participující instituce

Knihovna AV ČR, v. v. i. - hlavní koordinátor/řešitel

Středisko společných činností AV ČR, v. v. i.

Biotechnologický ústav AV ČR, v. v. i.

Archeologický ústav AV ČR, Praha, v. v. i.

Ústav fyzikální chemie J. Heyrovského AV ČR, v. v. i.

Ústav fyziky plazmatu AV ČR, v. v. i. (nový partner)

Knihovna AV ČR, v. v .i.

Koordinace a zajištění komunikace

- mezi členy projektového týmu
- s ústavy AV ČR
- s vedením AV ČR
- s dalšími VO
- s dalšími institucemi zajišťujícími podporu Open Science (knihovny, CESNET,...)
- s MŠMT
- s EOSC Association

Knihovna AV ČR, v. v .i.

Rozvoj Datového repozitáře AV ČR ASEP

- zajištění logické správy vědeckých dat
- zajištění persistentních identifikátorů (lidé, data, afilace,...) v souladu s národními i mezinárodními standardy
- zajištění podpory národních i mezinárodních standardů z hlediska struktury popisu dat
- interoperabilita s vnějšími systémy (OpenAIRE, Národní repozitář NTK, ...)
- možnost individuálního nastavení pro oborově specifické datové sety
- ...

Knihovna AV ČR, v. v .i.

Vytvoření specializovaného podpůrného týmu pro oblast Open Science/FAIR Data

- Data Stewards pro vědecká pracoviště AV ČR
- technicky orientovaní specialisté
- poskytování servisu ústavům v oblasti data managementu, včetně školení, napojení systémů pro automatizovaný přenos, webové rozhraní, atp.

Středisko společných činností AV ČR

Vytvoření datového úložiště

- zajištění bitové ochrany
- propojení s Datovým repozitářem AV ČR ASEP
- spolupráce s externími uložišti (např. CESNET)?

Participující ústavy

Definice požadavků na funkce datového repozitáře s ohledem na svou relevantní oblast věd a typy dat

- monitorování situace ve své oblasti věd
- specifikace potřeb oborového repozitáře
- specifikace potřeb z hlediska uspokojení poskytovatelů účelových dotací
- definice napojení na další projekty a infrastruktury (ELIXIR, Instruct, MOSBRI, Nanolab, NanoENVI, AIS a další infrastruktury, do kterých jsou zapojené)

Participující ústavy

Pilotní projekt návrh a implementace datového repozitáře (BTÚ)

- charakterizace všech produkováných vědeckých dat a jejich data managementu
- use cases definující kolik a jakých dat se archivuje, kde a jak (jen backup, skupinový/veřejný repozitář, LTP, nástroje pro živá data, ..., ústavní/AV repozitář, oborové/mezinárodní (PDB))
- snaha analyzovat a navrhnout přenositelné i pro ostatní LS ústavy

Otázky (prosinec 2021)

Nevíme jaký by byl **zájem ústavů** AV ČR o využití repozitáře

Nevíme jak velké **objemy dat** by mohly ústavy AV ČR do repozitáře ukládat

Nevíme zda jsou na ústavech **lidé pověřeni koordinací** této oblasti

Nevíme, zda lze počítat s podporou **zajištění udržitelnosti** ze strany zřizovatele po ukončení projektu

...

Zájem ústavů o využití repozitáře

ANO - 39 ústavů

NE - 15 ústavů

Zájem o služby podpůrného týmu OS

ANO - 40 ústavů

NE - 14 ústavů

Účast v přípravě OP JAK projektů

ANO - 14 ústavů

NE - 40 ústavů

Objemy dat - odhad na základě dotazníku

Iniciační dávka - 2 PB

Roční přírůstek - 0,4 PB

Účast v aktivitách souvisejících s EOSC

ANO - 14 ústavů

NE - 40 ústavů

<https://www.e-infra.cz/eosc/pracovni-skupiny-pro-implementaci-eosc-v-cr>

Zajištění udržitelnosti

???

Plánované vyhlášení výzev OP JAK

Individuální projekty systémové - VaV I - červen/červenec 2022

Alokace 2 000 000 000 Kč

Open Science I - leden 2023

Alokace 1 000 000 000 Kč

https://opjak.cz/wp-content/uploads/2022/03/Harmonogram_vyzev_OP-JAK_2022.pdf

Aktuálně příprava IPS projektů do OP JAK

Národní technická knihona - **projekt CARDS**

Masarykova univerzita (CESNET/e-INFRA) - **EOSC-CZ**

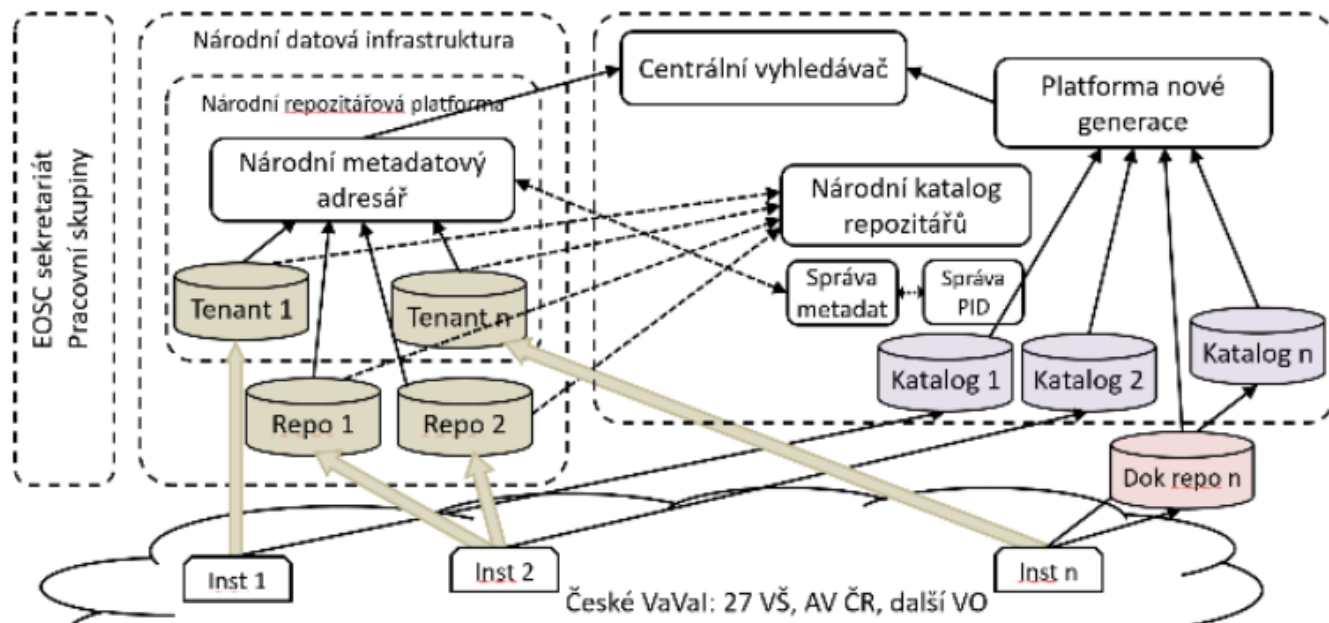
Komplementarita

EOSC-CZ

CARDS

Vědecká data

elektronické i tištěné dokumenty



Kontakt

Ing. Martin Lhoták

Knihovna AV ČR, v. v. i.

lhotak@knav.cz