

Správa výzkumných dat a iniciativa EOSC CZ

Martin Dvořák, Masarykova univerzita

23. 10. 2025



Spolufinancováno
Evropskou unií



MUNI
ICS



Krise reprodukovatelnosti výzkumu

Publikované studie nelze zopakovat s dosažením stejných výsledků.

- Výzkumníci za použití stejných dat a postupů nedokáží dospět ke stejným závěrům
- Zpochybnění důvěryhodnosti části vědecké produkce
- Příčiny:
 - Nedostatečně popsaná metodika výzkumu
 - Publikační tlak, malé vzorky, manipulace s daty či analýzou, změna podmínek
 - Nedostupnost/ztráta dat nebo kódu
 - Neexistence/nedodržování standardů pro správu dat a metadat

EOSC – European Open Science Cloud

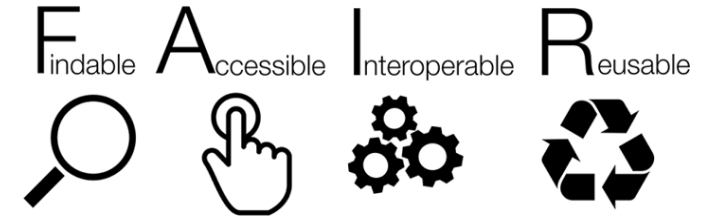
- Iniciativa na podporu správy výzkumných dat napříč zeměmi a vědeckými komunitami
- Federovaný ekosystém, který propojuje výpočetní, úložné a datové zdroje v Evropě
- Správa dat po dobu celého **životního cyklu**

→ Nejedná se o řešení krize, ale vytváří podmínky, prostředí, standardy pro zlepšení reprodukovatelnosti výzkumu

- **FAIR data**



FAIR data

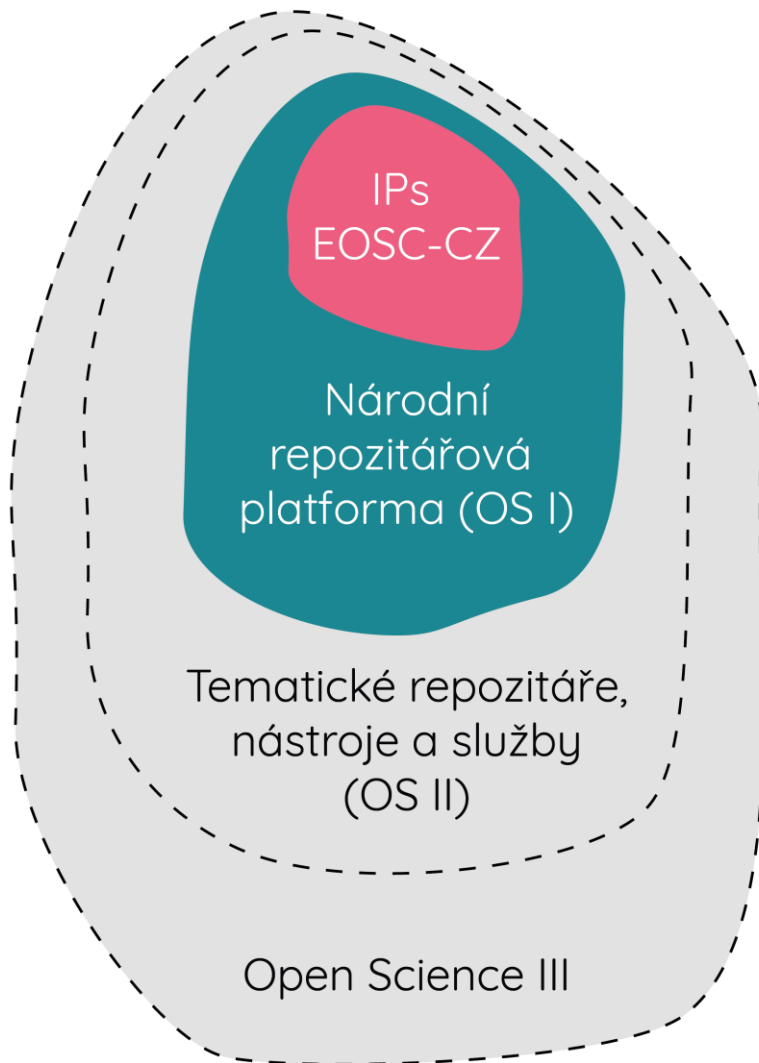


- Data, která jsou nalezitelná/vyhledatelná - findable
- Data, která jsou přístupná - accessible
- Data, která jsou dobře anotována/popsána – interoperable
- Jsou vyjasněny podmínky dalšího užití - reusable

Iniciativa EOSC v ČR

Hlavním výstupem je **Národní datová infrastruktura** pro výzkumná data.

- Společná **platforma** pro sdílení, správu a přístup k datům a výpočetním zdrojům
- Soustava projektů z operačního programu Jan Amos Komenský (OP JAK)



IPs EOSC-CZ (od r. 2023) – organizační a znalostní zázemí

- Organizační (**Sekretariát**) – <https://www.eosc.cz/o-nas/sekretariat-eosc-cz>
- Technická (**Národní metadatový adresář**) – <https://nma.eosc.cz/>
- Vzdělávací (**Školicí centrum**) – <https://www.eosc.cz/skoleni>

Národní repozitářová platforma (OS I, od r. 2024) – „železo”

- **Repozitářové systémy** (DSpace, CESNET Invenio, ASEP-ARL)
- **Pilotní repozitáře**
- **Základní služby** (PIDs, DSW, licence, ...)
- **Compliance** a UX (kyberbezpečnost, ServiceDesk, ...)
- **Školení** – technické aspekty

OS II (od r. 2025) – „oborově specifické služby”

- **Na základě vstupů od 8 tematických pracovních skupin**
Bio/Zdraví/Potraviny, Materiálové vědy, AI & ML, Sociální vědy,
Fyzika, Humanitní vědy a umění, Enviro, Citlivá data

OS III (od r. 2026) – „rozvoj znalostí a dovedností”

CARDS (od 2023) – knihovně-informační zázemí

Zapojení výzkumné komunity ČR

Vědci budují zázemí pro vědce



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

Pracovní skupiny EOSC CZ

- Hlavní komunita EOSC
- 12 pracovních skupin
- 4 průřezové: vstupy do NRP
- 8 oborových: vstupy do OS II
- Otevřené zájemcům (kdykoliv přijímá nové členy)
- Vědci pro vědce – tajemníci PS
- Fungování po celou dobu trvání iniciativy

Pracovní skupiny EOSC CZ

Praktické fungování PS EOSC CZ

- Pravidelné schůzky
- Záписy a záznamy
- Otevřené členství na bázi dobrovolnosti

Proč se zapojit?

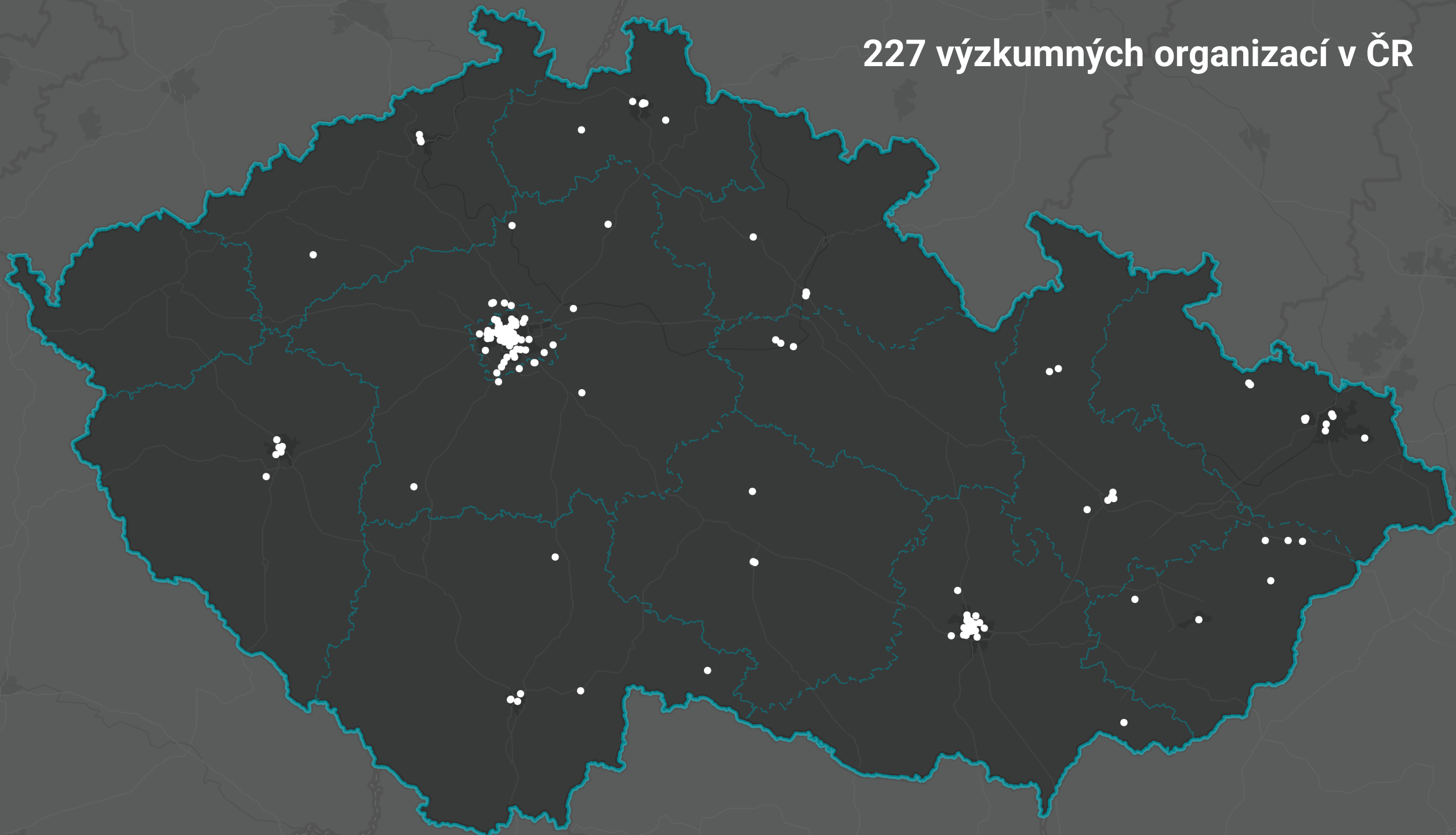
- Odbornost a informovanost
- Budování kontaktů
- Možnost ovlivňovat dění

Máte zájem stát se členem?



<https://www.eosc.cz/pracovni-skupiny>

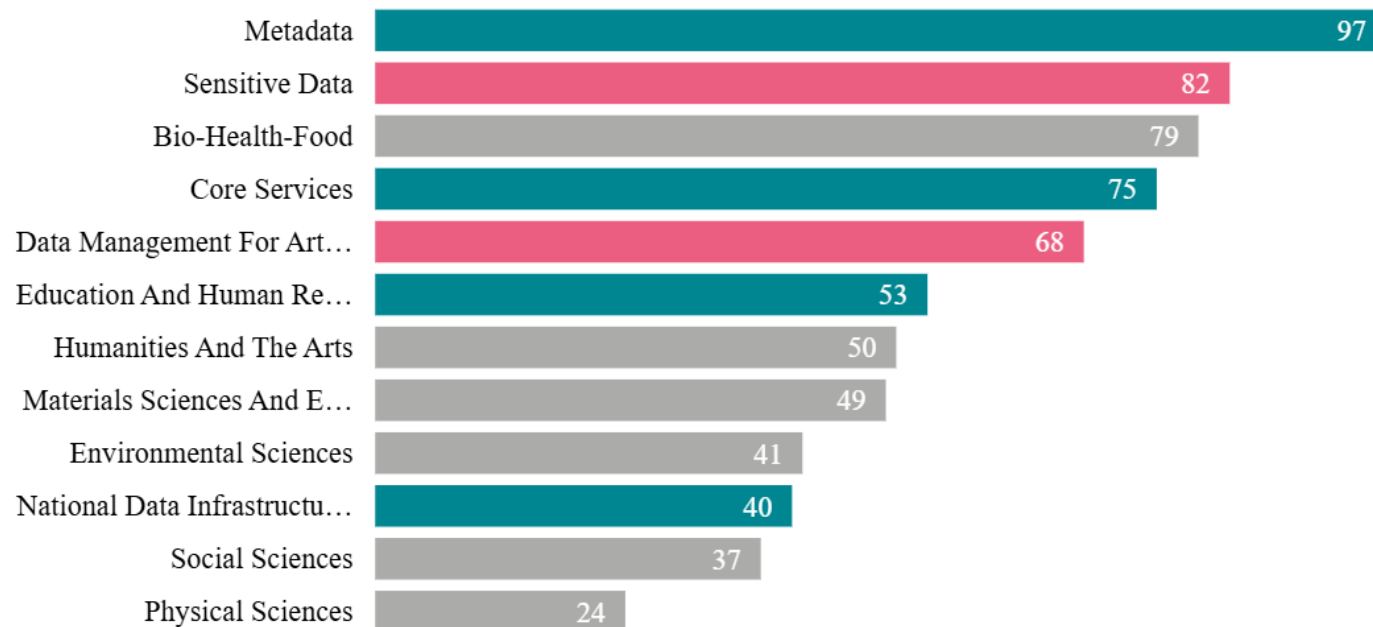
227 výzkumných organizací v ČR



227 výzkumných organizací v ČR
71 zapojeno do PS EOSC CZ
380 členů pracovních skupin

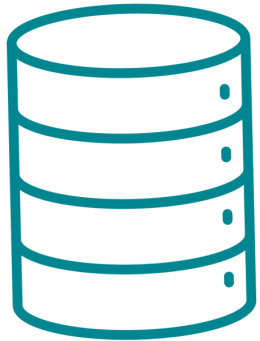
Zastoupení členů v pracovních skupinách

● průřezová ● specifická ● tematická



Národní datová infrastruktura

NDI



Úložné kapacity

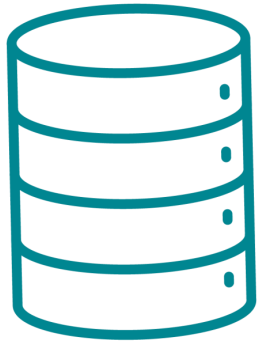


Nástroje a služby



Výpočetní kapacity

Praktické výstupy



Úložné kapacity



Nástroje a služby

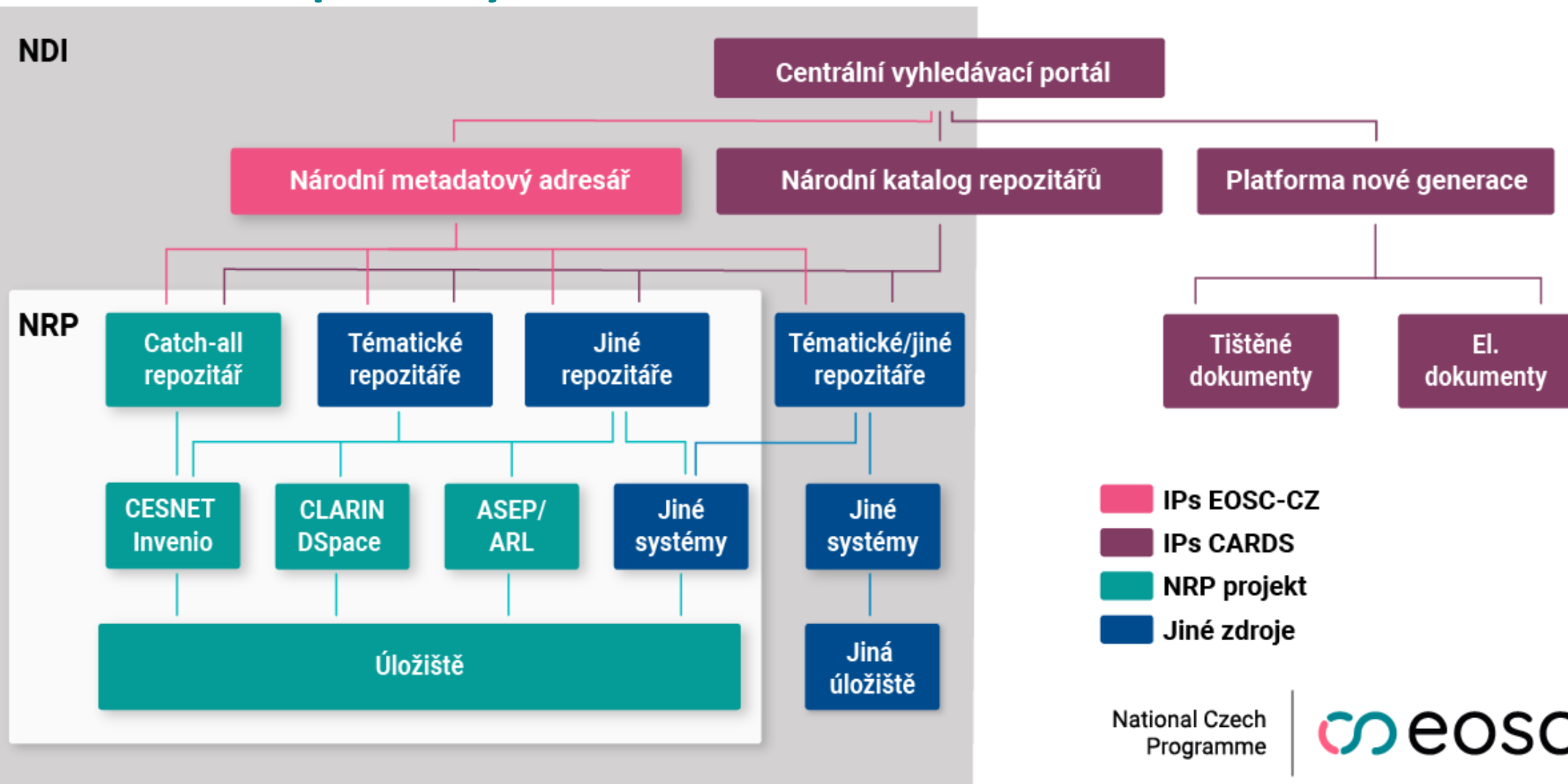


Výpočetní kapacity

Úložné kapacity

NDI

NRP



Úložné capacity

Hardware

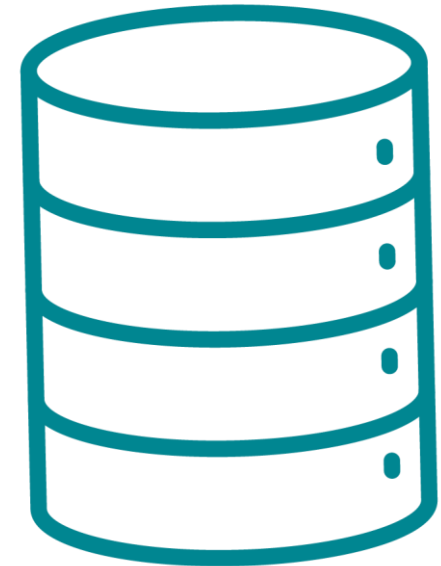
- Fyzická, distribuovaná úložná infrastruktura
- Uživatelská kapacita = 50 PB ve 3 kopiích

Repozitáře

- **Oborové (tematické) repozitáře**
 - 4 pilotní: Databáze molekulární biofyziky, Národní úložiště pro biodiverzitní data, Repožitář pro biologická zobrazovací data, ArchaeoVault
- **Catch-all repozitář**
- **Národní metadatový adresář**

Repozitářové systémy

- CESNET Invenio, CLARIN-DSpace, ASEP-ARL



Pilotní repozitáře

 Documentation Tutorials Deposit sign in with 



Molecular Biophysics Database

 All records 

Example: Lysozyme, NaCl, K_D, homo sapiens, xyhds-adf3t (record id) [Advanced search](#)



Molecular Biophysics Database collects raw data produced in experiments with biomolecular samples, biological material or other material, using molecular biophysics methods, such as Microscale Thermophoresis (MST), Biolayer interferometry (BLI), Surface Plasmon Resonance (SPR) and others.

MBDB development is supported by the project **MOSBRI** - Molecular Scale Biophysics Research Infrastructure of the European Commission, no. 101004806

Currently supported techniques

 MST
Microscale thermophoresis

 BLI
Bio-layer interferometry

 SPR
Surface plasmon resonance

 ITC
Isothermal titration calorimetry

<https://mbdb-data.org/>

Czech Repository of Herbarium Specimen Images





EOSC CZ

This repository is a part of the Czech National Repository Platform [NRP](#) [EOSC CZ](#)



Interoperability

Designed and maintained with interoperability support for [JACO](#) and [GBIF](#) as global aggregators.



Documentation

Please read the [documentation](#) to understand how the repository imports and publishes digital images.



Co-funded by
the European Union



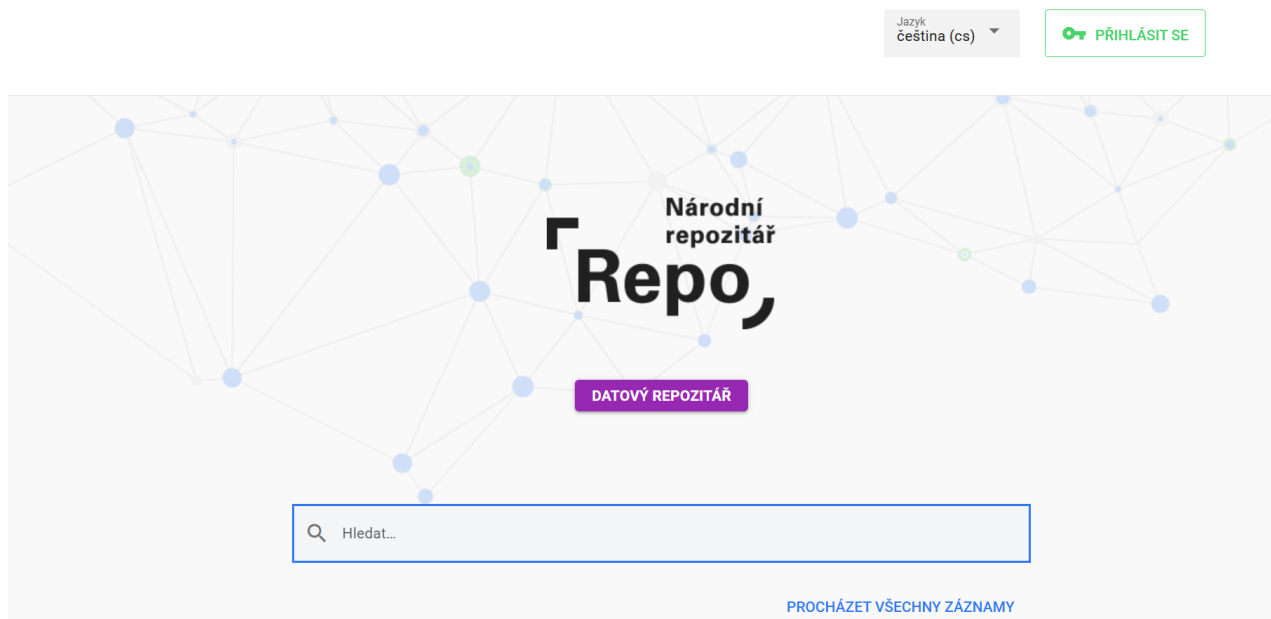
MINISTRY OF EDUCATION,
YOUTH AND SPORTS

National Czech
Programme

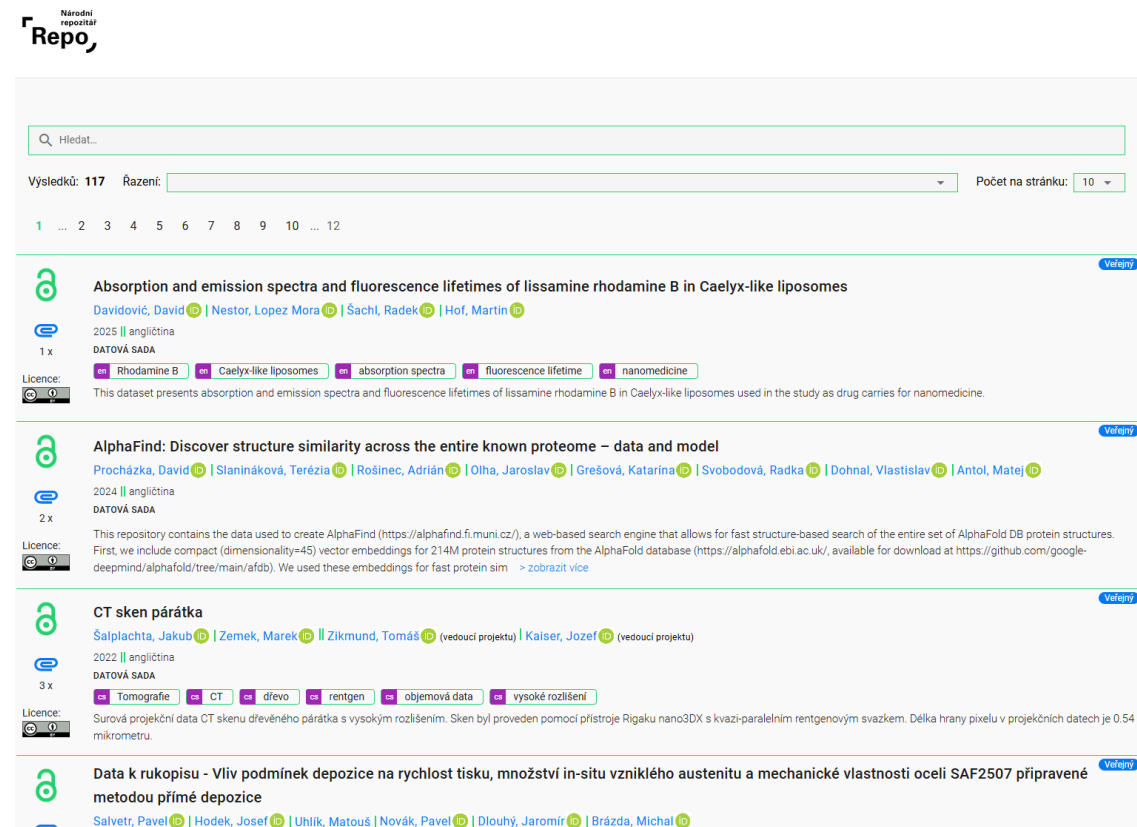


<https://herbarium.biodiversity.cz/>

Catch-all repozitář



<https://data.narodni-repozitar.cz/>



ČESKÉ DATOVÉ REPOZITÁŘE

ZOOLOGY



ENVIRO



BIOMONITORING



EXPOSOME



AI



IMAGING



SIMULATIONS



PHENOTYPE



BIOPHYSICS



SOCIAL



MAPS AND DATA



SENSITIVE SOCIAL DATA



REPO CZ



ISOARCH



PLANTS



AGROFOOD



BIO/CHEM



OMICS



FUNGI



MATERIALS



LINGUISTICS



BIOIMAGING



BIODIVERSITY



ARCHAEOVAULT



DUNE



LASER-PHYSICS



OBSERVATORY



CRYSTALOGRAPHY



PLASMA



ARTS



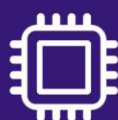
MOLECULAR BIOPHYSICS



LAUEDB



SENSORS



SOLAR



TELESCOPE



CORPUS



BIBLIOGRAPHY



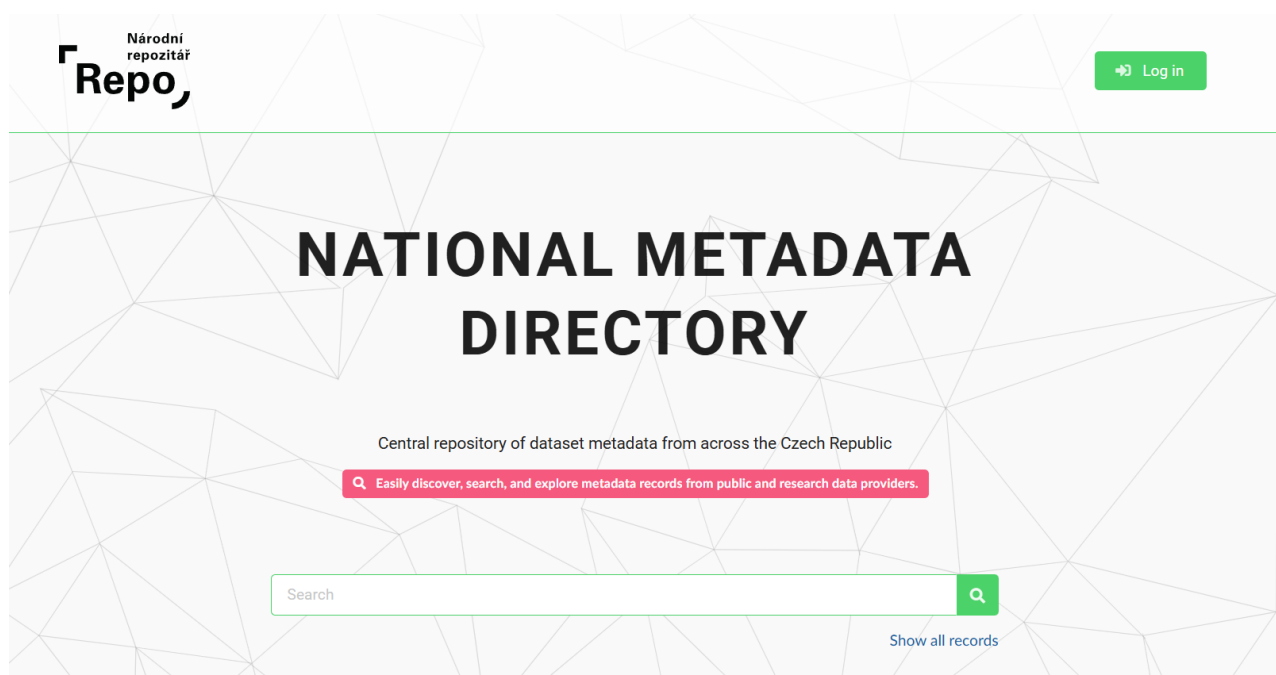
ARCHAEO-MAPS



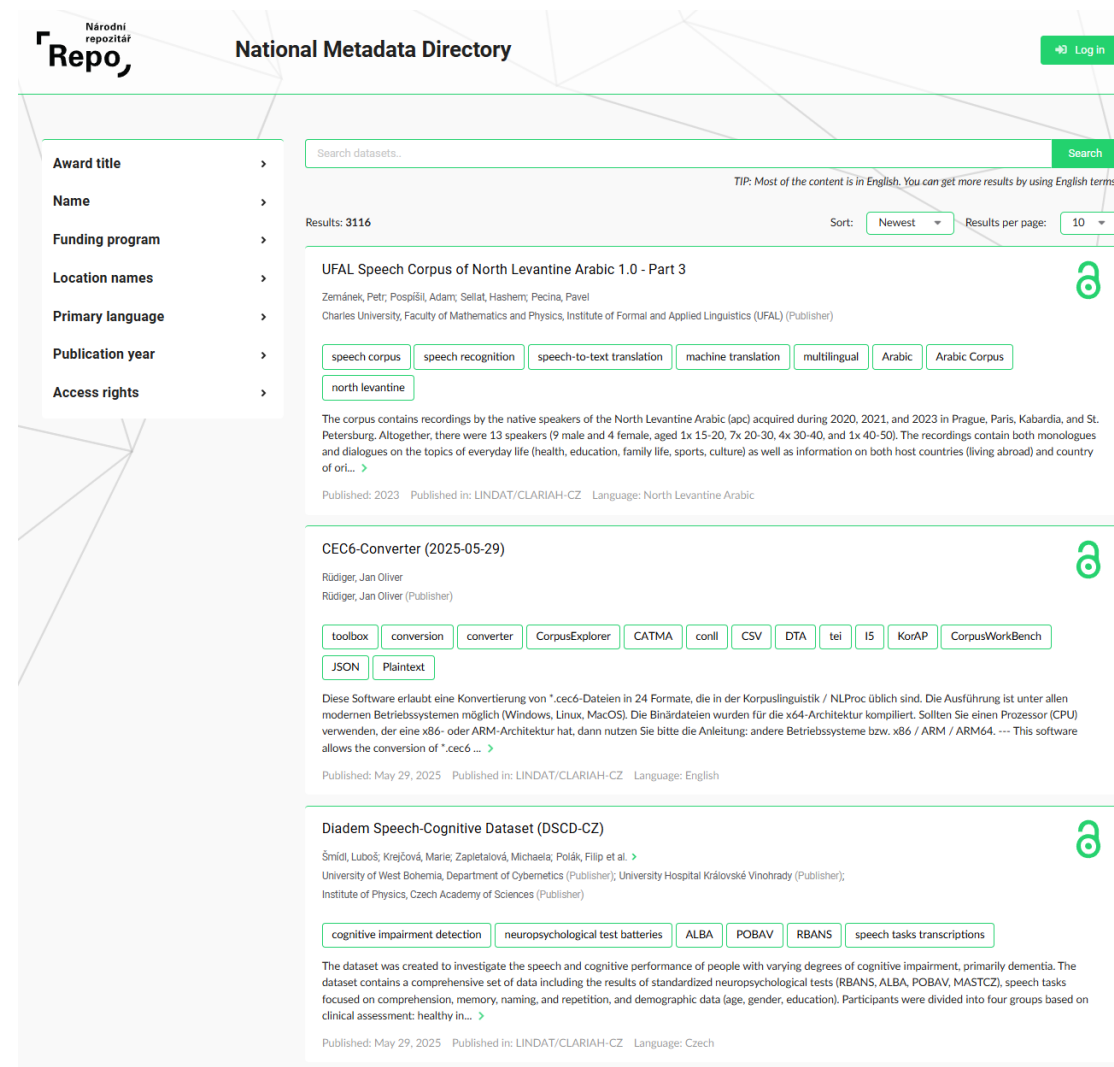
ARCHAEOLOGY



Národní metadatový adresář

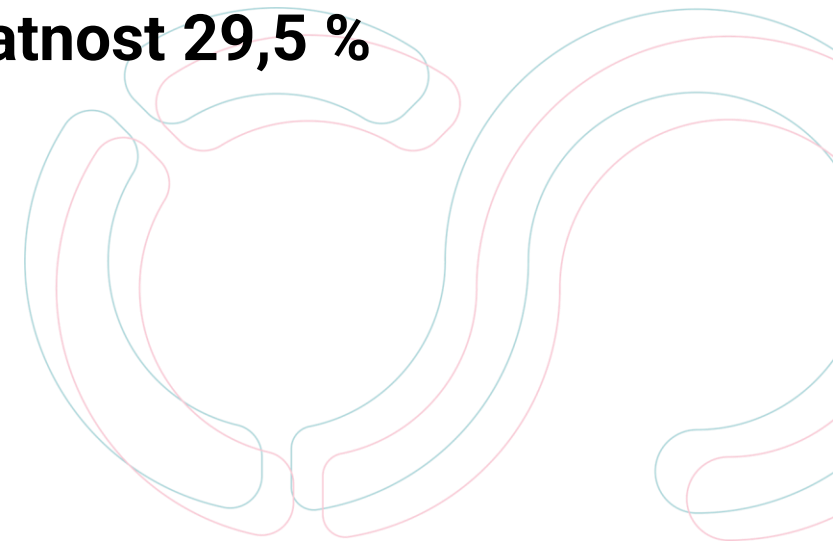


<https://nma.eosc.cz/>



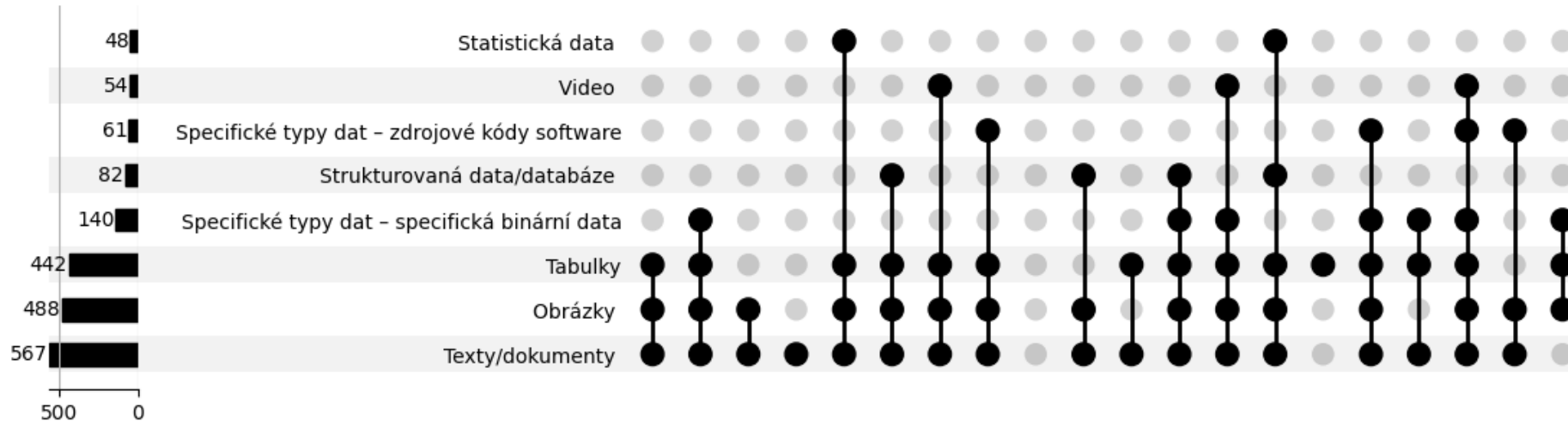
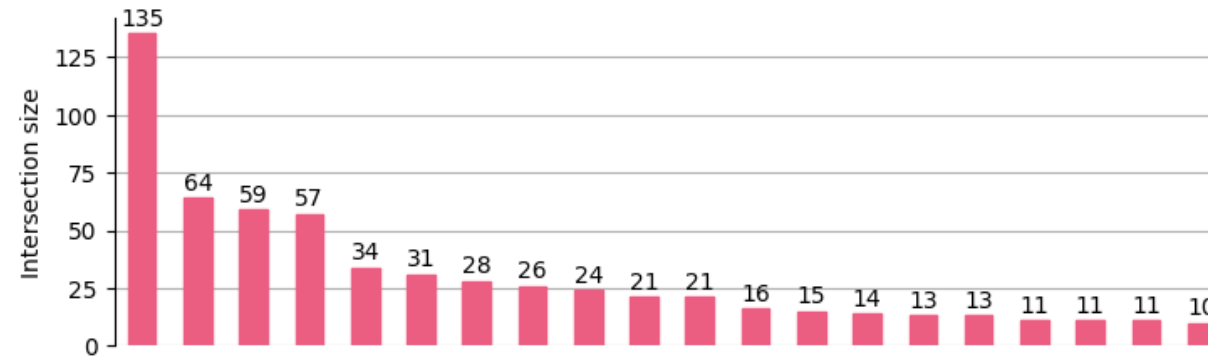
Národní průzkum „Správa výzkumných dat“

- Cíl: Zjistit stav práce s výzkumnými daty v ČR
- Osloveno 3 796 **hlavních řešitelů** výzkumných projektů
- Odpovědělo 1 121 respondentů (hlavních řešitelů); **návratnost 29,5 %**
- 50 otázek



Jaká data v projektu generujete?

N=1121



Jaký objem dat jste v rámci výzkumu vygenerovali?

N=1121

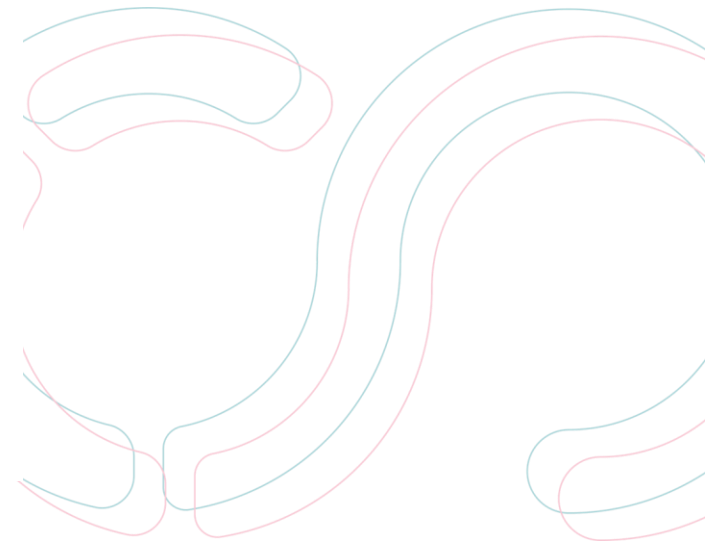
	Počet	Podíl (%)
Méně než 1 GB	233	21,1
<u>1 – 10 GB</u>	260	23,6
10,1 – 100 GB	219	19,8
100,1 GB až 1 TB	144	13,0
1,1 TB – 10 TB	93	8,4
10,1 TB – 100 TB	42	3,8
Více než 100 TB	12	1,1
Nevím	101	9,1
Celkem	1104	100,0

Metadata

- **59 % respondentů** opatřuje data metadaty
 - **7,2 %** Používání standardizovaných schémat
 - **36,3 %** Textový soubor readme
 - **23,2 %** Metadata z přístrojů
- **41 % respondentů** metadata nepoužívá
 - **6,2 %** Nevíme, jak na to
 - **2,4 %** Nemáme potřebné technické vybavení
 - **10,7 %** Nemáme kapacity (lidské, finanční)
 - **28,8 %** Pro projekt to není důležité

Očekávaná doba uložení dat

	Přírodní vědy	Technické vědy	Lékařské vědy	Zemědělské a veterinární vědy	Společenské vědy	Humanitní vědy
Do 3 let	1,6	7,7	2,8	10,9	4,5	5,3
<u>3 – 5 let</u>	14,3	22,6	22,2	21,8	20,9	10,6
<u>6 – 10 let</u>	20,5	23,2	21,3	21,8	13,6	5,3
Více než 10 let	18,9	11,9	28,7	12,7	10,0	9,7
Trvale	43,0	28,6	25,0	32,7	41,8	59,3
Nechceme / nepotřebujeme data uchovávat	1,6	6,0	0,0	0,0	9,1	9,7



Praktické výstupy



Úložné kapacity



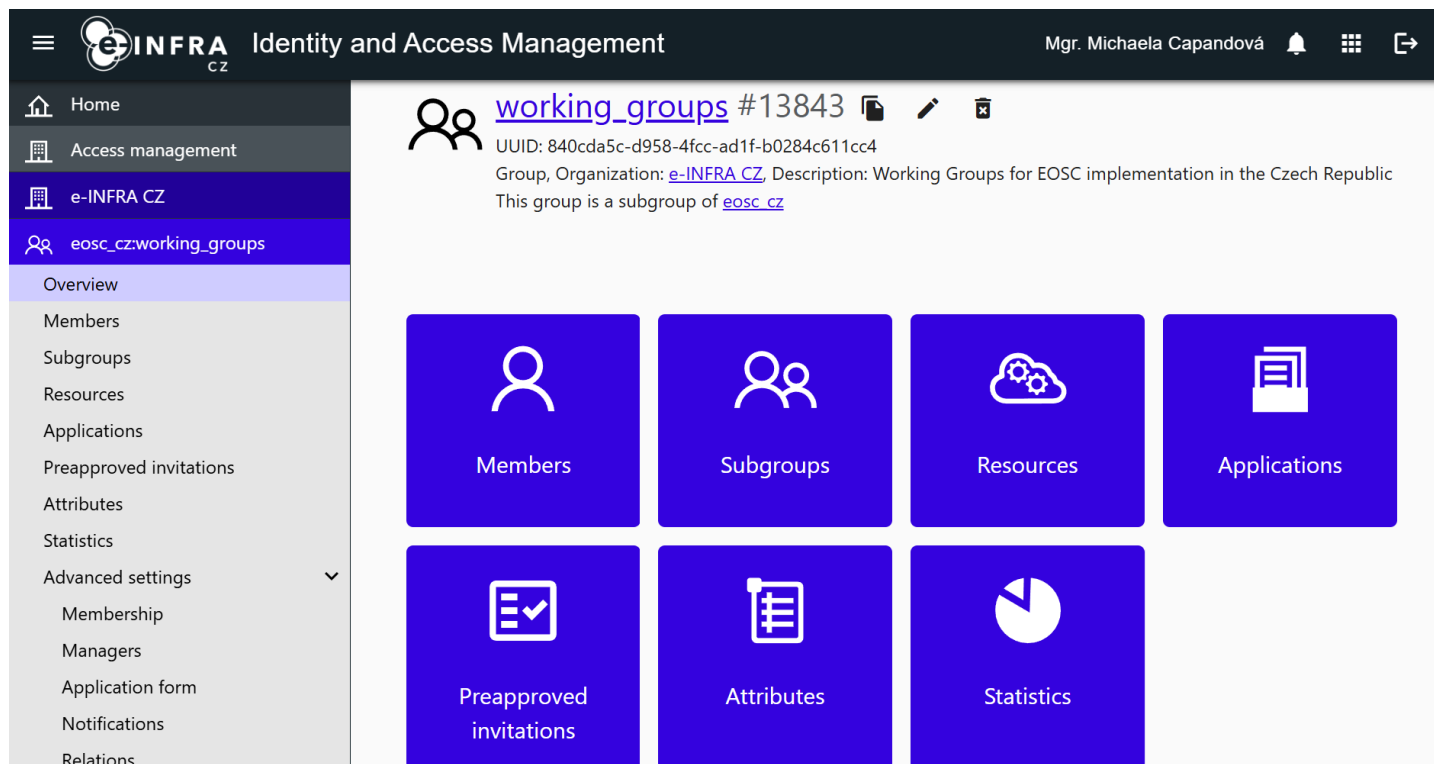
Nástroje a služby



Výpočetní kapacity

Autentizační a autorizační infrastruktura (AAI)

Umožňuje uživatelům z různých institucí jednoduchý přístup k datům a službám.

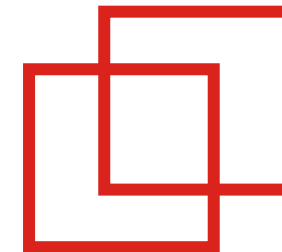


- Řízení přístupu a identit
- Řízení skupin a jejich rolí
- Přidělování práv



<https://perun-aai.org/>

Identifikátory CZ



 **identifikátory.cz**

Persistent Identifiers

[Persistent Identifiers](#) ▾ [Services](#) ▾ [About us](#) [News](#)



[Home](#) / [Persistent Identifiers](#)

Persistent Identifiers

Learn more about each persistent identifier (PID). Persistent identifiers are tools that are used to uniquely identify people, organisations, and other objects (e.g., books, articles, datasets) in a scholarly communication system.

**ORCID iD for
researchers**

DOI for objects

ISBN for books

ISSN for periodicals

**ISMN for notated
music**

**ROR for
organizations**

IGSN for samples

Other PIDs



SensitiveCloud



Zabezpečené prostředí pro ukládání, sdílení a zpracování dat

- Primárně určeno pro práci s vlastními citlivými daty
- Postupné rozšířené pro podporu řízeného sdílení dat
(hlavní technický prvek pro práci s citlivými daty v rámci NDI)
- Zahrnuje úložiště, výpočetní zdroje, podpora hotových webových aplikací

S jakými typy citlivých dat v projektu pracujete? N=304

	Ano
Osobní údaje (např. jméno, rodné číslo, zdravotní stav, biometrické údaje)	20,0 % 224
Obchodní tajemství (např. informace o klientech, strategické plány, finanční dokumenty)	6,0 % 67
Utajovaná data (např. data týkající se bezpečnosti státu)	1,2 % 13

Data Stewardship Wizard

Nástroj na plánování správy výzkumných dat (DMP), použití NRP instance zdarma



DS Wizard

Dashboard

Projects

Storage Costs Evaluator

DSW Guide

Test 1

Share

Questionnaire Metrics Preview Documents Settings

View Import replies

TODOs 1 Comments Version history

Current phase

Before Submitting the Proposal

Chapters

I. Administrative information 1

II. Re-using data 1

III. Creating and collecting ... 5

IV. Processing data ✓

V. Interpreting data ✓

VI. Preserving data 2

VII. Giving access to data 3

IV. Processing data

In the processing phase, the data will be undergoing the mostly automated steps for processing, before the analysis and interpretation.

In this chapter, many questions are focusing on the compute environment that is used to process the data and make it available for interpretation by project partners. Some of those questions (e.g. on workflow systems and data provenance) are also relevant for the work in the interpretation phase.

✓ IV.1 Will you be using a shared working space to work with your data?

Horizon Europe DMP

Will you be using a working space containing data and software specific to the project that is shared between all the people working on the data in the project? Sometimes such a system is called a *Virtual Research Environment*.

☒ Desirable: Before Submitting the Proposal

☐ a. No, participants in the project each have different collections of data and tools

☒ b. Yes

Clear answer

Answered 2 minutes ago by Antonin Zita.

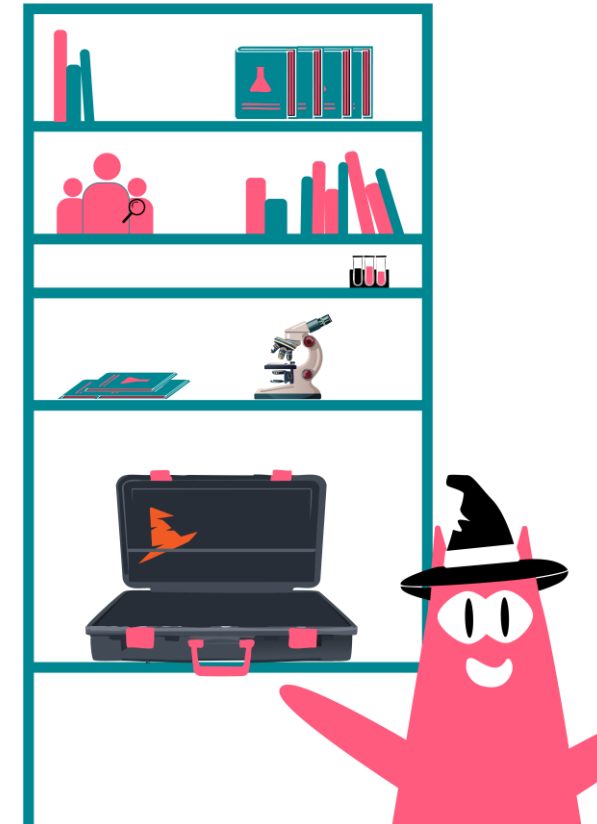
⤴ Collapse

⌛ IV.1.b.1 How will access to the work space be controlled?

Horizon Europe DMP

☐ a. Only all project members have read/write access to the data

☐ b. Anyone in the institute has read access to the data; all project members have read/write access.



Praktické výstupy:



Úložné kapacity

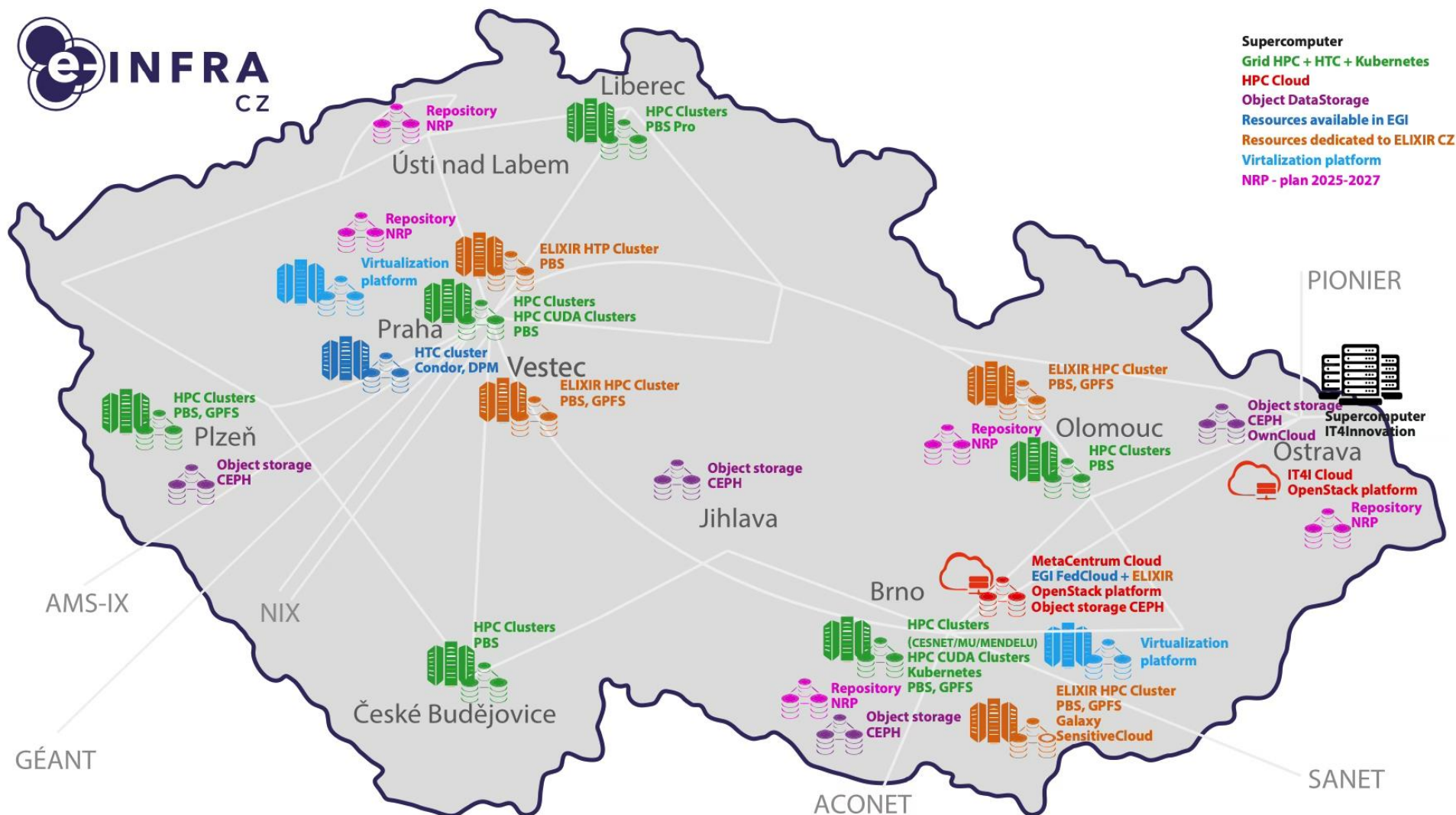


Nástroje a služby



Výpočetní kapacity

Zázemí velké výzkumné infrastruktury e-INFRA CZ



Gridová infrastruktura


metacentrum


a

Superpočítačové centrum

 **VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA**

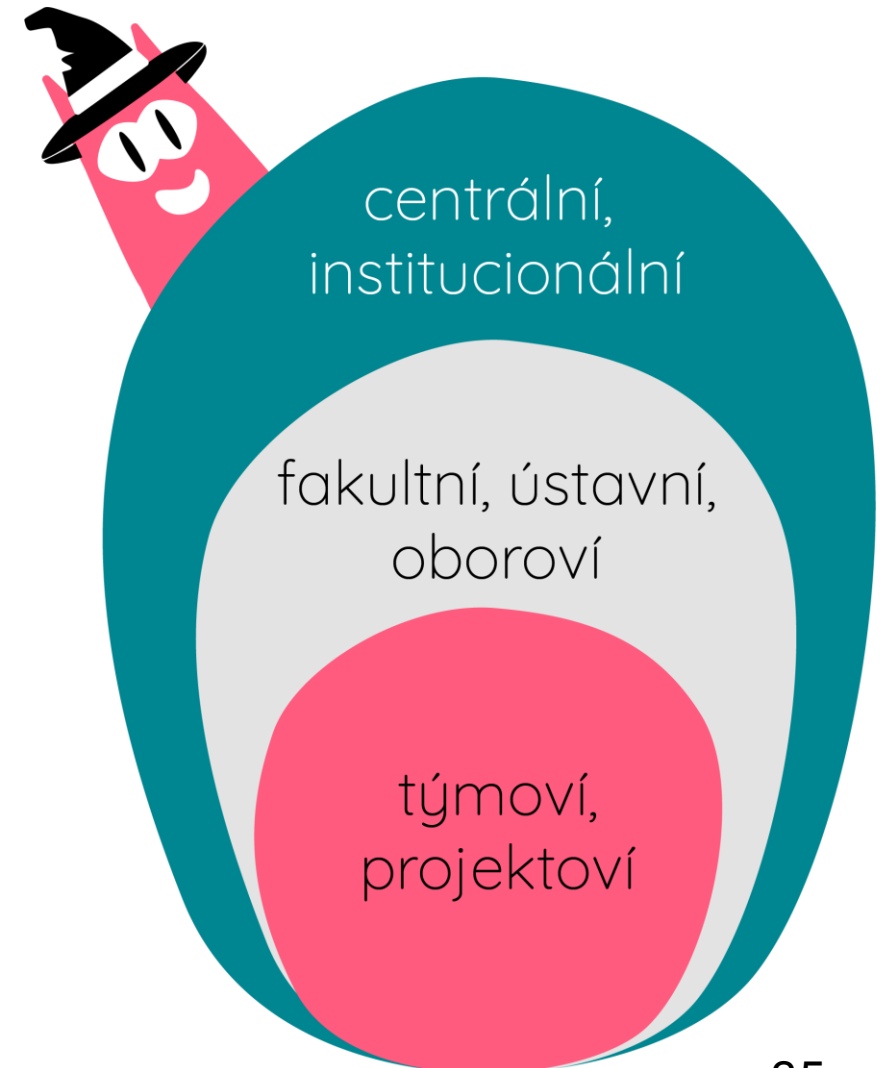
**IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ
CENTRUM**

Podpora práce s daty

Data stewardi

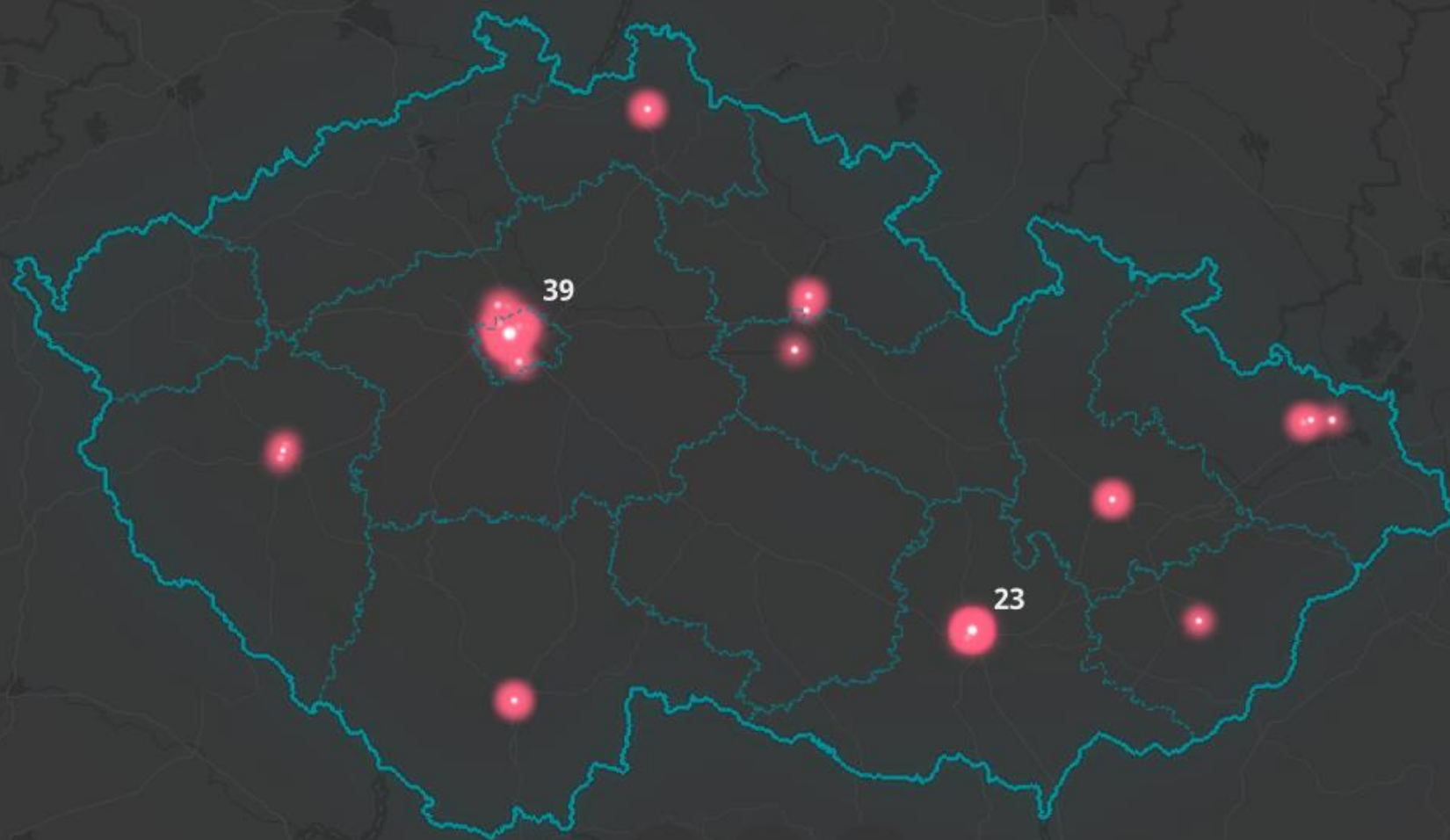
- Nový typ pozice zaměřené na **podporu správy výzkumných dat**
- Odpovědnost za nastavení a dodržování **osvědčených postupů při práci s daty**
- Zajištění **integrity a bezpečnosti dat** a dodržování **FAIR zásad**
- **Pojítka** mezi výzkumem, infrastrukturou a správou dat

Typy data stewardů



Mapa data stewardů v ČR

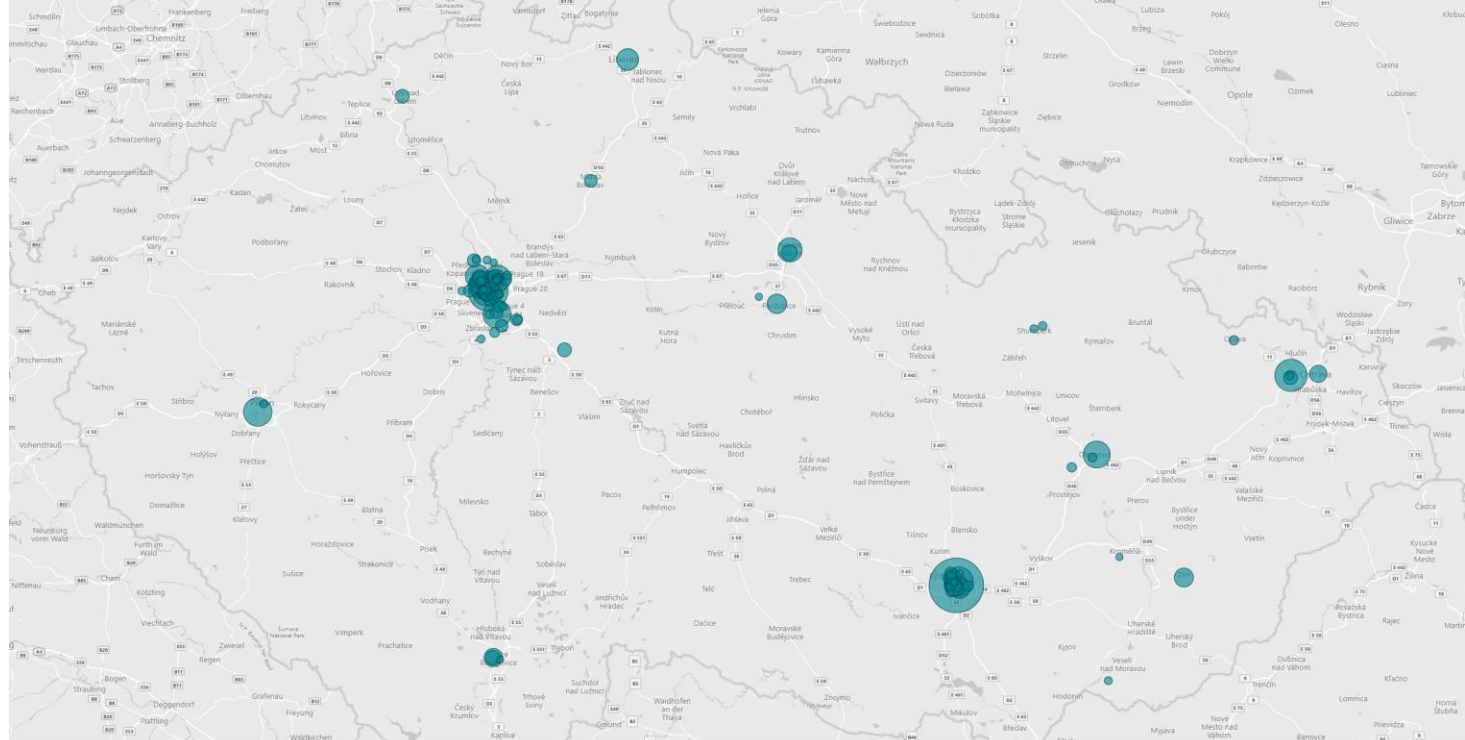
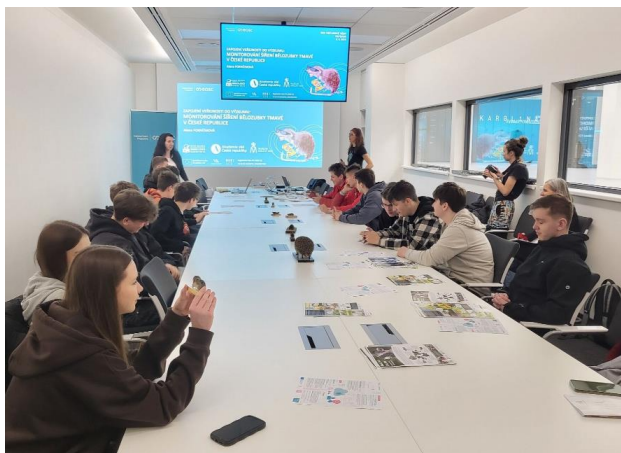
Institute NEBO jméno a příj...



Vzdělávání

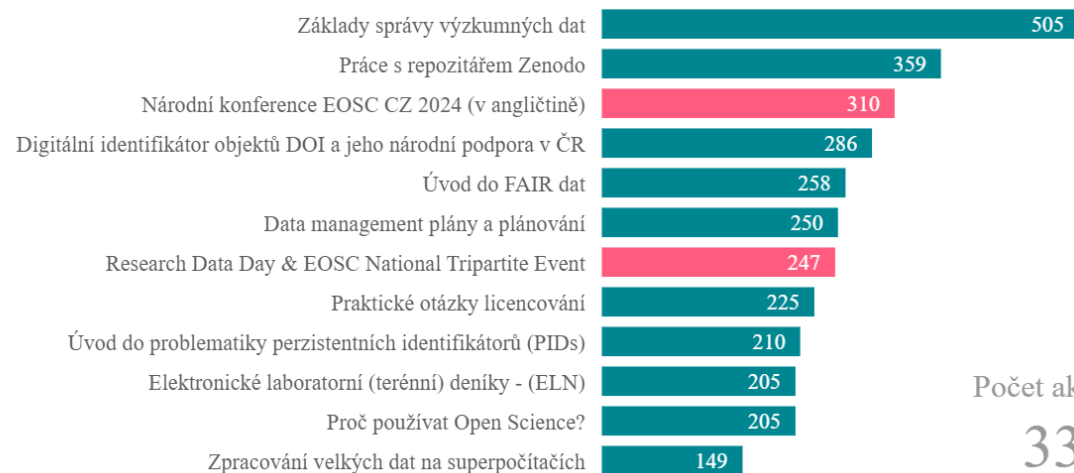
Stats

- Školení obsahově pokrývají témata související s **životním cyklem dat**
- 33 školení, 4 konference
- 5 000 registrací, 127 institucí



Počet registrací na jednotlivých aktivitách

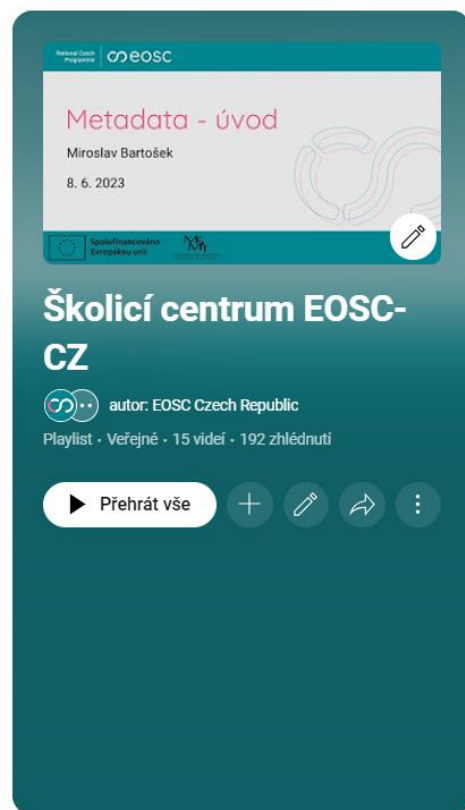
● konference ● seminář ● školení



Počet aktivit

33

Databáze školení na YouTube a webu



Metadata - úvod

EOSC Czech Republic • 450 zhlédnutí • před 1 rokem



Úvod do metadat II - metadaty v kontextu NRP

EOSC Czech Republic • 221 zhlédnutí • před 1 rokem



Úvod do organizace znalostí: řízené slovníky, ontologie

EOSC Czech Republic • 166 zhlédnutí • před 1 rokem



Systémy pro Národní repozitářovou platformu (NRP)

EOSC Czech Republic • 254 zhlédnutí • před 10 měsíci



Pilotování Data Stewardship kurzu pro doktorandy

EOSC Czech Republic • 119 zhlédnutí • před 1 rokem

[Záznamy školení](#)



Školení podle životního cyklu dat



Přehledové a úvodní kurzy



Plánování a příprava dat



Sběr dat



Zpracování dat



Analýza dat



Ukládání dat



Sdílení dat



Znovuvyužití dat



Ostatní

<https://www.eosc.cz/skoleni>

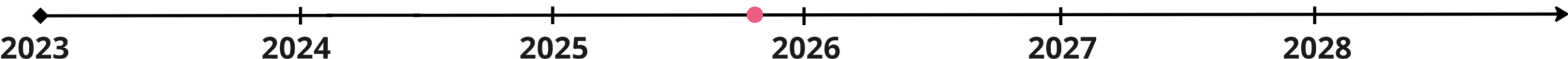


Co bude následovat

Konec první fáze

Fáze 1

- **Zapojení českých výzkumníků** a odborníků
- **Organizace pracovních skupin** a komunit
- **Kick-off třech hlavních projektů**
- **Navržení** základní infrastruktury a služeb
- **Spuštění prvních** služeb
- **Zvýšení povědomí**



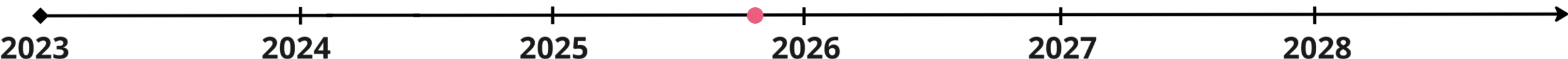
Přechod do druhé fáze

Fáze 1

- Zapojení českých výzkumníků a odborníků
- Organizace pracovních skupin a komunit
- Kick-off 3 hlavních projektů
- Navržení základní infrastruktury a služeb
- Spuštění prvních služeb
- Zvýšení povědomí

Fáze 2

- Spuštění infrastruktury
- Spuštění všech základních služeb
- Integrace prvních repozitářů
- Podpora data management dovedností a know-how
- Integrace národního a mezinárodního ekosystému
- Návrh modelu udržitelnosti



Důležité odkazy a kontakty

[Web EOSC CZ](#)



[Newsletter iniciativy EOSC CZ](#)



- Máte nějaké otázky?
 - info@eosc.cz
- Máte nápad na přednášku nebo školení?
 - events@eosc.cz
- Chcete kontaktovat naše komunikátory?
 - pr@eosc.cz



@EOSC Czech Republic



@eosccz.bsky.social

Děkuji za pozornost

Martin Dvořák

dvorak@ics.muni.cz

<https://www.linkedin.com/in/madvorak/>



Spolufinancováno
Evropskou unií



MUNI
ICS



VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA

IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ
CENTRUM