



# Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

# Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

Autoři: Martin Fařun, Ondřej Daniel

8. 7. 2022

(revize 2, 27. 7. 2022)

## Technologické centrum AV ČR

Projekt sdílených činností „Strategická inteligence pro výzkum a inovace“ (MS2104)

Technologické centrum AV ČR je neziskové zájmové sdružení právnických osob, které není pracovištěm Akademie věd ČR podle zákona č. 283/1992 Sb. a není financováno z rozpočtu Akademie věd ČR. Výstupy Technologického centra AV ČR obsahují nezávislé expertní názory a nevyjadřují tak oficiální postoj Akademie věd ČR ani jednotlivých členů sdružení.

## Obsah

1	Stručné shrnutí .....	4
2	Úvod .....	5
3	Evropský rámec EOSC a zapojení ČR .....	5
4	Rozvoj EOSC v EU.....	7
4.1	Vývoj evropských iniciativ .....	7
4.2	Aktuální stav .....	10
4.3	Situace a praxe ve vybraných zemích.....	11
4.3.1	Příklady států v pokročilé fázi implementace EOSC .....	12
4.3.2	Příklady států v počáteční fázi implementace EOSC .....	16
5	Předpoklady pro implementaci EOSC v ČR.....	18
5.1	Průzkum vědeckých repositářů v ČR .....	19
5.2	Přehled vybraných rozvinutých repositářů a infrastruktur .....	19
5.3	Rámec a aktuální stav přípravy národní datové infrastruktury EOSC .....	21
5.4	Předpoklady pro efektivní práci s metadaty výzkumných dat .....	23
5.5	Průzkum stavu a očekávání v oblasti archivace a zveřejňování vědeckých dat mezi členy pracovních skupin Architektura a Základní služby EOSC.....	24
5.6	Návrh typologie pracovních pozic pro implementaci EOSC .....	27
5.7	Hodnocení současného stavu a potřeb rozvoje – shrnutí rozhovorů s experty.....	29
6	Shrnutí a závěry .....	33
7	Přílohy.....	36
7.1	Klíčové principy pro implementaci a provoz EOSC.....	36
7.1.1	Reflexe principů FAIR.....	36
7.1.2	EOSC Interoperability Framework.....	38
7.1.3	Perzistentní identifikátory.....	39
7.1.4	Data management plány .....	39
7.2	Rozhovory s národními experty .....	40
7.3	Slovníček pojmů implementace EOSC v ČR .....	57
7.4	Seznam zkratk .....	59
8	Reference.....	61

## 1 Stručné shrnutí

Analýza shrnuje současný stav přípravy národní datové infrastruktury EOSC (European Open Science Cloud) a výsledky průzkumů realizovaných v nedávné době ohledně fungování vědeckých repozitářů a práce s vědeckými daty a informacemi. Dále se opírá o sadu strukturovaných rozhovorů s uživateli, experty a stakeholdery zapojenými do rozvoje EOSC, zaměřených na zjištění jejich potřeb a očekávání. Součástí studie je také rešerše stavu implementace EOSC v šesti vybraných zemích EU.

Stejně jako v jiných státech EU ani v ČR EOSC nevzniká na zelené louce. Existují zde pracoviště a infrastruktury, které jsou z hlediska FAIR práce s vědeckými daty na velmi pokročilé úrovni. Na druhém konci spektra je nicméně řada týmů a institucí, které mají ohledně FAIR práce s daty jen minimální povědomí, zkušenosti a motivaci. Výchozím předpokladem je proto **důsledná propagace myšlenky a výhod EOSC** a principů FAIR dat směrem k široké vědecké obci, tj. i do regionů mimo velká centra výzkumu v Praze, Brně, Ostravě nebo Olomouci.

**Navrhovaná výše podpory prostřednictvím programu OP JAK** se jeví v mezinárodním kontextu v porovnání se skupinou států v pokročilé fázi implementace EOSC (Německo, Nizozemsko, Francie), jako odpovídající a rovněž návrh struktury Národní datové infrastruktury se v Česku ubírá podobným směrem jako v těchto pokročilých zemích.

Provedené průzkumy i expertní rozhovory ukazují, že pro implementaci evropské politiky zacházení s vědeckými daty a širší přijetí FAIR principů je nezbytný koordinovaný přístup přes vědecké skupiny a disciplíny, který zahrnuje jak **společné budování federované infrastruktury pro ukládání anotovaných dat<sup>1</sup>**, tak výchovu a vzdělávání specialistů v této oblasti.

Je nezbytné od počátku zavádět postupy **užívání standardizovaných metadatových schémat**. Bez kvalitního popisu obsahu datové sady dokáže data najít jen ten, kdo zná autora nebo číslo projektu, ale zájemce nemá možnost data najít podle toho, k jakému tématu se vztahují.

Podstatná část zamýšlené podpory by měla být použita především na **vzdělávání a rozvoj lidských zdrojů** v oblasti práce s vědeckými daty a datovými sadami. Vzdělávání by mělo být zaměřeno na celoživotní výchovu vědců ke správnému nakládání s vědeckými daty. Jako druhý pilíř aktivit v této oblasti lidských zdrojů se jeví výchova datových specialistů s různými stupni kvalifikace.

Závěry z provedených průzkumů a expertních rozhovorů dále ukazují, že je účelné **využít již existujících řešení v rámci pokročilých oborově-vědních a institucionálních klastrů** a podpořit jejich integraci do Národní datové infrastruktury. Vzhledem ke stále majoritnímu počtu vědců a vědeckých týmů nezapojených do stávající struktury repozitářů je dalším důležitým krokem **vytvoření Národní repozitářové platformy jako multi-tenant platformy**, která umožní jednoduché zakládání a práci s asociovanými repozitáři institucím, oborově-vědním skupinám i dalším zájemcům bez ohledu na jejich vlastní znalosti technických principů péče o repozitáře a jejich zázemí.

Důležitá je také **specifická aktivita v oblasti citlivých dat napříč různými obory**, zaměřená na specifické technické požadavky z hlediska zabezpečení, specifické právní požadavky, a také specifické požadavky ze strany uživatelů, případně i státu, který citlivá data poskytuje.

V neposlední řadě je třeba v ČR vytvořit motivační prostředí pro **zapojení nových zájemců do EOSC Association Task Forces a mezinárodních projektů**, což zajistí kontinuální napojení na celoevropské aktivity a přinese další finanční prostředky pro budování EOSC v ČR.

---

<sup>1</sup> Tj. vědeckých dat s připojeným (nejlépe standardizovaným) metadatovým popisem.

## 2 Úvod

Studie byla zpracována týmem Technologického centra AV ČR v rámci projektu sdílených činností Strategická inteligence pro výzkum a inovace (MS2104) jako expertní podklad pro nastavení výzev OP JAK určených pro podporu implementace EOSC v ČR.

**Zapojení ČR do evropského rámce EOSC** je popsáno v kapitole 3. Dále studie sestává ze dvou hlavních součástí – rešerše stavu rozvoje EOSC v EU (kapitola 4) a analýzy předpokladů pro implementaci EOSC v ČR (kapitola 5).

**Rešerše stavu rozvoje EOSC v EU** (kapitola 4) se zabývá stavem implementace EOSC v zemích EU a zejména způsobem poskytování podpory na rozvoj EOSC z evropských i národních zdrojů. Součástí rešerše je podrobnější popis implementace EOSC na národní úrovni v šesti vybraných evropských zemích, z nich tři se již nacházejí v pokročilé fázi implementace EOSC (Německo, Nizozemsko, Francie) a další tři, velikostně srovnatelné s ČR, v počáteční fázi implementace EOSC (Rakousko, Portugalsko, Švédsko). Úkolem této části studie je poskytnout informace o tom, jak probíhá implementace EOSC jinde, nabídnout případnou inspiraci a umožnit porovnání situace v ČR a jiných zemích z hlediska předpokládané finanční podpory a obsahového zaměření implementace EOSC.

**Analýza předpokladů pro implementaci EOSC v ČR** (kapitola 5) shrnuje současný stav přípravy národní datové infrastruktury EOSC a výsledky průzkumů realizovaných v nedávné době ohledně fungování vědeckých repozitářů a práce s vědeckými daty a informacemi. V rámci přípravy této studie byla rovněž realizována sada strukturovaných rozhovorů s uživateli, experty a stakeholdery zapojenými do rozvoje EOSC, jejichž potřeby a očekávání jsou v této kapitole také shrnuty. Podrobnější výtahy ze všech realizovaných rozhovorů jsou k dispozici jako příloha 7.2 této studie.

Příloha 7.1 obsahuje klíčové principy pro implementaci a provoz EOSC. Cílem rešerše je upozornit na evropské know-how ve vybraných aspektech implementace EOSC, které jsou významné z hlediska požadované funkčnosti, vynaložených nákladů a dlouhodobé udržitelnosti implementace EOSC na národní úrovni, a měly by proto být v rámci připravovaných projektů akcentovány.

## 3 Evropský rámec EOSC a zapojení ČR

Ambicí **Evropského cloudu pro otevřenou vědu (European Open Science Cloud, EOSC)<sup>2</sup>** je vytvořit celoevropskou webovou strukturu pro FAIR<sup>3</sup> přístup k datům a službám pro vědu v Evropě. EOSC je koncipován jako multidisciplinární prostředí, kde mohou výzkumní pracovníci ukládat, zveřejňovat, vyhledávat a opakovaně využívat data, nástroje a služby, což jim umožní lépe provádět jejich práci.

EOSC vychází ze stávající infrastruktury a služeb podporovaných Evropskou komisí (EK), členskými státy, organizacemi a výzkumnými komunitami. Sdružuje je v rámci federativního přístupu do "systému systémů" a přidává hodnotu tím, že prostřednictvím metadatového popisu agreguje jejich obsah a umožňuje společné využívání dat i služeb.

---

<sup>2</sup> <https://www.eosc.eu/>

<sup>3</sup> FAIR jsou metodickým souborem pokynů pro publikování výzkumných dat. Podle FAIR principů by digitální data a jejich metadata měla být dohledatelná (Findable), přístupná (Accessible), čitelná a strojově zpracovatelná (Interoperable) a znovupoužitelná (Reusable). Podrobněji viz <https://www.go-fair.org/fair-principles/> a kap. 7.1.1: Reflexe principů FAIR.

Očekává se, že EOSC zkvalitní činnosti výzkumných pracovníků v mnoha ohledech, mezi něž patří zejména:

- bezproblémový přístup k datům a službám, s řízením přístupu prostřednictvím společné autentizační a autorizační infrastruktury (AAI),
- přístup k údajům z různých zdrojů, které jsou FAIR a v ideálním případě otevřené,
- přístup ke službám pro ukládání, výpočty, analýzu, uchovávání a další práci s daty,
- přijetí standardů, aby bylo možné kombinovat data a služby,
- helpdesk, školení a podpora pro lepší využívání EOSC.

Implementace EOSC je založena na dlouhodobém procesu koordinace, který EK od roku 2015 rozvíjí s mnoha různými stranami spoluutvářejícími evropské výzkumné prostředí. V počáteční fázi investovala EK přibližně 250 milionů EUR do prototypových složek EOSC prostřednictvím výzev k předkládání projektů v rámci programu Horizont 2020 (viz tab. 1). EK rovněž zahájila prozatímní řízení EOSC s cílem připravit strategické směry pro implementaci EOSC po roce 2020.

Současná fáze implementace (2021-2030) probíhá v kontextu Evropského partnerství EOSC<sup>4</sup> a podle **Strategické výzkumné a inovační agendy EOSC (SRIA)**<sup>5</sup>. Rozvoj EOSC tak přechází na přístup více řízený zúčastněnými stranami se společnou vizí, společnými cíli a doplňujícími se příspěvky na evropské, národní a institucionální úrovni. Tento rozvoj je koordinován **na principu tripartity (EOSC Tripartite Governance)**<sup>6</sup>, do níž je zapojena EU zastoupená EK, zúčastněné země zastoupené v Řídícím výboru EOSC (EOSC Steering Board) a výzkumná komunita zastoupená v Asociaci EOSC (EOSC Association). V programovém období 2021-2027 se předpokládají společné investice (s věcnými a finančními příspěvky) ze strany EU a partnerů mimo EU v celkové výši cca 1 mld. EUR (viz kap. 4.2).

Spolupráce na mezinárodní úrovni probíhá především prostřednictvím **Asociace EOSC**<sup>7</sup> (se sídlem v Bruselu), mezinárodní neziskové organizace založené podle belgického práva v červenci 2020 (převzala gesci za řídicí struktury EOSC z minulých let). Za ČR jsou v asociaci jako členové zastoupeny: sdružení CESNET (MŠMT toto sdružení zároveň pověřilo rolí národní „pověřené organizace“, tzv. Mandated organisation), MU (CERIT-SC), VŠB – TUO (IT4Innovations) – přičemž všechny tři výše uvedené organizace spoluvytvářejí infrastrukturu e-INFRA CZ – a dále Národní technická knihovna (NTK). Jako „pozorovatel“ (observer) je zapojen Ústav fyzikální chemie J. Heyrovského AV ČR.

Strategické a implementační priority EOSC jsou formulovány a aktualizovány prostřednictvím průběžné práce pracovních skupin (**EOSC Association Task Forces**), v nichž jsou zastoupeni jednotlivci z různých zainteresovaných výzkumných organizací ČR.

Mezi příklady mezinárodních projektů na podporu EOSC, v nichž jsou zapojeny také subjekty z ČR, lze uvést EOSC-Life, PaNOSC, EOSC Synergy, EOSC Future, EOSC-Hub, EUROCC, EGI ACE, nově HPLT aj. Obecně jde zejména o projekty primárně zaměřené buď tematicky (klastrové sdružování evropských velkých výzkumných infrastruktur na podporu EOSC) nebo o projekty cílené na budování celoevropských univerzálních e-infrastruktur pro FAIR data.

Návrh architektury a příprava implementace národní datové infrastruktury EOSC v ČR (viz kap. 5.3) pak úzce navazuje na aktivity evropské Asociace EOSC a společných mezinárodních projektů, což vytváří

---

<sup>4</sup> <https://eosc.eu/partnership>

<sup>5</sup> <https://eosc.eu/sria>

<sup>6</sup> <https://eosc.eu/tripartite-collaboration>

<sup>7</sup> <https://eosc.eu/>

dobré předpoklady pro využití zkušeností a pro zajištění kompatibility a koordinace mezi rozvojem EOSC v ČR a v ostatních evropských zemích.

I do budoucna je proto žádoucí podporovat aktivní zapojení zájemců zejména do práce Task Forces a dalších návazných pracovních skupin, které EOSC Association zaštiťuje. Očekává se, že spolufinancování EOSC na evropské úrovni bude realizováno především formou nových výzev a v nich podpořených projektů. Je proto nezbytné vytvořit motivační prostředí pro širší zapojení českých institucí a týmů do mezinárodních projektů. To jednak zajistí kontinuální napojení na celoevropské aktivity, jednak to přinese další finanční prostředky pro budování EOSC v ČR.

## 4 Rozvoj EOSC v EU

Tato kapitola se zabývá stavem implementace EOSC v zemích EU a zejména způsobem poskytování podpory na rozvoj EOSC z evropských i národních zdrojů. Součástí rešerše je podrobnější zmapování situace v šesti vybraných evropských zemích, z nich tři se již nacházejí v pokročilé fázi implementace EOSC (Německo, Nizozemsko, Francie) a další tři, velikostně srovnatelné s ČR, v počáteční fázi implementace EOSC (Rakousko, Portugalsko, Švédsko).

### 4.1 Vývoj evropských iniciativ

Podpora implementace EOSC z rámcového programu **Horizont 2020** (H2020) byla realizována především pod oblastí „Evropské výzkumné infrastruktury – včetně e-infrastruktur“ (*European research infrastructures - including e-Infrastructures*) v prioritě „Vynikající věda“ (*Excellent science*). Priorita „Vynikající věda“ podporuje excelentní výzkum a vytváří pro jeho provoz podmínky. Jedná se zejména o podporu vynikajících projektů hraničního výzkumu, vývoj nových a ambiciózních technologií, mobilitu vynikajících výzkumných pracovníků a špičkové infrastruktury. Výzkumné infrastruktury hrají důležitou roli v rozvoji znalostí a technologií, jejich rozšiřování a nabízejí vysoce kvalitní služby uživatelům z různých zemí.

Následující přehled uvádí nejvýznamnější výzvy a projekty, které byly zaměřeny výlučně na implementaci EOSC. Podpořeny byly jednak pilotní projekty a zkvalitnění a integrace existujících výzkumných institucionálních a tematických datových repozitářů, v další fázi potom už jde o přechod k vizi organizované panevropské interdisciplinární repozitářové platformy pro sdílení a správu výzkumných dat po dobu celého jejich životního cyklu.

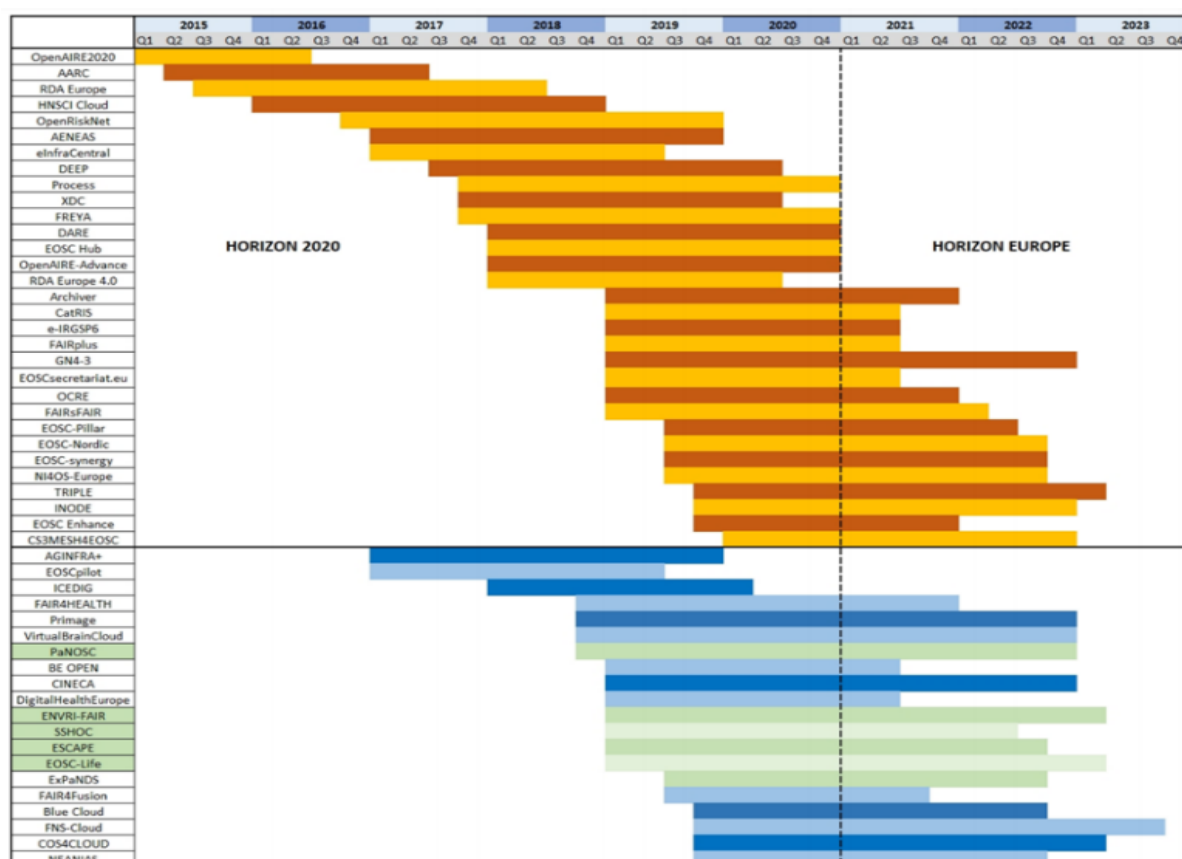
**Tab. 1: Výzvy programu H2020 zaměřené na implementaci EOSC**

Pracovní program (H2020)	Výzva	Alokace (EUR)	Datum vyhlášení	Datum ukončení
Pracovní program 2016-2017	Call - Development and long-term sustainability of new pan-European research infrastructures, H2020-INFRADEV-2016-2017 INFRADEV-04-2016: European Open Science Cloud for Research	10 000 000	15. 3. 2016	22. 6. 2016
	Call E-Infrastructures, H2020-EINFRA-2016-2017 EINFRA-12-2017: Data and Distributed Computing e-infrastructures for Open Science	26 000 000	14. 4. 2016	20. 9. 2016

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

<b>Pracovní program 2018-2020</b>  <b>Call - Implementing the European Open Science Cloud, H2020-INFRAEOSC -2018-2020</b>	INFRAEOSC-01-2018: Access to commercial services through the EOSC hub	12 000 000	05. 12. 2017	22. 3. 2018
	INFRAEOSC-02-2019: Prototyping new innovative services	28 500 000	16. 10. 2018	29. 1. 2019
	INFRAEOSC-03-2020: Integration and consolidation of the existing pan-European access mechanism to public research infrastructures and commercial services through the EOSC Portal	40 900 000	19. 11. 2019	18. 6. 2020
	INFRAEOSC-04-2018: Connecting ESFRI infrastructures through Cluster projects	95 000 000	5. 12. 2017	22. 3. 2018
	INFRAEOSC-05-2018-2019: Support to the EOSC Governance	30 000 000	8. 2. 2018	19. 6. 2018
	INFRAEOSC-06-2019-2020: Enhancing the EOSC portal and connecting thematic clouds	2 000 000	14. 11. 2018	20. 3. 2019
	INFRAEOSC-07-2020 - Increasing the service offer of the EOSC Portal	24 000 000	19. 11. 2019	18. 6. 2020

Zdroj: Pracovní programy „Evropské výzkumné infrastruktury – včetně e-infrastruktur“ 2016-2017 a 2018-2019; korekce podle Funding and Tender portálu



**Key:** Orange shades indicate coordination, best practice or technology projects. Blue shades indicate thematic science-based projects. Green shades indicate ESFRI cluster projects.

**Obr. 1: Projekty směřující k implementaci EOSC**

Zdroj: e-INFRA.cz



## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

Přehled podpory implementace EOSC z programu H2020 ukazuje, že iniciativa EOSC je významně napojena na oblast velkých výzkumných infrastruktur (VVI). Z důvodu robustnosti vytvářených a spravovaných systémů výzkumných dat stojí VVI logicky v popředí zájmu úsilí zaměřeného na FAIR principy.

Evropské strategické fórum pro výzkumné infrastruktury (ESFRI) bylo založeno roku 2002 z rozhodnutí Rady EU pro konkurenceschopnost. Od roku 2006 ESFRI zpracovává a publikuje tzv. Cestovní mapy ESFRI, zahrnující projekty evropských výzkumných infrastruktur, které byly identifikovány jako prioritní pro rozvoj Evropského výzkumného prostoru, a zprostředkující výsledky strategických debat k aktuálním otázkám tvorby politiky VVI.

**Cestovní mapa ESFRI 2021** zdůrazňuje význam propojení VVI s principy EOSC. VVI jsou vnímány jako klíčoví producenti výzkumných dat a zároveň jako “konzumenti” služeb EOSC napříč znalostními doménami, a jsou tedy prozatím nejvýraznější cílovou skupinou z hlediska kvalitního a udržitelného rozvoje EOSC. Dále dokument uvádí příklady nových tematických infrastruktur, v jejichž rámci se na evropské úrovni formuje klastrové pojetí implementace EOSC.

Tabulka níže poskytuje přehled VVI podpořených v rámci výzvy H2020-INFRAEOSC-2018-2020.

**Tab. 2: Přehled VVI podpořených v rámci výzvy H2020-INFRAEOSC-2018-2020**

Příklady projektů evropských klastrů velkých výzkumných infrastruktur zaměřených na EOSC	Rok zahájení	Rok ukončení	Rozpočet (EUR)
ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research (ENVI-FAIR)	1. 1. 2019	3. 6. 2023	18 997 878,75
Providing an open collaborative space for digital biology in Europe (EOSC-Life )	1. 3. 2019	31. 8. 2023	26 145 996,25
European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures (ESCAPE)	1. 2. 2019	31. 1. 2023	15 983 301,25
Photon and Neutron Open Science Cloud (PaNOSC)	1. 12. 2018	30. 11. 2022	11 953 516,99
Social Sciences & Humanities Open Cloud (SSHOC)	1. 1.2019	30. 4. 2022	14 493 360,36

Zdroj: Cordis

Mimo výše uvedeného ale výzkumné infrastruktury zařazené v Cestovních mapách ESFRI získaly ještě i další prostředky z jiných výzev EK, rovněž souvisejících s podporou budování EOSC.

## 4.2 Aktuální stav

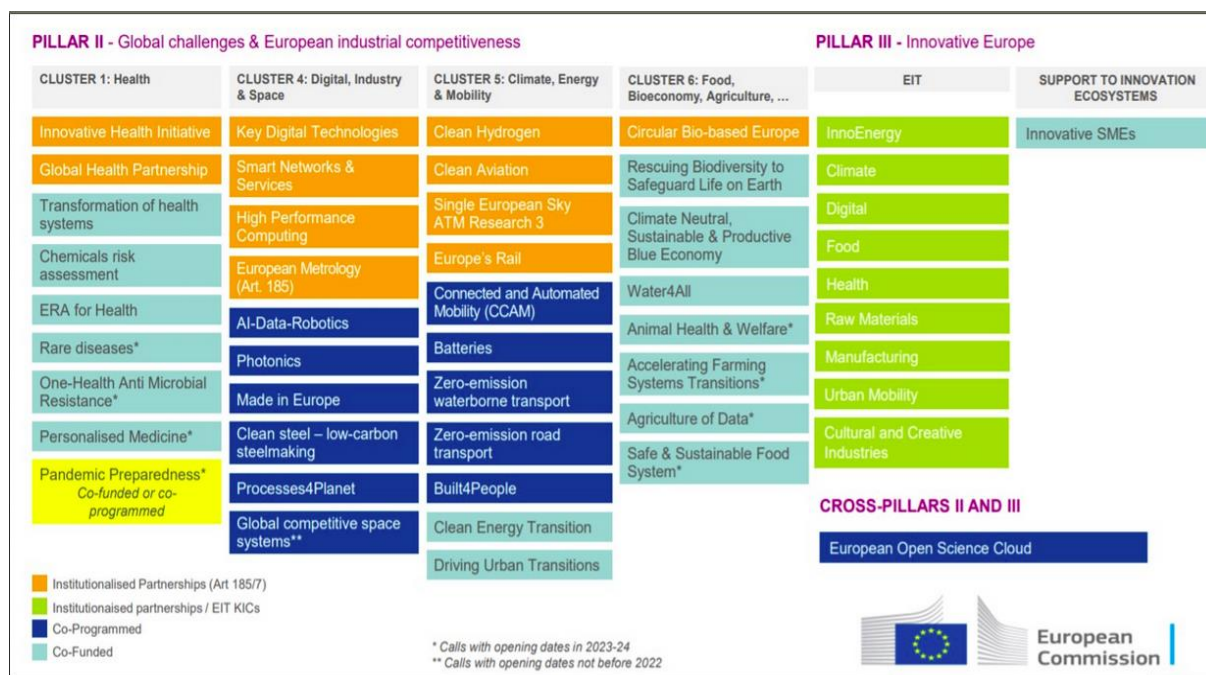
V současnosti probíhá na úrovni EU implementace EOOSC prostřednictvím nástroje Společně naprogramovaných evropských partnerství. Evropská partnerství jsou iniciativy, jejichž prostřednictvím Evropská unie v součinnosti s veřejnými i soukromými partnery podporuje rozvoj a implementaci programu na podporu výzkumu a inovací. Tato partnerství by měla být hlavním prostředkem v řešení složitých problémů, které potřebují integrovaný přístup za součinnosti velkého počtu aktérů z různých částí hodnotových řetězců a zemí. Vycházejí tedy ze společných vizí zakotvených ve strategických dokumentech (tzv. *Roadmaps*, Cestovních mapách) podporovaných všemi partnery.

Rámcový program Horizont Evropa (HE) zavádí v evropských partnerstvích strategičtější, ucelený přístup zaměřený na jeho dopady. Partnerství jsou zakládána v těch oblastech, v nichž mohou dosáhnout cílů programu HE účinněji než jinými prostředky rámcového programu. Partnerství institucionalizovaná Evropskou komisí (článek 185/7) budou v období 2021-2030 připravovat a vyhlášovat svoje výzvy napříč programem.

HE pracuje se třemi typy partnerství:

1. Společně naprogramovaná evropská partnerství (*Co-programmed European Partnerships*)
2. Spolufinancovaná evropská partnerství využívající akci spolufinancovanou programem (*Co-funded European Partnerships using a program co-fund action*)
3. Institucionalizovaná evropská partnerství (*Institutionalized European Partnership*)

Celkový přehled aktuálních evropských partnerství a pozici EOOSC v jejich rámci poskytuje obrázek níže.



**Obr. 2: Evropská partnerství v programu Horizont Evropa**

Zdroj: webový portál Horizont Evropa

Společně naprogramované evropské partnerství pro EOOSC bylo pro program HE zakotveno ve Strategickém plánu HE 2021-2024. Dalším stupněm v rozvoji tohoto partnerství bylo uzavření Memoranda mezi EK a Asociací EOOSC v červnu 2021, podrobně specifikující cíle, závazky, organizaci, trvání, indikátory, ad. V článku 2 tohoto dokumentu je pro období 2021-2030 deklarována ze strany

## Analýza aktuálního stavu kapacit pro implementaci EOOSC v ČR

Evropské komise předpokládána výše alokovaných finančních prostředků 490 mil. EUR. Tentýž článek deklaruje přibližně stejnou výši další podpory ze strany členských organizací Asociace EOOSC včetně dalších národních příspěvků (500 mil. EUR).

V souvislosti s tímto partnerstvím byla z programu HE vyhlášena mimořádná výzva na podporu sdílení dat pro výzkum variant koronaviru (viz tabulka níže).

Tab. 3: Výzva na podporu sdílení dat pro výzkum variant koronaviru

Výzva (Program HE)	Alokace (EUR)	Datum vyhlášení	Datum ukončení
FAIR and open data sharing in support to European preparedness for COVID-19 and other infectious diseases (HORIZON-INFRA-2021-EMERGENCY-01)	12 000 000	15. 4. 2021	20. 5. 2021

Zdroj: Funding and Tender portál

Z pracovních programů HE byl prozatím zveřejněn Pracovní program 2021-2022, jehož obsahem v oblasti „Výzkumných infrastruktur“ je prohlubování struktur evropského ekosystému EOOSC (viz tabulka níže).

Tab. 4: Výzvy na podporu EOOSC v Pracovním programu 2021-2022

Pracovní program (HE)	DESTINATION – ENABLING AN OPERATIONAL, OPEN AND FAIR EOOSC ECOSYSTEM (INFRAEOOSC)	Alokace (EUR)
Pracovní program 2021-2022	Call - HORIZON-INFRA-2021-EOOSC-01	59 000 000
	Call - HORIZON-INFRA-2022-EOOSC-01	30 000 000

Zdroj: Funding and Tender portál

V rámci EOOSC Procurement je vyčleněno dalších 35 mil. EUR na období 2022-2025 s cílem financovat vybrané služby<sup>8</sup>.

### 4.3 Situace a praxe ve vybraných zemích

Následující rešerše poskytuje příklady aktuální situace a běžné praxe ve vybraných členských státech EU. Jejich výčet není úplný, ale může poskytnout přehled o některých aktivitách, které probíhají na národní úrovni ve vybraných zemích EU.

<sup>8</sup> <https://ted.europa.eu/udl?uri=TED:NOTICE:234660-2022:TEXT:EN:HTML>

#### 4.3.1 Příklady států v pokročilé fázi implementace EOOSC

Do této skupiny byly zařazeny ty členské státy EU, které mají ve vztahu k EOOSC národní plán, národní program financování i základ budoucí národní platformy.

##### Německo

Páteří německého systému implementace EOOSC je národní infrastruktura výzkumných dat (*Nationale Forschungsdateninfrastruktur*, NFDI). NFDI má systematicky rozvíjet, trvale zabezpečovat a zpřístupňovat datové fondy vědy a výzkumu a propojovat je (mezi)národně. Bude vytvořena v procesu řízeném vědeckou komunitou jako síťová struktura konsorcií jednajících z vlastní iniciativy. Spolkové ministerstvo školství a výzkumu (*Bundesministerium für Bildung und Forschung*, BMBF) financující NFDI z 90 % spolu s odpovídajícími ministerstvy jednotlivých spolkových států (garantujících zbytkových 10 %, *Bund-Länder-Vereinbarung*) má v období 2021-2030 vyčleněno na podporu NFDI jako německého příspěvku k EOOSC vyčleněno 900 milionů EUR. (Zdroj: EOOSC Pillar)

Akční plán BMBF pro výzkumná data (*Aktionsplan Forschungsdaten*) sdružuje aktivity a projekty s cílem vytvořit kulturu sdílení a opětovného využívání výzkumných dat. Kromě nezbytných struktur bere akční plán v úvahu také potřebu standardizace při nakládání s výzkumnými údaji a podporu kompetencí ve vědě, průmyslu a technice. Akční plán je založen na třech základních prvcích, které společně posilují technologickou, ekonomickou a vědeckou suverenitu a umožňují spolehlivé využití dat:

1. Jako základ pro využití dat podporuje BMBF vývoj technologií a metod pro výzkum a vývoj související s daty. Za účelem zlepšení možnosti výměny dat a posílení jejich kvality bude cíleně podporováno zřízení vnitrostátních a evropských datových infrastruktur, jako je Národní infrastruktura pro výzkumné údaje (NFDI) a GAIA-X.
2. Dalším klíčovým prvkem akčního plánu BMBF pro výzkumná data je podpora inovací a nápadů vycházejících z výsledků výzkumu v podobě dat. Pro zvýšení inovační kapacity je důležité zajistit, aby všichni účastníci inovačního procesu měli co nejjednodušší přístup k výzkumným údajům. Za tímto účelem akční plán předpokládá například definování jednotných metod, standardů a datových formátů.
3. Výměna a využívání výzkumných údajů nakonec vyžaduje společnou úroveň znalostí o nakládání s těmito údaji, jejich shromažďování a využívání. Z tohoto důvodu BMBF sleduje cíl podporovat komplexní datové kompetence.



**Obr. 3: Akční plán Výzkumná data**

Zdroj: webový portál BMBF

Na základě položeného rozhovoru z dubna 2022 se zástupcem DFG (*Deutsche Forschungsgemeinschaft*, Německá agentura na podporu výzkumu) došlo k validaci konceptuálního přístupu k NFDI jako německého příspěvku k EOSC. Současně bylo zdůrazněno, že připravenost k implementaci EOSC je i v Německu výrazně proměnlivá podle různých disciplín. Výrazné zastoupení mají ale společenské a humanitní vědy. NFDI sdružuje až třicet konsorcií a je současně právnickou osobou a mandátní organizací, která tato konsorcia zaštiťuje a působí tedy jako kontaktní místo pro evropskou úroveň EOSC. Konsorcia mají velkou rozmanitost s ohledem na svoje způsoby financování.

#### Nizozemsko

Nizozemské Ministerstvo školství, kultury a vědy (*Onderwijs, Cultuur en Wetenschappen*, OCW) je významným hráčem v národním i evropském rozvoji iniciativy Open Science, proklamovaným na nejvyšší politické úrovni Evropskou komisí už v roce 2016, v době nizozemského předsednictví EU. Následovala mezirezortní debata ke sdílení výzkumných dat, na níž se mimo vědeckých a výzkumných institucí podílelo rovněž např. ministerstvo financí (*Economische Zaken*, EZK) a privátní sektor. Široká skupina zainteresovaných stran je koordinována několika dalšími klíčovými aktéry: *National Platform Open Science*, *National Coordination Point Research Data Management*, *Research Data Netherlands*, *Netherlands Federation of University Medical Centres*, významné infrastruktury napojené na ESFRI (ODISSEI, BBMRI-NL) atd. Mnohaúrovňová koordinace jistě přispívá k nalezení optimálního řešení pro národní EOSC, na druhé straně však objektivně znesnadňuje implementaci alespoň elementární hierarchie. (Heikkurinen 2020: 22) Základ poskytuje nizozemská národní platforma v oblasti Věd o živé přírodě (*Dutch Techcentre for Life Sciences*, DTLs), kterou lze pravděpodobně považovat za zárodek transdisciplinární iniciativy.

Národní program pro Open Science (*Nationaal Programma Open Science*, NPOS) má v Nizozemsku institucionalizovanou formu v čele s řídicím výborem (*NPOS Steering Committee*), národním

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

koordinátorem pro Open Science (*National Coordinator Open Science*) a poradním orgánem (*NPOS Advisory Council*). Činnost NPOS byla zahájena v roce 2017, od roku 2019 je pak NPOS funkčně napojen také na národní program veřejné podpory řízený a administrovaný nizozemskou radou pro výzkum (*Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO*). NPOS je členěn do tří hlavních programových sekcí: FAIR data, Open Access a Citizen Science.

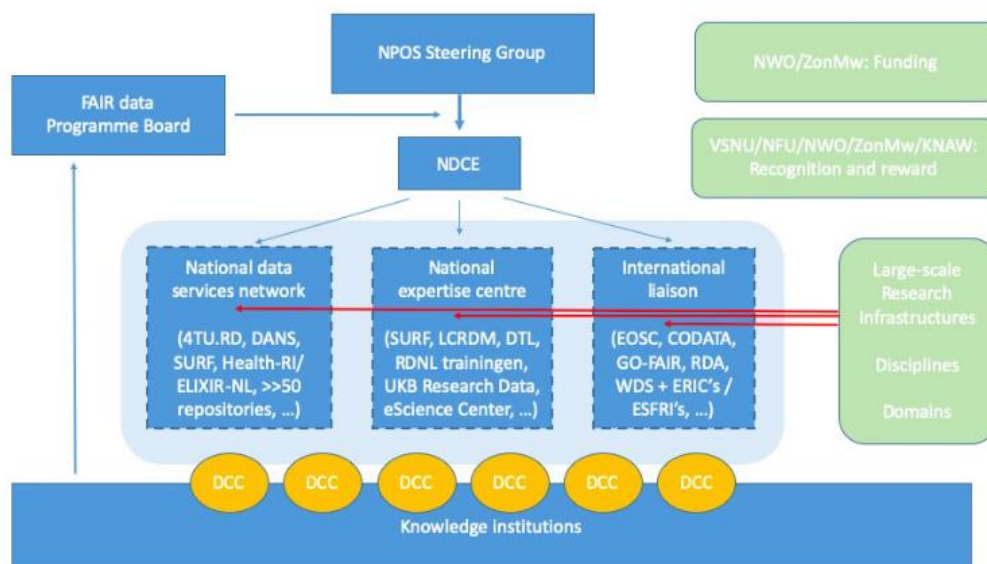
V roce 2020 NPOS zveřejnil dokument sekce FAIR data (*Exploring and optimising the Dutch data landscape*), který představuje plán národní implementace EOSC do roku 2025. Významnými milníky jsou:

- založení národního datového koordinačního a expertního centra (*National Data Coordination and Expertise Centre – NDCE*);
- jmenování poradního programového panelu pro FAIR data (*Programme Board for the FAIR data*);
- zřízení národní sítě všech poskytovatelů datových služeb a datových repozitářů (*National Open Science Cloud – National Data Services Network*) – katalog s jednotnými standardy služeb, koordinací, indikátory ad.;
- ustavení NDCE jako národního uzlu k transparentnímu napojení na EOSC EU a další mezinárodní organizace datových služeb ad.

Úplný přechod na FAIR principy v oblasti výzkumných dat pak NPOS deklaruje do roku 2030 (*Open Science 2030 in the Netherlands – NPOS2030 Ambition Document*).

Obrázek níže zachycuje strukturu nizozemské cesty k přechodu k FAIR principům.

A roadmap for the FAIR Data programme line



**Obr. 4: Cestovní mapa programové linie FAIR Data**

Pozn.: DCC – Digital Curation Centre

Zdroj: webový portál NPOS

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

NWO jako hlavní nizozemská národní organizace poskytující dotace do výzkumu a vývoje od roku 2019 navyšuje svůj dlouhodobý plán na financování e-infrastruktur a implementace EOSC v Nizozemsku (*Implementation Plan Investments Digital Research Infrastructure*). Pro rok 2019 počítal plán se vstupní částkou 40 mil. EUR, pro každý další rok implementace dále s dodatečnou podporou 20 mil. EUR<sup>9</sup>. Prozatím bylo z tohoto programu zahájeno financování 23 projektů (NWO 2022)<sup>10</sup>.

### Francie

Současným základním národním dokumentem k oblasti Open Science je ve Francii tzv. Druhý národní plán pro otevřenou vědu 2021-2024 (*Deuxième Plan national pour la science ouverte*), nastavující hlavní priority do roku 2024. Vyhlášen byl francouzským Ministerstvem pro vysoké školství, výzkum a inovace (*Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation*) v návaznosti na první plán z roku 2018. Jeho druhá osa je věnována souladu výzkumných dat s FAIR principy. Konkrétní opatření se týkají povinnosti jejich zveřejňování, vytvoření národní platformy pro výzkumná data (*Recherche Data Gouv*) a další podpory datových politik ve vztahu k celému životnímu cyklu výzkumných dat.

Za implementaci národního plánu pro Open Science odpovídá ve Francii osmičlenný Výbor pro otevřenou vědu (*Comité pour la science ouverte*), tvořený nejvyššími představiteli státu v agendě výzkumu a vývoje a prezidenty hlavních relevantních výzkumných a vysokoškolských asociací. Administrativně je výbor podporován Stálým sekretariátem pro otevřenou vědu (*Secrétariat permanent pour la science ouverte*), poradními orgány jsou pak sedm expertních kolegií a skupin (angl. *Publications College, Research Data College, European and International College, Skills and Training College, Software and source codes College, Open Scientific Publishing Expert Group, Legal Affairs Expert Group*) a široké profesní fórum.

Posledním významným článkem francouzského systému pro řízení otevřené vědy je Národní fond pro otevřenou vědu (*Le Fonds national pour la science ouverte*, FNSO), tj. dotační nástroj, jehož hlavní priority a cíle definuje Výbor pro otevřenou vědu. Nejobecnějším záměrem je rozvoj zdatného, transparentního, na datech založeného a společensky uznávaného vědeckého ekosystému. V souladu s národním plánem pro otevřenou vědu jsou hlavními prioritami FNSO:

- otevřený přístup,
- otevřená data,
- otevřený software,
- a lidské zdroje pro otevřenou vědu spolu s podporou osvěty.

Tabulka níže představuje přehled francouzských aktivit financovaných ze strany FNSO, které mají vztah k FAIR principům.

**Tab. 5: Francouzské národní aktivity FAIR, 2021-2023**

Iniciativa /projekt	Trvání	Rozpočet (EUR)
GT Certification des entrepôts de données	2021	100 000
GT Data Sharing Plan Rennes	2021	78 000
Recherche Data Gouv Plateform	2021-2023	4 110 000

<sup>9</sup> <https://www.nwo.nl/en/researchprogrammes/implementation-plan-investments-digital-research-infrastructure> - Programová dokumentace je prozatím zveřejněna do konce r. 2022.

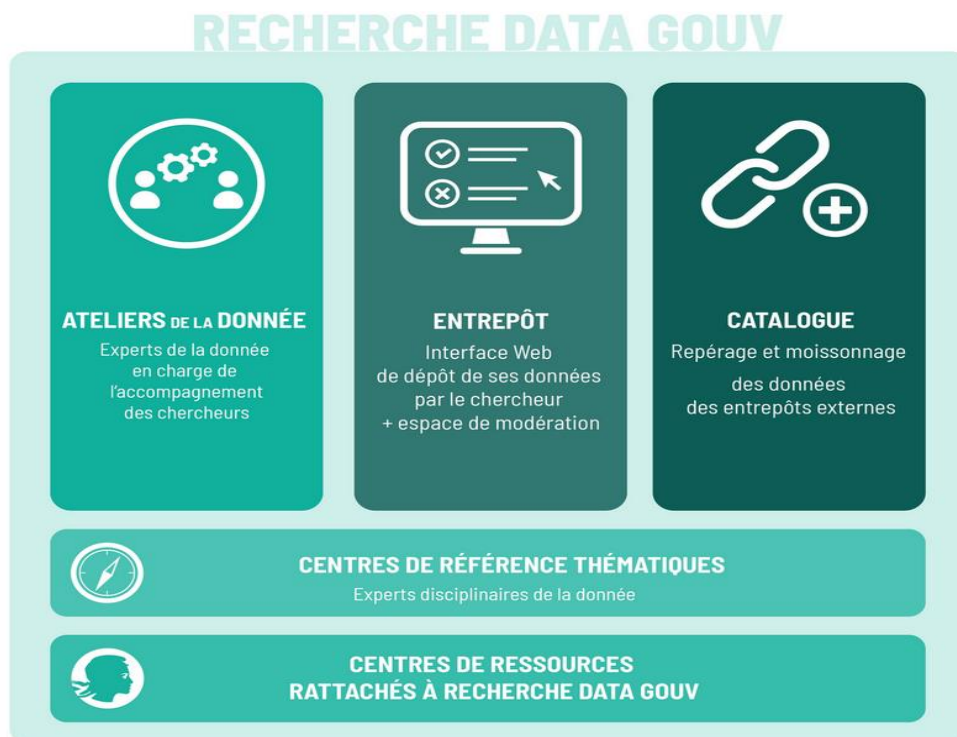
<sup>10</sup> [https://www.nwo.nl/en/projects?0&f%5B0%5D=nwo\\_projects\\_program%3A55724](https://www.nwo.nl/en/projects?0&f%5B0%5D=nwo_projects_program%3A55724)

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

I-FAIR IR	2021-2023 (30 měsíců)	298 193,4
FAIREST	2021-2022 (24 měsíců)	227 059,20
OpenAgroécologie	2021-2022 (24 měsíců)	140 000

Zdroj: Webový portál *Ouvrir la science*

Grafika níže přibližuje schéma repozitářové infrastruktury vznikající pod iniciativou *Recherche Data Gouv Plateform*.



**Obr. 5: Schéma francouzské repozitářové infrastruktury**

Legenda:

- Datové semináře (datoví experti zajišťující podporu výzkumných pracovníků),
- Repozitář (webové rozhraní pro výzkumné pracovníky, kde mohou ukládat svá data + prostor pro moderaci),
- Katalog (popis a výběr dat z externích úložišť),
- Tematická referenční centra (odborníci na oborová data),
- Datové zdroje připojené do repozitářové infrastruktury

Zdroj: webový portál *Ouvrir la science*

### 4.3.2 Příklady států v počáteční fázi implementace EOSC

Pokud byla první skupina zemí popsána jako ta, která už má národní plán, program financování a základ budoucí národní platformy, zemím ve druhé skupině jeden nebo více těchto elementů zatím chybí. Přesto ale může být užitečné zmínit i příklady tří zemí, které jsou počtem obyvatel srovnatelné s Českou republikou.

#### Rakousko

Rakouská podpora EOSC vychází z aktualizace Cestovní mapy výzkumných infrastruktur ESFRI uveřejněné v roce 2018. V lednu 2021 vytvořila Technická univerzita ve Vídni (TU Wien) novou



organizační jednotku, EOSC and International Liaison Office, se sídlem v knihovně TU Wien a v únoru 2021 požádala asociace AConet<sup>11</sup> o členství v asociaci EOSC jako rakouská pověřená organizace.

EOSC Support Office Austria a jeho řídicí struktura byly formálně založeny v říjnu 2021 jako provozní jednotka rakouské pověřené organizace EOSC, právně zastoupená sdružením ACONET a zajištěná smluvní odpovědností ze strany TU Wien. O členství v asociaci EOSC AIBL požádaly čtyři rakouské výzkumné infrastruktury: Přírodovědné muzeum ve Vídni, Technická univerzita ve Štýrském Hradci (TU Graz), TU Wien a Vídeňská univerzita. O status pozorovatele požádalo Centrum pro změnu klimatu v Rakousku (CCCA), Univerzita v Linci (JKU) a Akademie výtvarných umění ve Vídni. V únoru 2022 byla přijata politika otevřené vědy Rakouska (*Open Science Policy Austria. Österreichische Policy zu Open Science und der European Open Science Cloud*). Neformálním způsobem výměny informací je v Rakousku organizace tzv. EOSC Café.

Na základě položeného rozhovoru z dubna 2022 se zástupcem Spolkového ministerstva školství, vědy a výzkumu (*Bundesministerium für Bildung, Wissenschaft und Forschung, BMBWF*) došlo k validaci konceptuálního přístupu k datové infrastruktuře jako páteři rakouského EOSC. V této souvislosti je třeba zmínit zřízení FAIR Office Austria, při TU Graz, TU Wien a Vídeňské univerzitě. V rozhovoru byly dále zmíněny některé slabiny přístupu k otevřeným datům, a to především v souvislosti s lékařskými obory a dále s komerčním sektorem.

### Portugalsko

Aktivity EOSC v Portugalsku navazují na dokument Hlavní zásady otevřené vědy přijatý v únoru 2016. Od tohoto dokumentu se odvíjí Národní politika pro otevřenou vědu<sup>12</sup>. Klíčovým hráčem je v této souvislosti Nadace pro vědu a technologie (*Fundação para a Ciência e a Tecnologia, FCT*), která podporuje výzkumné infrastruktury strategického zájmu a politiku otevřeného přístupu.

Konkrétní informace lze nalézt v portugalské účasti v ESFRI. (Heikkurinen 2020: 12-14) FCT vyvíjí národní síť pokročilých počítačů, která bude integrovat zdroje ze společného podniku EuroHPC. Investice se odhadují ve výši 13 až 18 milionů EUR a 1,4 až 2 miliony EUR ročně na provoz. Pokud jde o datovou infrastrukturu, rozpočty za rok 2020 zahrnují portugalský uzel EUDAT a také Dataverse for the long tail of research data. Odhadovaný rozpočet odpovídá 100 000 EUR. FCT odpovídá za infrastrukturu RCAAP (*Repositórios Científicos de Acesso Aberto de Portugal*). RCAAP nabízí mimo jiné služby SaaS hostingu pro repozitáře (založené na DSpace) a časopisy s otevřeným přístupem (založené na OJS). Roční částka na provoz odpovídá přibližně 500 000 EUR.

Na základě položeného rozhovoru z května 2022 se zástupkyní FCT, lze v portugalském případě zmínit poměrně významnou integraci národních výzkumných infrastruktur v rámci ESFRI. Slabinou je nedostatečné povědomí výzkumné komunity o EOSC, a především jeho nevyváženost napříč oborovými klastry. Portugalská vláda se zavázala podporovat integraci EOSC na období 10 let (2021-2030) ve výši 1 milionu EUR přímé podpory a 0,5 milionu EUR nepřímo ze strany především FCT. Do budoucna Portugalsko také počítá se zřízením národního kontaktního místa pro EOSC.

### Švédsko

Centrálním uzlem pro FAIR správu výzkumných dat je Švédská národní datová služba (*Svensk nationell datatjänst, SND*). Ta spolupracuje už se 40 vysokými školami a veřejnými výzkumnými organizacemi napříč vědními obory, a poskytuje jim vysoce kvalitní vyhledávací a popisný systém pro sdílení a správu

<sup>11</sup> <https://www.aco.net/>

<sup>12</sup> [https://www.ciencia-aberta.pt/files/ugd/a8bd7c\\_1ca622bff7f34abbad228ac94e1eda16.pdf](https://www.ciencia-aberta.pt/files/ugd/a8bd7c_1ca622bff7f34abbad228ac94e1eda16.pdf)

jejich dat (systém DORIS). Metadata mohou být následně sdílena, jak na národní, tak mezinárodní úrovni, prostřednictvím národního SND katalogu výzkumných dat.

V kontaktu mezi jednotlivými institucemi (SND Network) je preferován přístup zachování a podpory lokálních funkčních specifíků. Za tímto účelem jsou průběžně zakládána regionální přístupová střediska (*Data Access Units* – DAUs). Na stejném principu je následně ze strany SND zprostředkována školicí podpora zaměstnanců DAU, a také školicí podpora pro výzkumníky co do tvorby data management plánů, tak co do bezpečné správy a publikování dat skrze národní SND katalog.

Hlavním orgánem SND je řídicí výbor (*Steering committee*), složený ze zástupců SND konsorcia (*SND Consortium*) tvořeného 9 švédskými univerzitami. V této formě je SND řízena od roku 2018, ačkoli založena byla z iniciativy vládní agentury pro výzkum (*Vetenskapsrådet*, VR) už v roce 2008. Univerzity zastoupené v konsorciu rovněž vysílají do SND své experty na tzv. hlavní domény specializace SND. Dalším důležitým orgánem SND je rada DAU (*DAU Council*), v níž má každé přístupové středisko svého zástupce.

Hlavní domény specializace SND:

- *social sciences*
- *humanities*
- *public health and epidemiology*
- *environmental data*
- *sensitive data*
- *spatial data - geodata and GIS data*
- *materials science*
- *life science*
- *registry-based medical and other research*
- *interdisciplinary population-based research related to registries and biobanks*

V roce 2019 byla publikována Švédská národní cestovní mapa pro evropský výzkumný prostor 2019–2020 (*Swedish National Roadmap for the European Research Area 2019–2020*), jejíž pátou prioritou je Open Science. V této prioritě je zakotveno, že výsledky veřejného výzkumu musejí být v souladu s FAIR principy, a to nejpozději do r. 2026 – zpracováním fází implementace, a to v co nejrychlejší variantě, byla pověřena VR spolu s národní knihovnou (*Kungliga biblioteket*).

Finanční plán významné veřejné podpory národní implementace EOSC je ve Švédsku aktuálně v přípravě.

## 5 Předpoklady pro implementaci EOSC v ČR

Tato kapitola shrnuje současný stav přípravy národní datové infrastruktury EOSC a výsledky průzkumů realizovaných v nedávné době ohledně fungování vědeckých repozitářů a práce s vědeckými daty a informacemi. V rámci přípravy této studie byla rovněž realizována sada strukturovaných rozhovorů s uživateli, experty a stakeholdery zapojenými do rozvoje EOSC, jejichž potřeby a očekávání jsou v této kapitole také shrnuty.

## 5.1 Průzkum vědeckých repozitářů v ČR

Průzkum byl realizován v květnu a červnu 2020 týmem Ústavu výpočetní techniky Masarykovy univerzity (Bartošek a kol., 2020). Adresáty průzkumu byly vysoké školy, výzkumné organizace, akademické a veřejné knihovny. Realizátorům se sešlo 93 relevantních odpovědí od 43 institucí. Nejvíce odpovědí bylo z VŠ (17) veřejných knihoven (10), AV ČR a samostatných výzkumných organizací (9). Dotazník měl dvě části – část pro repozitář (57 odpovědí) a část pro datové sady uchovávané mimo repozitář (36 odpovědí).

Následuje stručný výběr výsledků zajímavých s ohledem na EOSC:

Naprostou většinu **repozitářů** bylo možné charakterizovat jako repozitáře s převážně textovým obsahem (37), uchování a zpřístupnění výzkumných dat se primárně týkalo pouze dvou repozitářů (Český sociálně vědní archiv a LINDAT/CLARIAH).

Pouze u čtvrtiny repozitářů respondenti uvedli, že k nim existuje nějaká institucionální směrnice upravující politiku otevřeného přístupu. Jen menší část repozitářů (41 %) byla registrována v některém mezinárodním repozitářovém registru. Více než polovina repozitářů ale umožňovala sklizení svého obsahu (data harvesting) některými dalšími systémy (nejčastěji metadat prostřednictvím protokolu OAI-PMH). Nedobrá byla naopak situace ohledně jednoznačné globální identifikace osob (výzkumníků) – jen 17 % repozitářů uvedlo, že umožňují použití identifikátoru ORCID (z části jako nepovinný údaj).

Jak nejčastější výzvy a problémy při provozu repozitářů byly identifikovány:

- nedostatek technického / odborného personálu či know-how (18),
- nejasná či chybějící Open Access / FAIR data / Open Science politika (16),
- nedostatečné financování (11),
- nedostatečný zájem ze strany vedení instituce (5),
- nedostatečné technické zázemí instituce (4).

V části dotazníku věnované **datovým sadám** se velikost datových sad pohybovala ve škále od desítek MB až po desítky TB. Nejobjemnějšími sadami byly Mapová sbírka UK (19 TB) a Obrazová studie MRI Centra neurověd CEITEC MU (12 TB). Dvě třetiny odpovědí se týkaly živých datových sad, tj. dat, která jsou dále aktivně modifikována. Více než polovina odpovědí uváděla, že výzkumná data nejsou opatřena žádnými metadaty. Téměř dvě třetiny vlastníků datových sad uvedly, že by uvítaly umístění datové sady do repozitáře (lepší viditelnost, dostupnost, zajištění).

Více než tři čtvrtiny vlastníků dat se domnívaly, že jejich data jsou potenciálně přínosná pro širší národní či mezinárodní výzkumnou komunitu, a je proto účelné usilovat o vhodné zajištění jejich dohledatelnosti a využitelnosti. Ukázalo se však, že problematika správy výzkumných dat nebyla zatím v akademickém prostředí v ČR rozvinutá a systematicky podchycená. Až na ojedinělé výjimky nebyly identifikovány šířeji využívané datové repozitáře a výzkumná data byla ukládána spíše ad-hoc na různých místech v organizaci, pod různou správou – a jejich rychlé dohledání a zmapování se ukázalo prakticky nereálné.

## 5.2 Přehled vybraných rozvinutých repozitářů a infrastruktur

Jakkoli průzkum z r. 2020 podchytil spíše převažující řadu týmů a institucí, které měly ohledně FAIR práce s daty jen minimální povědomí, zkušenosti a motivaci, existují v ČR pracoviště a infrastruktury, které jsou z hlediska FAIR práce s vědeckými daty na velmi pokročilé úrovni. Jedná se zejména o entity napojené na větší evropské projekty či infrastruktury, které jim umožňují sbírat potřebné zkušenosti a

mnohdy poskytují i technické a procesní zázemí pro FAIR práci s vědeckými daty. Mezi rozvinuté repozitáře a infrastruktury poskytující přístup k vědeckým datům provozované nebo využívané v českém prostředí patří následující:

**ELIXIR CZ** (<https://www.elixir-czech.cz/>) je česká národní infrastruktura pro biologická data. Jedná se o distribuovanou výzkumnou infrastrukturu pro bioinformatiku, která vznikla z pokročilého výpočetního prostředí. Nabízí bezplatné využívání specializovaných databází a unikátních nástrojů národní i mezinárodní komunitě vědců zabývajících se výzkumem v oblasti přírodních věd.

ELIXIR CZ je tvořen organizacemi provádějícími výzkum napříč Českou republikou, se sídlem v Ústavu organické chemie a biochemie Akademie věd ČR. Mimo Českou republiku je ELIXIR CZ partnerskou organizací Evropské výzkumné infrastruktury pro data z oblasti věd o živé přírodě (European Research Infrastructure for Life Science Data), mezivládní organizace, která koordinuje a rozvíjí zdroje z oblasti věd o živé přírodě v celé Evropě. ELIXIR CZ úzce spolupracuje s dalšími velkými výzkumnými infrastrukturami, konkrétně s INSTRUCT, EATRIS, BBMRI, EuroBioimaging a EU-OPENSREEN, a synergicky tak propojuje vědní obory z oblasti biologie.

**LINDAT/CLARIAH-CZ** (<https://lindat.mff.cuni.cz/cs>) je české centrum zaměřené na data, které poskytuje certifikované úložiště a služby počítačového zpracování jazyka. Jedná se o výzkumnou infrastrukturu, která se zabývá především jazykovými daty, ale také dalšími digitálními zdroji a nástroji pro jejich využívání, údržbu a jejich rozšiřování a nabízí je výzkumné komunitě, průmyslu pro rozvoj aplikací a ve specifických případech, jako je např. jazyková kultura, také přímo veřejnosti.

LINDAT/CLARIAH-CZ je spojeným distribuovaným českým národním uzlem celoevropských sítí CLARIN ERIC (Common Language Resources and Technology Infrastructure) a DARIAH ERIC (Digital Research Infrastructure for Arts and Humanities). Na infrastrukturu participuje 11 výzkumných organizací, které působí v oblasti humanitních věd a umění v ČR - v lingvistice, historii a historické bibliografii, kultuře a vědě o kultuře, dějinách umění, filozofii, filmové kultuře, výtvarném umění, muzikologii a dějinách hudby, etnologii, folkloristice, archeologii a také v některých mezioborových oblastech.

**Český sociálněvědní datový archiv (ČSDA)** (<https://archiv.soc.cas.cz/cz/>) shromažďuje data z českých sociálněvědních šetření, výzkumů veřejného mínění a mezinárodních šetření s českou účastí. Hlavním cílem je zpřístupnění dat pro účely akademického výzkumu a výuky. Mezi další cíle patří podpora sekundární analýzy dat a podpora speciálních výzkumných projektů. ČSDA je členem sítě evropských sociálněvědních datových archivů CESSDA (Council of European Social Science Data Archives) a sdružení střeoevropských archivů REGIO a podílí se na programech mezinárodní výměny dat.

**GENASIS (Global ENvironmental ASsessment Information System)** (<https://www.genasis.cz/>) je globální informační systém pro hodnocení životního prostředí, který poskytuje přehledné informace o kontaminaci životního prostředí chemickými látkami, zejména perzistentními organickými polutanty (POPs). Informační systém GENASIS vznikl na Masarykově univerzitě ve spolupráci Centra pro výzkum toxických látek v prostředí (RECETOX) a Institutu biostatistiky a analýz (IBA MU).

**BBMRI CZ** (<http://www.bbmri.cz/>) je národní výzkumná infrastruktura biobank a biomolekulárních zdrojů. Kolekce všech českých biobank jsou součástí katalogu BBMRI-ERIC Directory. Adresář BBMRI-ERIC je nástroj, který shromažďuje a zpřístupňuje informace o biobankách v celé Evropě, které jsou ochotny sdílet svá data a/nebo vzorky a spolupracovat s jinými výzkumnými skupinami. Adresář je jednou ze služeb, které BBMRI-ERIC Common Services for IT (CS-IT) nabízí celosvětové komunitě biobank, a byl vytvořen ve spolupráci s národními uzly BBMRI a partnery. Výzkumným pracovníkům Adresář nabízí možnost vyhledávání vzorků a dat, zatímco biobankám nabízí platformu pro sdílení existence jejich fondů a služeb a pro navázání kontaktů s výzkumnými pracovníky, kteří o ně mají zájem.

**Ústav molekulární a translační medicíny LF UP v Olomouci (IMTM)** (<https://imtm.cz/>) je národním uzlem velké výzkumné infrastruktury pro translační medicínu (European Translational Medicine Infrastructure, EATRIS-ERIC) a spravuje rozsáhlý repozitář medicínských dat, který zahrnuje např. data z extenzivního testování na SARS-CoV-2 způsobující COVID-19, data o projektu Českého genomu a o řadě dalších klinických studií a registrů. Na svém internetovém Data Stewardship Portalu (<https://portal.imtm.cz/>) pracoviště nabízí také sadu softwarových řešení pro správu dat v klinických studiích, registrech a dalších klinických/vědeckých databázích a dalšího proprietárně vyvinutého softwaru. IMTM nabízí kromě specifických nástrojů pro data stewardship uživatelům i přístup k datovým sadám pro hodnocení biologické aktivity chemických sloučenin, multiomickým profilům a anonymizovaným informacím z jednotlivých preklinických a klinických studií.

**CERN Open Data Portal** (<https://opendata.cern.ch/>) shromažďuje více než dva petabajty dat z oblasti částicové fyziky. Repozitář CERN je využíván Fyzikálním ústavem AV ČR a dalšími členy infrastruktury CERN-CZ, která organizuje a podporuje účast české vědecké komunity v mezinárodní laboratoři částicové fyziky CERN. V rámci infrastruktury CERN je mj. provozován i otevřený repozitář Zenodo (<https://zenodo.org/>) vyvinutý v rámci evropského projektu OpenAIRE.

### 5.3 Rámec a aktuální stav přípravy národní datové infrastruktury EOOSC

Základní dokument popisující **architekturu implementace EOOSC v ČR**<sup>13</sup> byl vytvořen v rámci práce platformy EOOSC při MŠMT v roce 2021. Na přelomu roku 2021 a 2022 byl na základě této platformy ustaven **Koordinační výbor EOOSC** jako poradní orgán MŠMT. Pro vlastní rozvoj EOOSC v ČR je plánována podpora z Operačního programu Jan Amos Komenský (OP JAK) prostřednictvím výzev Open Science v celkové výši cca 2,5 miliardy Kč v následujících šesti letech.

Vedle toho je plánována podpora zastřešujících aktivit na centrální úrovni prostřednictvím dvou individuálních projektů systémových (IPs).

Hlavním cílem systémového projektu EOOSC-CZ (**IPs EOOSC-CZ**) je vytvoření společného zázemí pro implementaci datově orientovaných součástí strategie Open Science v ČR, a to především v rámci pilířů FAIR data a EOOSC. V jedné rovině má být tohoto cíle dosaženo zřízením **Sekretariátu EOOSC-CZ** a zajištěním informačního a školicího zázemí pro celou oblast, monitoringem realizace a využití Národní datové infrastruktury (NDI), podporou pracovních skupin a poskytnutím expertního zázemí. Ve druhé rovině pak implementací a zajištěním provozu klíčových společných komponent NDI, především Národního metadatového adresáře (NMA), autentizačních a autorizačních služeb (pokročilá správa přístupu k datům i metadatům), provozem zázemí pro perzistentní identifikátory (PID) pro všechny digitální objekty v NDI, nastavením kyberbezpečnostních politik a řešení incidentů. Projekt by měl být realizován infrastrukturou e-INFRA CZ.

Projekt je komplementární k projektu CARDS (**IPs CARDS**, Czech Academic and Research Discovery Services), jehož nositelem je Národní technická knihovna (NTK). Ambicí tohoto systémového projektu je vytvoření unifikovaného prostředí pro sdílení a efektivní správu informačních zdrojů a podílení se na vytvoření společného zázemí pro implementaci datově orientovaných součástí strategie Open Science v ČR, a to především v rámci pilířů FAIR data a EOOSC. Stanoveného cíle by mělo být dosaženo jednak zřízením centrální platformy nové generace (PNG) pro správu všech typů informačních zdrojů společné pro knihovny všech vysokých škol a dalších výzkumných organizací i knihovny s celostátní působností, a za druhé komplexní metodickou podporou práce s metadaty výzkumných dat a s perzistentními

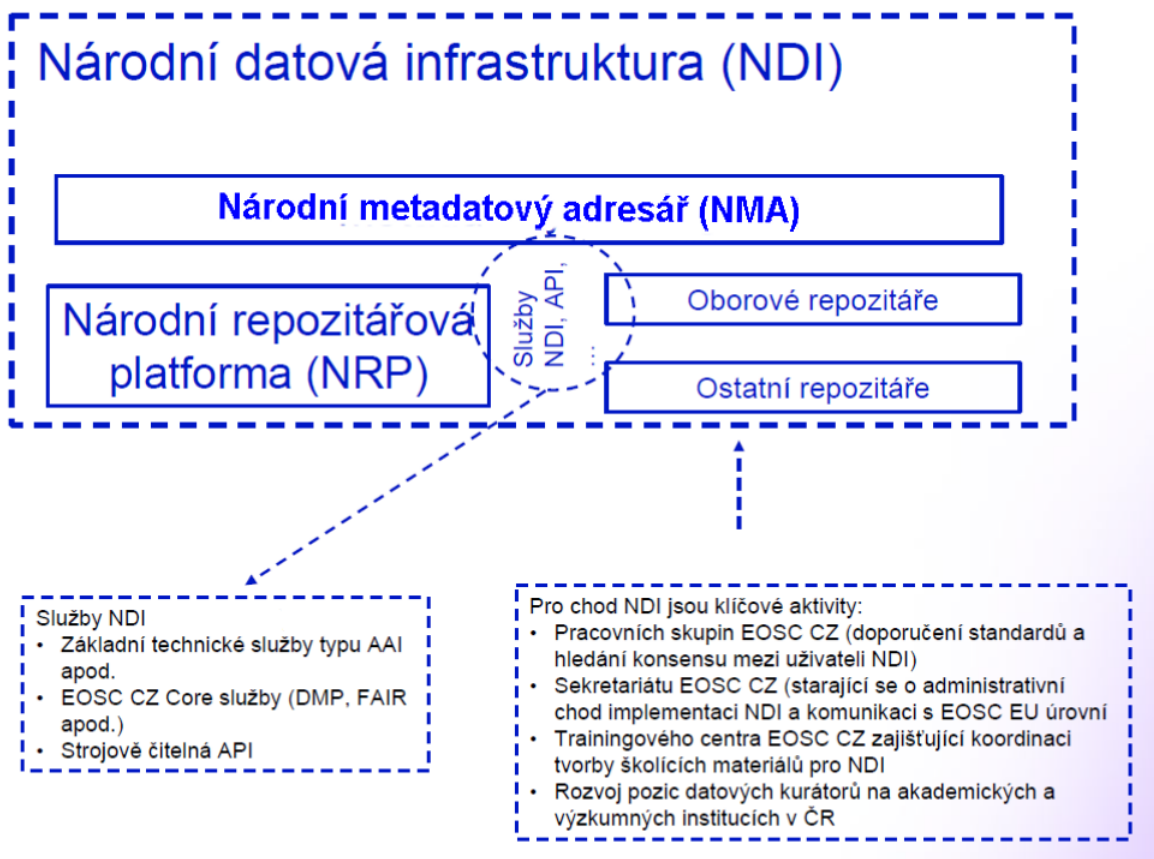
---

<sup>13</sup> [https://www.e-infra.cz/file/8cbdb401961e0ab98fe1ac3595b63e3e/475/EOOSC\\_architektura%20implementace.pdf](https://www.e-infra.cz/file/8cbdb401961e0ab98fe1ac3595b63e3e/475/EOOSC_architektura%20implementace.pdf)

identifikátory. Nástrojem pro zajištění metodické podpory nejen práce s metadaty, ale s repozitáři obecně a jejich interoperabilitou, bude Národní katalog repozitářů (NKR) zřízený v rámci projektu IPs CARDS, který bude evidovat existující repozitáře v ČR a informace o nich (např. metadatové formáty, ontologie apod.).

Komplementarita projektů CARDS a EOSC-CZ spočívá v tom, že společně vytvářejí předpoklady pro naležitelnost a přístupnost datových sad, tedy první dva požadované atributy FAIR dat (tj. findability a accessibility). Projekt CARDS díky zohledňování sémantické stránky práce s metadaty a podpoře perzistentních identifikátorů zajišťuje klíčový předpoklad využitelnosti uložených dat, totiž jejich interoperabilitu, tedy atribut třetí. Atribut čtvrtý, znovupoužitelnost (reusability) je v rukou tvůrců, resp. vkladatelů dat a závisí na rozsahu a kvalitě poskytnutých metadat.

Oba systémové projekty (IPs) tak mají vytvořit společné základní zázemí pro realizaci dalších plánovaných **výzev Open Science** v rámci OP JAK. Ty by měly podporovat implementaci federované národní datové infrastruktury EOSC na úrovni výzkumných institucí, resp. konsorcií těchto výzkumných organizací. Ve výzvách by takto mělo být podpořeno vybudování a provoz Národní repozitářové platformy (NRP) a podpořeny oborově-vědní repozitáře, to vše jako součást Národní datové infrastruktury ČR. Klíčovou součástí podpory musí být i zvyšování kapacit a kompetencí lidských zdrojů pro implementaci a následný provoz a využívání EOSC a jeho služeb.



**Obr. 6: Schéma návrhu Národní datové infrastruktury**

Zdroj: e-INFRA CZ

Na návrhu různých aspektů NDI v současné době pracují čtyři průřezové pracovní skupiny:

- **Metadata** – skupina má zajišťovat především podporu a konzultace k managementu metadat výzkumných dat, jak na obecné úrovni, tak pro vědecké obory;
- **Architektura NDI** – skupina má za cíl návrh a architektonický dohled nad implementací národní datové infrastruktury, specificky nad platformou pro datové repozitáře a bezprostředně navazující služby, jako je technická implementace zacházení s perzistentními identifikátory, sběr a indexace metadat, vyhledávání v metadatech, binárně spolehlivé ukládání datových sad a integraci těchto funkcionalit do národní e-infrastruktury jako celku;
- **Základní služby** – skupina má za cíl definovat základní služby NRP pro efektivní práci s vědeckými daty v ČR;
- **Vzdělávání a lidské zdroje** – skupina si definovala dvě hlavní oblasti zájmu, a to vzdělávání a podpora pro výzkumné pracovníky a vzdělávání pro “data support” (data steward, data curator).

Komplementárně k nim vznikly tematické (oborově-vědní) pracovní skupiny, které se zabývají návrhem oborově-vědních repozitářů a přípravou oborově-vědních projektů:

- **Bio/Health/Food** – integruje instituce a výzkumné infrastruktury z oblasti biologických, lékařských věd, potravin a výživy;
- **Environmental Sciences** – tematické zaměření na environmentální vědy;
- **Physics** – skupina sdružuje zástupce některých velkých výzkumných infrastruktur, které zajišťují účast na velkých mezinárodních experimentech, i mnohých menších skupin z oblasti fyziky;
- **Material Sciences and Technology** – cílem připravovaného projektu je vybudování, implementace a správa repozitáře pro oborově-vědní klastr Materiálové vědy a technologie;
- **Data Management for Artificial Intelligence and Machine Learning** – datová infrastruktura pro podporu rozsáhlých dat spolu se superpočítačovými schopnostmi pro vývoj nové generace pokročilých modelů strojového učení, které umožní nové aplikace pro analýzu dat v éře exascale computing;
- **Humanities and the Arts** – skupina je navrhována jako konzultační skupina pro všechna česká pracoviště zabývající se výzkumem v oblasti humanitních věd a umění;
- **Social Sciences** – skupina se zaměřuje na výzkumná data relevantní pro sociální vědy, bez ohledu na jejich disciplinární původ a metodologii;

Průřezový charakter napříč vědními obory má pracovní skupina zaměřená na správu citlivých dat a osobních údajů:

- **Sensitive Data** – skupina je zaměřena na správu citlivých dat a osobních údajů v celém svém měřítku, spolupracuje se všemi ostatními pracovními skupinami, a vytváří tak komplexní řešení založené na bohatých zkušenostech z různých vědních oborů.

Všechny pracovní skupiny jsou otevřené všem zájemcům o aktivní zapojení do práce těchto skupin<sup>14</sup>.

#### 5.4 Předpoklady pro efektivní práci s metadaty výzkumných dat

Pro ukládání i následné vyhledávání datových sad v rámci NRP i mimo ni bude nutné věnovat pozornost systematické práci s metadatovým popisem. Jak již bylo zmíněno, metadatové podpoře se bude

---

<sup>14</sup> <https://www.e-infra.cz/eosc/pracovni-skupiny-pro-implementaci-eosc-v-cr>

věnovat projekt IPs CARDS (vytvoření a provoz odborného metodického pracoviště pro podporu práce s metadaty výzkumných dat) ve spolupráci s pracovní skupinou Metadata projektu EOSC-CZ.

Hlavním předpokladem pro efektivní práci s metadaty je vytvoření sítě odborníků, která může poskytnout širší podporu a rozsah znalostí (zejména oborově specifických standardů). Tato síť je důležitá zejména v počátcích implementace EOSC v ČR a budování NDI a NRP, než se zvýší počet specializovaných pracovních pozic pro práci s výzkumnými daty v ČR (data curators, data stewards – viz kap. 5.6). Tuto síť bude koordinována odborným metodickým pracovištěm při NTK a bude zahrnovat odborníky z pracovní skupiny Metadata, z jednotlivých tematických skupin a případně i další. Tato síť odborníků se bude podílet na metodické podpoře a konzultacích v otázkách managementu metadat výzkumných dat a možnostech jejich ukládání do repozitářů či napojení zdrojů do NMA.

Předpokladem pro naležitelnost a interoperabilitu (meta)dat (první dva atributy FAIR principů) je kvalita a synchronizace metadatového popisu. Tu by měl na základní úrovni zajistit společně definovaný metadatový model pro popis výzkumných dat (dále jen model). Model určí minimální povinné údaje pro indexaci záznamů datových sad do Národního metadatového adresáře, čímž bude zajištěna interoperabilita základních údajů o datových sadách s napojenými repozitáři, ale také naležitelnost a přístupnost dat podle FAIR principů. Model by měl být opatřen dokumentací, udržován/rozvíjen a zároveň ověřen a otestován definovanou skupinou uživatelů z různých oborů a výše zmiňovanou sítí odborníků. Tento model by pak mohl sloužit také jako výchozí bod pro metadatový popis v nově vznikajících oborových repozitářích.

Předpokladem pro efektivní interoperabilitu repozitářů se zachováním sémantiky je jejich včasná evidence včetně technických parametrů, informací o použitých metadatových schématech, ontologiích apod. Pro zajištění této interoperability bude vytvořen Národní katalog repozitářů (NKR). NKR by měl sloužit vytvořené síti odborníků jako významný referenční zdroj základních informací o repozitářích a kontextu jejich obsahu, a zároveň by měl být adresářem relevantních zdrojů, ze kterých by měla být indexována metadata do NMA.

## 5.5 Průzkum stavu a očekávání v oblasti archivace a zveřejňování vědeckých dat mezi členy pracovních skupin Architektura a Základní služby EOSC

V březnu a dubnu 2022 byl realizován průzkum stavu a očekávání v oblasti archivace a zveřejňování vědeckých dat. Průzkum proběhl mezi členy pracovních skupin Architektura a Základní služby EOSC, je tedy omezen na respondenty, kteří jsou v problematice infrastruktur pro vědecká data aktivní. Tím se zásadně odlišuje od předchozího průzkumu provedeného v roce 2020 (Bartošek a kol., 2020) v gesci Rady pro výzkum, vývoj a inovace, který se snažil o maximální pokrytí vysokých škol v ČR. Oproti tomuto byl také současný průzkum zaměřen výrazně více technicky a jasně směřoval ke sběru požadavků pro návrh implementace EOSC v ČR. V neposlední řadě zkoumal rovněž v současnosti dostupnou expertízu na jednotlivých institucích.

### **Respondenty průzkumu byli:**

- Czech National Corpus Research Infrastructure (sociální a humanitní vědy)
- Digitální archiv Archeologické mapy České republiky, repozitář pro data o archeologických terénních výzkumech (archeologie)
- Astronomický ústav AV ČR, Skupina fyziky slunečních erupcí a protuberancí – dalekohled SORT (fyzika)
- Astronomický ústav AV ČR, Sluneční patrola (fyzika)



## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

- Masarykův onkologický ústav, žadatelé o vzorky a data uchovávaná v biobankách (medicínský výzkum)
- Ústav fyziky plazmatu AV ČR, sekce Fúzní plazma (fyzika)
- Ústav fyzikální chemie J. Heyrovského AV ČR, Heyrovsky Open Science Club (fyzikální chemie)
- CEITEC, CF BioData – NCBR, Research group Structural bioinformatics (strukturní bioinformatika)
- CEITEC, CF CELLIM (světelná mikroskopie)
- CEITEC, Výzkumná skupina Roberta Váchy (počítačová simulace biomolekul)
- CEITEC, Laboratoř multimodálního a funkčního zobrazování – MAFIL (medicínské zobrazování, neurozobrazování, mapování mozku)
- CEITEC, Bioinformatics Core Facility (sekvencování, bioinformatika)
- MENDELU, obecně za univerzitu
- Český sociálně-vědní datový archiv, Sociologický ústav AV ČR (sociální a humanitní vědy)
- CESNET / e-INFRA CZ

Je tedy patrné, že jednotlivé odpovědi často reprezentují entity neporovnatelné velikostí, počtem uživatelů i objemem dat. To nicméně nebylo cílem.

Průzkum byl rozdělen do několika **problémových oblastí**. Následuje jejich agregovaný přehled:

- Specifika přístupu k datům zejména z hlediska potřeby řízení přístupu
- Stav ukládání a případného zveřejňování dat
- Pokud skupina provozuje repozitář, na jakém softwarovém systému
- Pro živá data (tj. data, se kterými se aktivně pracuje)
  - Kde jsou uložena
  - Jak jsou zálohována
  - Jsou strojově čitelná
- Kde jsou zpracovaná data zveřejňována
- Jak se řeší přístupová práva k datům pro řízené zveřejnění
- Metadatové popisy dat
- Objemy dat
- Vyhledatelnost (ve vlastním vyhledávači, nástrojích jako Google Scholar a podobně)
- Stav integrace s nástroji jako laboratorními deníky (zejména aby nebylo nutné ručně zadávat strojově dostupná metadata)
- Lidské zdroje pro správu dat
- Formáty dat a metadat, původ dat, transformace datových formátů, archivace, přidělování perzistentních identifikátorů

Jednotlivá témata byla pro vyplnění nepovinná, nelze pokrýt celou rozmanitost současného stavu jednotným dotazníkem. Instrukce nabádaly k vyplnění částí, které jsou pro danou skupinu relevantní.

Výše uvedené problémové oblasti byly poptávány ve dvou relativně samostatných režimech, a to:

- Současný stav
- Představa o ideálním stavu cca v roce 2025

**Současný stav** poskytuje jednak přehled o současném fungování jednotlivých skupin, jednak o dostupné expertíze v jednotlivých tématech. Přestože většina respondentů byla v představě stavu pro rok 2025 relativně stručná, tato část dává vhled do jejich uvažování o potenciálním rozvoji (byť

poplatné době průzkumu – vývoj problematiky je velmi dynamický) a zejména zachycuje témata, která by neměla v plánované infrastruktuře národní repozitářové platformy zůstat opomenuta.

Výstupy této analýzy jsou silně diverzifikované. Obecně lze říci, že úroveň v současnosti používaných řešení silně kolísá, od dat na lokálních discích, přes ad-hoc metody ukládání dat na místě nebo na zdroje e-infrastruktury, až po depozici do oborového repozitáře včetně smysluplných metadat. Většina respondentů v současnosti ukládá živá data pouze do souborových systémů, je tak obtížné udržet konzistenci dat a příslušných metadat. Skupiny budou také potřebovat nástroje pro řízení přístupu uživatelů a pro řízení životního cyklu zveřejnění dat, tyto v současnosti mají k dispozici pouze v základní podobě (pro soubory, nikoli pro datové sady jako logické entity). Pouze tři respondenti ukládají finalizovaná data do repozitářů (každý s vlastním technickým řešením). Jeden respondent je součástí mezinárodního týmu vyvíjejícího a provozujícího rozsáhlou infrastrukturu datových repozitářů. Další mají pro data systémy, které nějaké rysy repozitáře vykazují (nicméně nazývat je repozitáři by bylo zcela neadekvátní). Majorita respondentů nicméně i finalizovaná data ukládá v souborových systémech (se všemi riziky ztráty konzistence dat a metadat), případně pro zveřejnění přístupná přes webové rozhraní.

Na široké škále se pohybuje také úroveň technické péče o data, typicky je to na vlastních hardwarových prostředcích se zálohováním do e-infrastruktury. Jen v jednotlivých případech jsou data opatřena perzistentními identifikátory či metadata propagována k agregátorům metadat. Žádný z respondentů nevytváří archivní balíčky ani nemá manipulaci s daty provázanou s plánem správy dat. Vzácně narazíme na nějakou automatizaci pořizování metadat, obvykle v podobě ad-hoc skriptů. Ohledně personálního zajištění typicky nejsou k dispozici specialisté věnující se převážně nebo výhradně datové problematice.

Z provedeného průzkumu vyplývá, že ani skupiny, které explicitně projevují zájem o správu vědeckých dat (což projevují zapojením v pracovních skupinách) nejsou v naprosté většině v současnosti připraveny na požadavky FAIR dat. Obvykle provozují specializované řešení vytvořené na míru lokálních potřeb. To vede na duplikaci činností a snižuje efektivitu vědecké práce. Výsledky průzkumu dokládají, že pro implementaci evropské politiky zacházení s vědeckými daty a širší přijetí FAIR principů je naprosto nezbytný koordinovaný přístup přes vědecké skupiny a disciplíny, který zahrnuje jednak společné budování federované infrastruktury pro ukládání anotovaných dat<sup>15</sup> a jednak výchovu a vzdělávání specialistů v této oblasti, kteří pak ve správě dat budou vědeckým skupinám k dispozici.

Z představ respondentů o ideálním stavu **cca v roce 2025** lze odvodit, že by pro naplnění očekávání vědeckých skupin měly být k dispozici následující komponenty:

- Národní metadatový adresář, který agreguje metadata z repozitářů v Národní repozitářové platformě a umožňuje výměnu metadat s mezinárodními infrastrukturami. Základní metadatový adresář dále bude umožňovat vyhledávání; bude rovněž sklizen do vyhledávače vytvořeného v PNG a agregujícího metadata všech elektronických i tištěných dokumentů, a tedy i datových sad v ČR na jednom místě.
- Pro ukládání záznamů do repozitářů jsou nezbytné perzistentní identifikátory<sup>16</sup>. Infrastruktura bude proto provozovat technický systém pro jejich přidělování datovým sadám.
- Národní repozitářová platforma bude sloužit pro vlastní ukládání dat do repozitářů. To zahrnuje provoz hardwarové infrastruktury, vytvoření a provoz systému pro vytváření tenantů

<sup>15</sup> Tj. vědeckých dat s připojeným metadatovým popisem.

<sup>16</sup> Obstarávání všech perzistentních identifikátorů (s výjimkou ISBN a ISMN, které zajišťuje Národní knihovna ČR) zajistí projekt CARDS.

rezpozitářů pro potřeby uživatelských skupin a institucí, včetně zajištění catch-all tenantu – instance obecného datového rezpozitáře. Národní rezpozitářová platforma bude zajišťovat archivaci uložených dat, optimálně v geograficky oddělených lokacích a s podporou periodických kontrol binární konzistence.

- Optimální by byla dostupnost prostředí pro citlivá data (osobní, medicínská).
- Celý systém bude správou identit uživatelů navázán na národní e-infrastrukturu, identity uživatelů budou sloužit k řízení přístupu minimálně pro tenanty Národní rezpozitářové platformy.
- K použití Národní rezpozitářové platformy a zejména vytváření instancí/tenantů v ní prostřednictvím API bude k dispozici dokumentace a také pokročilá uživatelská podpora.
- Budou k dispozici nástroje pro přenosy dat do rezpozitářů.
- Celý systém bude plně integrován do dohledu v oblasti kyberbezpečnosti ve spolupráci s národní e-infrastrukturou.

V oblasti **rozvoje lidských zdrojů** pro práci s daty respondenti předpokládají jejich posílení. Očekávané jsou zejména znalosti z oblasti péče o data. Některé instituce budou u svých datových specialistů vyžadovat rovněž určitou míru znalosti konkrétní problematiky výzkumu, která jim umožní data správně zpracovat a popsat. Jiné instituce předpokládají v tomto ohledu spolupráci mezi datovými specialisty a výzkumníky, která ovšem více zatíží výzkumníky a vyžádá si jejich edukaci v oblasti péče o data a datové sady.

### 5.6 Návrh typologie pracovních pozic pro implementaci EOSC

Pracovní skupina pro vzdělávání a lidské zdroje navrhuje pro další komunikaci a rozvoj dané oblasti pracovat s níže uvedenou typologií pracovních pozic potřebných pro implementaci EOSC v ČR<sup>17</sup>. Tato typologie by měla podpořit jak srozumitelnost komunikace dané problematiky z hlediska interního rozvoje v jednotlivých institucích, tak i spolupráci na národní a mezinárodní úrovni. Daná typologie, včetně příkladů náplně práce by měla pomoci i při intervencích dotačních programů (nejen OP JAK) a stanovování mzdového a kvalifikačního rámce pro vytváření nových pracovních pozic, které budou vznikat v českých institucích. Stanovená typologie je pragmatickým výtahem zejména zahraničních zdrojů<sup>18</sup>, provedena byla též “jobs analysis” v relevantních portálech zahraničních poskytovatelů pracovních pozic.

- **Vědec** – Každý vědec by měl mít následující dovednosti: práce s daty dle oborového standardu (zálohování, popis); práce s daty v souladu s institucionální politikou, ale i institucionální podporou; chápání významu a cest k naplňování FAIR principů; základy bezpečného nakládání s daty; terminologické základy (data, metadata, PID atd.); zásady citování datových výstupů; základy problematiky licencování dat (co mohu a nemohu); základy transferu a komercializace technologií; základy problematiky datových formátů; práce s daty v intencích projektové podpory; etika práce s výzkumnými daty (např. nakládání s osobními údaji, problematika sběru dat od vulnerabilních skupin apod.)

---

<sup>17</sup> Práce na typologii pracovních pozic se účastnili zástupci těchto organizací: Archeologický ústav AV ČR, ČSDA SOC AV ČR a 1. LF UK, e-INFRA CZ/MUNI, FLU AVCR – ČZU, FZÚ AV ČR, IPHYS Bioimaging facility, FGU AV CR, Knihovna AV ČR, LIB CZU Prague, Masarykova univerzita, Mendelova univerzita v Brně, Ministerstvo školství, mládeže a tělovýchovy, Národní technická knihovna, RECETOX MU, SKRVŠ ČZU v Praze, Univerzita Karlova, Univerzita Palackého v Olomouci, Ústav vědeckých informací 2. LF UK a FN Motol, VŠB-TUO, VŠCHT Praha.

<sup>18</sup> Identifying digital skill sets for EOSC. <https://www.eoscsecretariat.eu/news-opinion/identifying-digital-skill-sets-eosc>

- **Data scientist** – Poučenější vědec se širší znalostí ve specifických oblastech, jako jsou statistika, grafika, fotografie, text mining, prostorové zobrazení apod. Tato typová pozice je pracovní skupinou vnímána jako zástupce hlavní cílové skupiny implementace politiky EOSC, tj. spíše jako evoluční stupeň obecného vývoje vědce.
- **Data steward** – Správce dat, metadat a jejich zpracování, k jehož dovednostem patří: tvorba data management plánů; podpora FAIRifikace dat vytvářených vědci; problematika volby vhodných datových formátů; schopnost navrhovat a organizovat pracovní datová úložiště; znalost nároků kladených na data ze strany EOSC/NRP; orientace v dostupných standardech, metadatových schématech a heslářích; schopnost posuzovat kvalitu dat a metadat (formálně a technicky); dobrá orientace v problematice PID; orientace v etických a právních otázkách spojených s daty; schopnost základních datových transformací (slučování, párování, selekce, filtrování, převod mezi různými formáty atd.); schopnost být mediátorem datové politiky v rámci instituce; podpora výzkumníků v akademické sféře (otázky a odpovědi na témata správy dat, datových repozitářů, metadatového popisu, FAIR data a Open Science principů); zkušenosti s metadatovým popisem výzkumných dat; zkušenosti s praxí ukládání výzkumných dat do institucionálních a oborových datových repozitářů (hledání vhodného oborového repozitáře, oborové standardy i mimo ČR); mediace komunikace mezi výzkumníky a IT podporou pro práci s daty (ukládání, zpracování); praxe práce s daty v národních (GAČR, TAČR) a mezinárodních (Horizon Europe apod.) projektech (požadavky na data management plán apod.); praxe práce s daty při publikování (datové publikace, proces zpřístupňování dat pro recenzní řízení apod.); tvorba a podpora plnění data management plánů.
- **Data curator** – Specialista, který pečuje o IT archiv, definuje datové standardy a spravuje repozitář (validace formátů, správa a workflow). K jeho dovednostem patří: schopnost organizovat budování, provoz a rozvoj datových repozitářů; široká orientace v oblasti standardů, možných technických řešení, ontologií apod.; znalost mezinárodního prostředí a best practice ve vztahu k datové správě; schopnost provazovat nezávislé datové zdroje a budovat nad nimi nadstavbová řešení; schopnost finančního a kapacitního plánování ve vztahu k datové správě a její udržitelnosti; přehled o možnostech data re-use a schopnost podpory vědeckého využití dat.
- **Datový právník** – Právník pro oblast IT, k jehož dovednostem patří: znalost národní a mezinárodní legislativy, principy LTP; integrace s AAI a jinými systémy; podpora uživatelů repozitáře; compliance s předpisy na provoz repozitářů (CoreTrusSeal apod.); kyberbezpečnost, monitoring provozu; znalost transferu a komercializace technologií; znalost autorského práva a licencování obsahu (včetně licence Creative Commons pro díla a open source licence pro programy); znalost databázového práva; znalost ochrany osobních údajů; znalost právních dopadů etických aspektů výzkumu; znalost právních aspektů provozování softwarových služeb.

Pracovní skupina hodnotila i role typu data librarian, data ombudsman, manažer akademického výzkumu apod., ty však byly vyhodnoceny jako méně jasně určené, pro české prostředí matoucí a jejich činnost je zahrnována buď do výše uvedených pracovních pozic, nebo evolučním procesem navazují na činnosti a pozice "analogové vědy". Např. manažer akademického výzkumu je pozice, která existuje již dlouho (byť v českém prostředí je často nazývána např. vedoucí odboru vědy).

## 5.7 Hodnocení současného stavu a potřeb rozvoje – shrnutí rozhovorů s experty

V rámci přípravy studie proběhla na přelomu května června 2022 série řízených rozhovorů s uživateli, experty a stakeholdery zapojenými do rozvoje EOSC.

Uživatelskou skupinu respondentů tvořili:

- **prof. RNDr. Jan Hajič, Dr.** (ÚFAL MFF UK), zástupce infrastruktury LINDAT/CLARIAH-CZ a vedoucí oborově-vědní pracovní skupiny Humanities and the Arts
- **RNDr. Jiří Chudoba, Ph.D.**, FZU AV ČR, vedoucí oborově-vědní pracovní skupiny Physics
- **prof. RNDr. Jana Klánová, Ph.D.**, (RECETOX MUNI) zástupkyně infrastruktur RECETOX a EIRENE a vedoucí oborově-vědní pracovní skupiny Environmental Sciences
- **doc. RNDr. Jiří Vondrášek, CSc.** (ÚOCHB AV ČR), zástupce infrastruktury ELIXIR CZ a vedoucí oborově-vědní pracovní skupiny Bio/Health/Food

Expertní skupinu tvořili:

- **RNDr. David Antoš, Ph.D.** (CESNET), vedoucí pracovní skupiny Architektura NDI
- **PhDr. Radka Římanová, Ph.D.** (Ústřední knihovna UK), vedoucí pracovní skupiny Vzdělávání a lidské zdroje

Skupinu stakeholderů tvořili:

- **doc. MUDr. Marian Hajdúch, Ph.D.** (Ústav molekulární a translační medicíny Lékařské fakulty UP v Olomouci), zástupce infrastruktury EATRIS-ERIC, člen Rady pro výzkum, vývoj a inovace
- **RNDr. Jan Hrušák, CSc.** (ÚFCH J. Heyrovského AV ČR), zástupce ESFRI
- **prof. RNDr. Luděk Matyska, CSc.** (ISC MUNI), zástupce infrastruktury e-INFRA CZ
- **Ing. Martin Svoboda**, ředitel NTK, zástupce projektu CARDS (IPs CARDS)

Všem respondentům byla položena **společná sada otázek** s cílem zjistit jejich potřeby, očekávání a názory na potřebné směry rozvoje EOSC. Poslední otázka byla zaměřena na udržitelnost a možnou spoluúčasť uživatelů na provozu EOSC. Respondenti odpovídali na tyto otázky:

- Co od EOSC očekáváte? Proč je dobré EOSC v ČR rozvíjet?
- Jaký je váš současný model práce s vědeckými daty? Na jaké problémy narazíte? Jak rozvoj EOSC změní vaši praxi? (otázka jen pro skupinu uživatelů)
- Jak vnímáte současný stav implementace EOSC v ČR? Jaká jsou specifika v ČR oproti jiným státům?
- Jakým směrem by se měl rozvoj EOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOSC?
- Jaké technologie navrhujete využít? Jaké nové služby implementace EOSC přinese? Jaké požadavky budou kladeny na uživatele? (otázka jen pro skupinu expertů)
- Jaké aktivity jsou pro rozvoj EOSC v ČR klíčové? V jakém časovém horizontu by měly být realizovány?
- Jak hodláte dál rozvíjet své systémy v kontextu EOSC? Do čeho, v jakém časovém horizontu a kolik hodláte investovat? (otázka jen pro skupinu uživatelů)
- Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčasť uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOSC)?

Následující přehled přináší stručné shrnutí odpovědí získaných od respondentů. Detailnější záznamy z rozhovorů jsou k dispozici v příloze.

**Očekávání spojená s EOOSC:**

- Posune ČR a její výzkumnou komunitu směrem k implementaci principů Open Science v oblasti strukturované práce s výzkumnými daty. Dojde ke konsolidaci směrem k mezinárodním standardům a zvyklostem.
- Jedná se o dostupnost dat pro vědu i ostatní uživatele. Potřebujeme, aby se data neukládala náhodně, ale aby existovala datová úložiště, do kterých se data ukládají řízeně s velmi dobrými metadatovými popisy. Už je překonaná představa, že se vytvoří centrální úložiště a databáze, do kterých budou lidé všechny vzorky ukládat. Je třeba vytvářet systém úložišť, který umožní provádět federované analýzy. Tj. bez potřeby přenášet data jinam mohou uživatelé v povolené míře k datům přistupovat, na místě je analyzovat a získat výsledek.
- Přinese povědomí, že data jsou cenná a má smysl je popisovat a sdílet. Uživatelé by měli zjistit, že když mají data v pořádku a jsou připraveni tato data otevřít dalším zájemcům, lépe se jim bádá a spolupracuje.
- Na mezinárodní úrovni umožní zapojení naší vědecké komunity do mezinárodní spolupráce. Dobře popsaná (ideálně s využitím standardů) a přístupná data představují kompetitivní výhodu pro spolupráci.
- Přinese i možnost dohledání vědeckých dat v případě podezření z manipulace nebo potřeby ověření hypotézy, která byla na konkrétních datech postavena.
- Výhodou by mělo být, že EOOSC představuje bezpečný a transparentně regulovaný prostor, ve kterém je možné s daty operovat. Na rozdíl od komerčních služeb, které přinášejí nejistotu bezpečnosti a kontinuity.
- V rámci spolupráce ve velkých mezinárodních projektech se péče o data daří, ale jsou i jiné projekty, kde na to nejsou zdroje ani kapacity. Pak jsou tu i lokální skupiny, které data sice produkují, ale nemají je kde uchovávat. Po skončení grantu bohužel data mizí nebo nejsou k nalezení. EOOSC by měl přinést na národní úrovni hardwarové zdroje pro archivaci těchto dat a zároveň lidi, kteří by ta data dokázali popsat a zdokumentovat pro další využití.
- Schopnosti IT zpracovávat čím dál větší data budou mít čím dál větší vliv i na základní principy vědecké komunikace. Například ve vědách o živé přírodě se mluví o tom, že lineární vyprávění o nějaké znalosti formou publikace začínají vnímat jako zbytečnost, že znalosti jsou rozumně zachycené jen v datech. To pak vyžaduje velmi komplexní podporu pro jejich zpracování.
- Principy EOOSC by měly mít přesah i k datům ve veřejné správě, ať už na evropské nebo národní úrovni. Tato data by měla být rovněž spravována obdobným způsobem. I data z veřejné správy jsou využitelná ve výzkumu.

**Pohled na současný stav v ČR:**

- Některé české subjekty jsou v EOOSC aktivitách na mezinárodní úrovni zapojeny již dlouho. S výjimkou těchto subjektů se však implementace EOOSC v ČR zatím nerozvinula.
- Ponětí o EOOSC ve vědecké komunitě zejména v regionech mimo Prahu, Brno, Olomouc a Ostravu je zatím poměrně slabé.
- Prostřednictvím pracovních skupin pro rozvoj EOOSC se postupně daří aktivovat lidi, kteří se chtějí podílet na systematické přípravě dalšího rozvoje.
- Ani většina zájemců o zapojení do pracovních skupin pro rozvoj EOOSC nemá v současné době velké povědomí o kultuře práce s daty, životním cyklu dat nebo data management plánech. Je třeba zajistit profesionály, které tyto znalosti institucím zprostředkují.
- Politické rozhodnutí o podpoře rozvoje EOOSC prostřednictvím OP JAK umožňuje plánovat další rozvoj.

- Rozvoj EOSC se týká především akademické sféry. Průmysl má svoje specifika a vzhledem k EOSC je na místě určitá obezřetnost. Větší potenciál z hlediska EOSC má přesah k veřejným datům ve veřejné správě, kde však zatím nejsou žádné významnější aktivity patrné.
- Na některých vysokých školách stále chybí zpracované datové politiky. Velký rozdíl je v připravenosti mezi univerzitami zejména v Praze, Brně a Ostravě a jinými univerzitami, kde nejsou k dispozici lidské zdroje, a tato oblast se zatím stále řeší spíš na bázi ad hoc případů náhlé potřeby než komplexně.
- Z hlediska implementace EOSC chybí jasné strategické stanovisko k tomu, co lze a nelze v rámci intervence uplatnit a jakým směrem má instituce svou činnost napřít, aby nedocházelo k duplicitnímu financování.
- Současný přístup je dobrý v tom, že má vzniknout repozitářová platforma, která umožní jednoduché zakládání a práci s asociovanými repozitáři institucím, oborově-vědním skupinám i dalším zájemcům bez ohledu na jejich vlastní znalosti technických principů péče o repozitáře a jejich zázemí..
- Naprosto klíčovým tématem je spolupráce. V tom je projekt dobře nastaven, že většinu energie nevěnuje na technický návrh, ale na domluvu na standardech mezi aktéry.
- Máme docela silnou a vysoce integrovanou e-infrastrukturu, takže z technologického hlediska je budování infrastruktury pro EOSC spíše evolučním krokem. Náročným úkolem při tom však je hledání široké multioborové shody, jak má taková infrastruktura vypadat a co přesně má dělat, aby pak výsledek byl použitelný pro velmi heterogenní obory.
- Česko je slušně připravené díky výzkumným infrastrukturám. Pracoviště, která vstupují do mezinárodních projektů, mají zkušenosti z mezinárodních konsorcií. Díky tomu zde existuje struktura lidí, kteří vědí, co mají dělat.
- Vzhledem k malé připravenosti vědecké komunity i implementace EOSC může vzniknout problém s povinným ukládáním dat do repozitářů již v rámci dnes vypisovaných výzev. Bylo by proto dobré zvážit delší přechodné období, abychom tím vědeckou komunitu od smysluplného nakládání s daty spíše neodradili.

### **Potřebné směry rozvoje:**

- Vytvoření spolehlivého zázemí, které budou lidé moci pohodlně používat (metadatový adresář, národní repozitářová platforma, oborové clustery). Nemůže to být jediné unifikované řešení, musí být respektován současný stav a potřeby oborových clusterů.
- Je nezbytné od počátku v realitě zavádět postupy užívání standardizovaných metadatových schémat. Bez kvalitního popisu obsahu datové sady dokáže data najít jen ten, kdo zná autora nebo číslo projektu, ale jiný zájemce nemá možnost data najít podle toho, k jakému tématu se vztahují.
- Specifickou oblast, kterou je třeba řešit, tvoří citlivá data, se kterými pracují nejen lékaři, ale třeba i ekonomové nebo sociologové.
- Je potřeba pečlivě zvažovat, jak dlouho má smysl konkrétní data uchovávat. To se liší v rámci oborů i v rámci jednotlivých projektů.
- Uživatelé by měli být motivováni a za dobrou práci s daty odměněni např. oceněním v rámci kariérního růstu. Dnešní systém hodnocení vědy oceňuje různé vědecké výsledky (např. publikace), ale neoceňuje práci s daty.
- Nedává smysl hledat zcela univerzální přístup pro všechny vědní obory. Už jen sjednocení metadat napříč různými obory nedává smysl, protože i metadata jsou v různých oborech logicky různá. Např. pro jazyková data potřebujeme jiné údaje než pro data chemická.

Současně chemik bude mít málokdy zájem propojovat chemická data s jazykovými. Dobrá a užitečná integrace dat se tedy dá udělat spíše po oborech.

- Potřebujeme uskutečnit kulturní a systémovou změnu. V ČR máme strukturovanou metodiku hodnocení vědy a bylo by potřeba zavést do ní motivační parametr pro sdílení dat. Další motivací pro vědce je přístup k budoucím finančním zdrojům navázaným na EOOSC. Pokud se takto podaří vědce motivovat, po určité době se projeví přínosy sdílení dat i ve vědních oborech, kde zatím nejsou patrné.
- Je třeba podpořit nejen národní infrastrukturu a oborové klastry, ale i jednotlivé instituce. Na počátku diskuse byl významně akcentován oborový prvek, ale ne všechny oblasti jsou takto jednoznačně klasifikovatelné. Centrální národní řešení nemůže roli konkrétních instituce plnohodnotně nahradit, protože se jedná o individuální dlouhodobou práci s jednotlivými vědci v rámci těchto institucí. Dalším důvodem jsou citlivá data např. z medicínského výzkumu ve fakultních nemocnicích, která by neměla opustit institucionální repozitář.
- Je třeba zamezit bezbřehé debatě. Při implementaci EOOSC je potřeba vyjít z reálných zkušeností, a až pak si nechat takto vytvořený rámec připomínkovat od dalších uživatelů.

### **Klíčové aktivity pro rozvoj:**

- Práce s lidskými zdroj. Vysvětlit vědecké komunitě, že dobrá práce s daty je nedílnou součástí běžné vědecké práce.
- Důležité je dobré PR, které každého vědce seznámí s konceptem EOOSC tak, aby se s ním dokázal identifikovat buď na základě vlastních potřeb, nebo potřeb své instituce.
- Nejdůležitější aspekt je rozvoj lidských zdrojů. Na druhou stranu je rovněž zajímavý přístup k (centrálně) spravovanému technickému vybavení, které poskytuje prostor pro ukládání vlastních dat..
- Důležité je pokrýt práci s daty již ve studijních programech a pokračovat celoživotním vzděláváním.
- Základem jsou lidské zdroje. Najít vědce, kteří s daty dříve aktivně pracovali, a jsou ochotni upustit od osobní vědecké kariéry a začít pracovat s daty pro celek, není vůbec snadné. Musí mít logické myšlení a nadhled a přitom rozumět podstatě dat v konkrétním oboru. Na to bude zřejmě potřeba větší investice na straně oborových klastrů.
- Důležitá je koordinace aktivit z hlediska zákonů (např. autorské právo) a dobrých praxí.
- Je potřeba rozvíjet autentizační a autorizační služby.
- Důležitá je specifická aktivita kolem citlivých dat napříč různými obory, zaměřená na specifické technické požadavky z hlediska zabezpečení, specifické právní požadavky, a také specifické požadavky ze strany uživatelů, případně i státu, který citlivá data poskytuje. Tato aktivita by měla vést ke vzniku kompetenčních center, která poskytnou zabezpečené prostředí pro nakládání s citlivými daty a pro jejich zpracování.

### **Udržitelnost a spoluúčast uživatelů:**

- Výzkumné infrastruktury mají vlastní plány udržitelnosti. O potřebné velikosti Národní repozitářové platformy a potřebě její podpory získáme přesnější údaje kolem r. 2025.
- Data a EOOSC jsou důležitým argumentem v diskusi, zda dále investovat veřejné prostředky do rozvoje infrastruktur. Důležité je vysvětlovat, že vstupujeme do věku, kdy jsou dobře spravovaná a maximálně otevřená data pro vědu životně důležitá.
- Většina vědy se dělá ve veřejném prostředí, takže případné spolufinancování je jen otázka přesměrování veřejných zdrojů.



- Argument placení obvykle pomáhá k zefektivnění systému. Na začátku je dobré poskytnout prostředky na rozjezd a pak se může zvážit spolufinancování, případně plně samofinancování, pokud je vidět jasný přínos dané aktivity. Zejména u malých institucí to ale může znamenat, že s koncem grantu skončí i činnost.
- Teď je třeba nabídnout práci datových expertů (lidské zdroje). Kontaktem výzkumníků s nimi se know-how rozšíří a výzkumníci už si dokáží plánovat kapacitu lidských zdrojů na další projekty. Spoluúčast výzkumníků bude v tom, že věnují vlastní čas na komunikaci s datovými experty.
- Podle zkušenosti z projektu CzechELib (NTK) je rozumný model dělby nákladů 50 % stát a 50 % uživatelé. Pokud nemusíte platit, nebudete si toho vážit. Poskytnutí zcela zdarma pro rozjezd je rozumné, ale mělo by být limitováno na nějakou omezenou dobu.
- Naproti tomu CESNET má dlouhodobou zkušenost, že kdekoli došlo na spoluúčast uživatelů, tak to spíše používání infrastruktury poškodilo, než podpořilo. Samozřejmě data dokážou růst nade všechny meze a jejich skladování rozhodně zadarmo není, takže nějaký model udržitelnosti potřeba bude, nicméně modelem určitě není spoluúčast uživatelů. Spíše potřebujeme kvalifikované datové kurátory, kteří budou schopni rozhodovat pro každý typ dat o tom, jak dlouho je má smysl uchovávat.
- Vhodné by bylo oddělit přístup k uživatelům z komerční a akademické sféry. Nebylo by správné, aby výzkumníci z veřejné sféry platili za data vytvořená z veřejných prostředků.

## 6 Shrnutí a závěry

Stejně jako v jiných státech EU ani v ČR **EOSC nevzniká na zelené louce**. Existují zde pracoviště a infrastruktury, které jsou z hlediska FAIR práce s vědeckými daty na velmi pokročilé úrovni (viz 5.2 Přehled vybraných rozvinutých repositářů). Jedná se zejména o entity napojené na větší evropské projekty či infrastruktury, které jim umožňují sbírat potřebné zkušenosti a mnohdy poskytují i technické a procesní zázemí pro FAIR práci s vědeckými daty. Na druhém konci spektra je nicméně řada týmů a institucí, které mají ohledně FAIR práce s daty jen minimální povědomí, zkušenosti a motivaci. Výchozím předpokladem je proto **důsledná propagace myšlenky a výhod EOSC** směrem široké vědecké obci, tj. i do regionů mimo velká centra výzkumu v Praze, Brně, Ostravě nebo Olomouci.

Na návrhu různých aspektů implementace EOSC v současné době pracuje několik průřezových a oborově-vědních pracovních skupin, které se formují z řad zájemců o aktivní zapojení do návrhu a rozvoje EOSC. Lze tedy očekávat, že **návrh celé struktury EOSC i jednotlivých dílčích projektů bude výsledkem širší odborné diskuse** a bude respektovat potřeby aktivně zapojených účastníků.

Navrhovaná **výše podpory prostřednictvím programu OP JAK** se jeví v mezinárodním kontextu v porovnání se skupinou států v pokročilé fázi implementace EOSC (Německo, Nizozemsko, Francie, viz 4.3.1 Příklady států v pokročilé fázi implementace) jako odpovídající a rovněž návrh struktury Národní datové infrastruktury se v Česku ubírá podobným směrem jako v těchto pokročilých zemích. Význam poskytnutí této podpory spočívá také v tom, že právě díky ní je možné koncepční plánování dalšího rozvoje EOSC na úrovni, která umožní udržet českou vědu z hlediska práce s daty na potřebné evropské úrovni.

Provedené průzkumy i expertní rozhovory ukazují, že pro implementaci evropské politiky zacházení s vědeckými daty a širší přijetí FAIR principů je nezbytný koordinovaný přístup přes vědecké skupiny a

disciplíny, který zahrnuje jednak **společné budování federované infrastruktury pro ukládání anotovaných dat**<sup>19</sup> a jednak výchovu a vzdělávání specialistů v této oblasti.

Je žádoucí od počátku zavádět postupy **užívání standardizovaných metadatových schémat**. Bez kvalitního popisu obsahu datové sady dokáže data najít jen ten, kdo zná autora nebo číslo projektu, ale jiný zájemce nemá možnost data najít podle toho, k jakému tématu se vztahují.

Podstatná část zamýšlené podpory by měla být použita především na **vzdělávání a rozvoj lidských zdrojů** v oblasti práce s vědeckými daty a datovými sadami. Vzdělávání by mělo být zaměřeno na celoživotní výchovu vědců ke správnému nakládání s vědeckými daty. Nedílnou součástí těchto znalostí musí být i **principy tvorby a používání data management plánů** (DMP, viz 7.1.4 Data management plány). Jako druhý pilíř aktivit v oblasti lidských zdrojů se jeví **výchova datových specialistů** s různými stupni kvalifikace (viz 5.5 Návrh typologie pracovních pozic pro implementaci EOSC) a podpora jejich zapojení do práce výzkumných týmů. Kapacitně lze odhadnout, že z předpokládané dotace 2,5 mld. Kč pro výzvy Open Science OP JAK lze pro následujících šest let zajistit výchovu a působení datových specialistů napříč všemi významnými výzkumnými obory. Zatím otevřenou otázkou ovšem je, zda se podaří získat dostatek vhodných zájemců o tyto pozice a rovněž jakým způsobem bude řešeno návazné financování těchto specialistů v období po skončení intervence. Bude nezbytné najít vhodnou rovnováhu mezi zvyšováním a rozšiřováním kompetencí samotných vědeckých pracovníků a výchovou nových specialistů.

Závěry z provedených průzkumů a expertních rozhovorů dále ukazují, že je účelné **využít již existujících řešení v rámci oborově-vědních a institucionálních klastrů**, nevymýšlet věci znovu, ale naopak zúročit již existující zkušenosti s budováním a provozem oborových repozitářů a podpořit jejich propojení s nově připravovanou Národní repozitářovou platformou (NRP).

Vzhledem ke stále majoritnímu počtu vědců a vědeckých týmů nezapojených do stávající struktury vědně oborových repozitářů je významným krokem implementace EOSC **vytvoření Národní repozitářové platformy jako multi-tenant platformy**, v níž je možné vytvářet konkrétní datové repozitáře asociované se všemi ostatními složkami NDI a která zajistí i základní společné služby spojené s tvorbou metadat a řízením přístupu k datům. Tento přístup umožní poměrně jednoduché zakládání a práci s asociovanými repozitáři institucím, vědeckým oborovým skupinám i dalším zájemcům bez ohledu na jejich vlastní znalosti technických principů péče o repozitáře a jejich zázemí. Navíc NRP vytvoří základ, včetně společného systému řízení přístupu k uloženým objektům, na něž se mohou další (oborově-vědní) repozitáře napojovat.

Důležitá je také **specifická aktivita v oblasti citlivých dat napříč různými obory**, zaměřená na specifické technické požadavky z hlediska zabezpečení, specifické právní požadavky, a také specifické požadavky ze strany uživatelů, případně i státu, který citlivá data poskytuje. Tato aktivita by měla vést ke vzniku kompetenčních center, která poskytnou zabezpečené prostředí pro nakládání s citlivými daty a pro jejich zpracování.

Mezi další nezbytné předpoklady pro úspěšnou implementaci EOSC v ČR patří **soulad s principy EOSC Interoperability Framework** (viz 7.1.2 EOSC Interoperability Framework) ve všech definovaných oblastech interoperability, tj. v oblasti technické, syntaktické, sémantické, organizační i právní. V neposlední řadě je zde třeba zmínit i potřebu **zajištění přístupu ke standardizovaným perzistentním identifikátorům** pro data, osoby i organizace (viz 7.1.3 Perzistentní identifikátory) a jejich důsledného používání při práci s daty a datovými sadami.

---

<sup>19</sup> Tj. vědeckých dat s připojeným (nejlépe standardizovaným) metadatovým popisem.

**Otázka dlouhodobé udržitelnosti** Národní datové infrastruktury přesahuje období implementace OP JAK a z provedených rozhovorů vyplývá, že k ní v současné době neexistuje jednotný přístup. Zatímco část respondentů upřednostňuje zkušenost, že nebudou-li se uživatelé NRP a oborových repozitářů na jejich fungování alespoň symbolicky finančně podílet, nebudou si jejich služeb dostatečně vážit se všemi důsledky z toho plynoucími, druhá část respondentů poukazuje na fakt, že v případě nutnosti spolufinancování by mohla řada výzkumníků a výzkumných týmů o využívání těchto služeb ztratit zájem. Konečné rozhodnutí v této věci se bude zřejmě utvářet až po důkladné analýze zkušeností s rutinním provozem, kapacitními potřebami a provozními náklady NRP a dalších repozitářů v období po roce 2024.

Na evropské úrovni je žádoucí podporovat **aktivní zapojení zájemců do práce EOOSC Association Task Forces** a dalších návazných pracovních skupin, které EOOSC Association zaštiťuje. Očekává se, že spolufinancování EOOSC na evropské úrovni bude nadále realizováno především formou nových výzev a v nich podpořených projektů. Je proto nezbytné vytvořit motivační prostředí pro **širší zapojení českých institucí a týmů do mezinárodních projektů**. To jednak zajistí kontinuální napojení na celoevropské aktivity, jednak to přinese další finanční prostředky pro budování EOOSC v ČR.

## 7 Přílohy

### 7.1 Klíčové principy pro implementaci a provoz EOOSC

Cílem rešerše je upozornit na zahraniční know-how a dobré praxe ve vybraných aspektech implementace EOOSC, které jsou významné z hlediska požadované funkčnosti, vynaložených nákladů a dlouhodobé udržitelnosti implementace EOOSC na národní úrovni.

#### 7.1.1 Reflexe principů FAIR

Principy FAIR popisují, jak by data měla být zpracována, aby byla lépe **vyhledatelná (Findable), dostupná (Accessible), interoperabilní (Interoperable) a opětovně využitelná (Reusable)**, a jsou podporovány některými významnými poskytovateli financí, například Evropskou Komisí. Principy FAIR vychází z velké části ze Zásad a pokynů pro přístup k výzkumným údajům z veřejných zdrojů. (OECD 2007) Evropská komise zjistila, že roční náklady plynoucí z neaplikování principů FAIR stojí evropské hospodářství **nejméně 10,2 miliardy EUR ročně**. Kromě toho je také řada důsledků neexistence systému FAIR, které nebylo možné spolehlivě odhadnout, jako je **dopad na kvalitu výzkumu, hospodářský obrat nebo strojovou čitelnost výzkumných údajů**. Na základě hrubé paralely s evropskou ekonomikou otevřených dat Evropská komise dospěla k závěru, že tyto nevyčíslené prvky by mohly představovat **dalších 16 miliard EUR ročně** nad rámec odhadovaných hodnot. (European Commission 2019, p. 4)

Hlavní zásady FAIR (Wilkinson 2016):

Vyhledatelnost:

- F1. Datové sadě včetně ji popisujících metadat je přiřazen globálně jedinečný a trvalý identifikátor (PID).
- F2. Data jsou popsána bohatými metadaty (definovanými v R1 níže).
- F3. Metadaty jasně a explicitně obsahují identifikátor dat, která popisují.
- F4. (Meta)data jsou registrována nebo indexována v prohledávatelném zdroji.

Dostupnost:

- A1. (Meta)data jsou vyhledatelná podle svého identifikátoru pomocí standardizovaného komunikačního protokolu.
- A1.1. Protokol je volný, otevřený a univerzálně implementovatelný.
- A1.2. Protokol umožňuje v případě potřeby procesy autentizace a autorizace.
- A2. Metadaty jsou přístupné, i když data již nejsou k dispozici.

Interoperabilita:

- I1. (Meta)data využívají formální, přístupný, sdílený a široce použitelný jazyk pro reprezentaci znalostí.
- I2. (Meta)data využívají slovníky, které se řídí zásadami FAIR.
- I3. (Meta)data obsahují kvalifikované odkazy na jiná (meta)data.

Opětovná využitelnost:

- R1. (Meta)data jsou bohatě popsána množstvím přesných a relevantních atributů.
- R1.1. (Meta)data jsou zveřejněna s jasnou a přístupnou licencí na používání dat.

- R1.2. (Meta)data jsou spojena s původem dat.
- R1.3. (Meta)data splňují komunitní standardy relevantní pro danou oblast.

V rámci iniciativy EOSC sledovala pracovní skupina FAIR vývoj metrik pro data a certifikované služby FAIR, aby mohla vydat doporučení, co z nich by mělo být přijato v rámci EOSC. V této oblasti již probíhá značná aktivita. Pracovní skupina **RDA FAIR Data Maturity Model** koordinuje mezinárodní činnost s cílem definovat společný soubor základních kritérií hodnocení FAIRness a obecný a rozšiřitelný model sebehodnocení pro měření úrovně vyspělosti datové sady. Skupina uspořádala řadu seminářů a výstupy zveřejní ve stejném časovém horizontu jako pracovní skupina FAIR.

Projekt **FAIRsFAIR** má za úkol podporovat společný vývoj a zavádění certifikačních systémů pro datová úložiště. Doporučení a výsledky tohoto projektu budou převzaty v rámci národní iniciativy INFRAEOSC a klastrových projektů. Pracovní skupina FAIR úzce spolupracuje s FAIRsFAIR a dalšími subjekty, které provádějí činnosti týkající se posuzování služeb, aby bylo zajištěno, že EOSC je důvěryhodným prostředím s odpovídajícími funkcemi. (EOSC Work Plan 2020)

Výzkum financování e-infrastruktur v rámci projektu **EOSC Pillar** se v prvním kroku tým EOSC-Pillar obrátil na zástupce každé cílové skupiny a otázkou: „Jak jste obeznámeni se zásadami FAIR týkajícími se dat?“ Respondenti byli požádáni, aby své odpovědi uvedli na čtyřbodové stupnici od „velmi dobře znám“ po „vůbec neznám“. Zástupci všech cílových skupin v průměru převážně uváděli, že jsou obeznámeni nebo velmi dobře obeznámeni s konceptem FAIR údajů. **81 % respondentů zastupujících e-infrastrukturu uvedlo, že jsou v průměru „velmi dobře obeznámeni“ nebo „obeznámeni“ se zásadami FAIR.** Ze všech zástupců e-infrastruktur uváděli francouzští respondenti nejčastěji, že jsou (velmi) obeznámeni s principy FAIR pro data (92 %), následovali němečtí respondenti (86 %), italské respondenti (82 %), belgičtí respondenti (75 %) a rakouští respondenti. 77 % všech zástupců financujících subjektů uvedlo, že jsou (velmi) dobře obeznámeni s problematikou FAIR. Ve všech zemích přibližně **tři čtvrtiny všech zástupců univerzit** uvedli, že jsou (velmi) obeznámeny se zásadami FAIR. Přesto jsou rozdíly mezi jednotlivými zeměmi u vysokých škol značné. Zatímco všichni zástupci univerzit z Rakouska (N = 12) a Belgie (N = 9) v průzkumu uvedli, že jsou (velmi) obeznámeni se zásadami FAIR, zástupci ostatních tří zemí uvedli nižší obeznámenost. 68 % z 19 zástupců francouzských univerzit odpovědělo, že jsou (velmi) dobře obeznámeni se zásadami FAIR, stejně jako 58 % z 19 italských univerzit a 52 % z 54 německých univerzit. Ve všech zemích uvedlo přibližně **70 % všech zástupců výzkumných infrastruktur**, že jsou (velmi) obeznámeni s konceptem FAIR dat. Znalost konceptu FAIR mezi výzkumnými infrastrukturami je nejvyšší v Belgii, kde 85 % respondentů uvedlo, že jsou (velmi) obeznámeni s konceptem FAIR dat, následuje Francie (78 %), Německo (72 %) a Itálie (67 %). (Bodlos et al., s. 81-82)

**Interpretace celkových nákladů spojených s absencí výzkumných údajů FAIR** jako jediné hodnoty opomíjí mnoho nekvantifikovatelných přínosů FAIR. Nicméně měřitelné náklady na neaplikování principů FAIR na zacházení s vědeckými daty v Evropě ve výši **10,2 miliardy EUR ročně** jsou přesvědčivým argumentem ve prospěch zavedení zásad FAIR. Pro představu, výdaje na výzkum v Evropě v roce 2016 činily 302,9 miliardy EUR. Zatímco minimální skutečné náklady spojené s absencí FAIR tak lze považovat za pouhých 3 % všech výdajů na výzkum, tvoří tato částka zároveň 78 % ročního rozpočtu programu Horizont 2020. Poznatky o ekonomice otevřených dat navíc naznačují, že dopad FAIR na inovace by mohl k odhadovaným minimálním nákladům přidat dalších 16 miliard eur. Tato studie sice nezohledňuje náklady na zavedení FAIR, ale pokud bychom předpokládali, že dodatečné náklady určené na správu dat činí až 2,5 % všech výdajů na výzkum, zůstalo by z implementace zásad FAIR kladné saldo ve výši cca 2,6 miliardy EUR ročně. (PwC EU Services 2018)

### 7.1.2 EOSC Interoperability Framework

Dosažení interoperability v rámci EOSC je nezbytné, aby federace služeb, které budou tvořit EOSC, poskytovala uživatelům služeb přidanou hodnotu. V kontextu se zásadami FAIR se o interoperabilitě hovoří v souvislosti s tím, že "**výzkumná data je obvykle třeba integrovat s jinými daty; kromě toho je třeba, aby data spolupracovala s aplikacemi a procesy pro analýzu, ukládání a zpracování**".

**Evropský rámec interoperability (EIF)** definuje interoperabilitu jako "schopnost organizací spolupracovat na dosažení vzájemně prospěšných cílů, což zahrnuje sdílení informací a znalostí mezi těmito organizacemi prostřednictvím obchodních procesů na bázi výměny dat mezi jejich ICT systémy" (European Commission 2017):

- **Technická interoperabilita.** Vlastnost informačního systému, jehož rozhraní jsou dobře dokumentovaná, aby mohl v současnosti nebo v budoucnu spolupracovat bez omezení s jinými IT systémy.
- **Syntaktická interoperabilita.** Pokud dva nebo více systémů používají společné datové formáty a komunikační protokoly a jsou schopny vzájemně komunikovat pomocí otevřených standardů.
- **Sémantická interoperabilita.** Zajišťuje, že přesný formát a význam vyměňovaných dat a informací je při výměně mezi stranami zachován a pochopen, jinými slovy "co je odesláno, je to, co je pochopeno".
- **Organizační interoperabilitou** se rozumí způsob, jakým organizace sladují své obchodní procesy, odpovědnosti a očekávání, aby dosáhly společně dohodnutých a vzájemně prospěšných cílů.
- **Právní interoperabilita** spočívá v zajištění toho, aby organizace působící podle různých právních rámců, politik a strategií mohly spolupracovat. To může vyžadovat, aby právní předpisy nebránily zřizování evropských veřejných služeb v rámci členských států a mezi nimi a aby existovaly jasné dohody o tom, jak řešit rozdíly v právních předpisech přesahující hranice, včetně možnosti zavést nové právní předpisy.

Následující přehled shrnuje minimální doporučení pro **EOSC Interoperability Framework**, uspořádaná podle jednotlivých vrstev (EOSC 2021a, s. 27)<sup>20</sup>:

Technická interoperabilita:

- Otevřené specifikace pro služby EOSC.
- Společný rámec pro zabezpečení a ochranu soukromí (včetně infrastruktury pro autorizaci a autentizaci).
- Snadno srozumitelné dohody o úrovni služeb pro všechny poskytovatele zdrojů EOSC.
- Snadný přístup ke zdrojům dat dostupným v různých formátech.
- Hrubé a jemné nástroje pro vyhledávání datových sad (a jiných výzkumných objektů).
- Jasná politika EOSC PID.

Sémantická interoperabilita:

- Jasná a přesná, veřejně dostupná definice všech pojmů, metadat a datových schémat.
- Sémantické artefakty nejlépe s otevřenými licencemi.
- Související dokumentace k sémantickým artefaktům.

---

<sup>20</sup> Syntaktická interoperabilita není v tomto dokumentu řešena.

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

- Úložiště sémantických artefaktů, pravidla s jasným rámcem správy.
- Minimální metadatový model (a křížové přechody) pro usnadnění vyhledávání v existujících federativních výzkumných datech a metadatach.
- Možnosti rozšiřitelnosti, které umožní využití oborových metadat.
- Jasně protokoly a stavební bloky pro federaci/sklizeň sémantických katalogů artefaktů.

### Organizační interoperabilita:

- Doporučení pravidel spolupráce zaměřených na interoperabilitu.
- Doporučení pro používání standardizovaných datových formátů a/nebo slovníků a jim odpovídajících metadat.
- Přehledná správa stálých názvů a funkcí organizací<sup>21</sup>.

### Právní interoperabilita:

- Standardizované lidsky a strojově čitelné licence s centralizovaným zdrojem znalostí a podpory v oblasti autorských práv a licencí.
- Volné licence pro metadata (a pokud možno i pro data, kdykoli je to možné).
- Identifikace různých částí datové sady s různými licencemi.
- Jasně označené případy prošlých nebo neexistujících autorských práv, jakož i osiřelých dat.
- Jasný seznam licencí doporučených EOSC a jejich kompatibilita s licencemi doporučenými členskými státy.
- Sledování vývoje licencí v čase u datových sad.
- Harmonizovaná politika a pokyny pro řešení případů, kdy by zveřejněním mohlo dojít k ohrožení patentové přihlášky nebo obchodního tajemství.
- Soulad s GDPR pro osobní údaje.
- Další omezení přístupu k údajům a jejich používání uplatňovaná pouze v případech platných právních předpisů nebo z oprávněných důvodů.
- Harmonizované podmínky používání napříč úložišti.
- Sladění vnitrostátních právních předpisů členských států a EOSC.

### 7.1.3 Perzistentní identifikátory

Problematika perzistentních identifikátorů v rámci EOSC je zakotvena např. v dokumentech Hellström a kol. 2019, nebo European Commission 2020a a 2020b. Perzistentní identifikátory mohou identifikovat mnoho různých výzkumných subjektů, které mohou být digitální (např. dokumenty, data, software, služby – jinak známé jako digitální objekty – a sbírky z nich vytvořené), fyzické (např. lidé, nástroje, artefakty, vzorky) nebo konceptuální (např. organizace, projekty, slovníky). PID podporující FAIR výzkum musí být celosvětově jedinečný, trvalý a rozlišitelný.

Podle britské studie založené na základě současné úrovně zavádění PID pro články (DOI) a osoby (ORCID ID) existuje významný přínos – včetně úspory nákladů – při vytvoření národního konsorcia PID, který se odhaduje na 5,67 milionu liber v průběhu pěti let, pokud se podaří splnit cíle zavádění PID ve výši 67 % do třetího roku a 85 % do pátého roku. (Brown et al., 2021, s. 5–7)

### 7.1.4 Data management plány

**DMP (Data Management Plan)** je formální dokument, který poskytuje popis nakládání s datovým materiálem v průběhu výzkumného projektu a po jeho skončení. (CESSDA 2020) Způsob, jakým bude

---

<sup>21</sup> Do popředí se zde dostává identifikátor ROR (Research Organization Registry).

DMP po dokončení vypadat, není univerzální. Je to „živý“ dokument, který průběžně reflektuje vývoj projektu a jeho účastníků. V průběhu projektu se aktualizuje, aby bylo zajištěno, že odráží aktuální stav projektu.

Jedná se o **účinný způsob, jak si výzkumník a jeho tým mohou ujasnit potřeby projektu na práci s daty ještě před zahájením projektu samotného**. V průběhu projektu je to pak referenční materiál, jak by se mělo s daty zacházet, a tedy udržovat v datech pořádek a kontrolu nad všemi procesy práce s nimi. Bez ohledu na velikost týmu je potřeba snadno najít informace týkající se dat, pokud jde o umístění souborů, konvence pojmenování, standardy, popis projektu, projektové role, zálohovací režimy, verzování atd. Sepsáním DMP může výzkumný pracovník zajistit, aby byl materiál po dobu výzkumu dobře spravován, což také usnadní přípravu materiálu k archivaci a umožní tak další výzkum po skončení výzkumného projektu. Obvykle je také snazší dokumentovat výzkumný materiál již v průběhu výzkumu, v těsné blízkosti kroků výzkumného procesu, v jejich rámci jsou výzkumná data vytvářena nebo měněna.

DMP umožňuje předem naplánovat, jak poskytnout datovou sadu do datového úložiště způsobem, který bude **co nejvíce odpovídat principům FAIR**. DMP zjednodušuje proces strukturování a dokumentace datových sad, čímž usnadňuje interním i externím uživatelům v budoucnu nalezení a pochopení materiálu, kodifikuje formát dat, který je nejvhodnější pro opakované použití, a specifikuje licenci pro využití dat dalšími subjekty.

Jedním ze způsobů, jak zajistit, aby se data nestala z dlouhodobého hlediska nepoužitelnými, je vybrat si **(důvěryhodné) datové úložiště**, které má podporu naplňování FAIR principů implementovanou ve své infrastruktuře pro předkládání datových sad. V zájmu dodržení principu FAIR je, aby výzkumní pracovníci ukládali svá data spolu s veškerou dokumentací potřebnou pro jejich pochopení a opětovné použití do (důvěryhodného) repozitáře výzkumných dat, který si stanoví jako výslovný cíl uchování dat a který disponuje potřebnými odbornými znalostmi pro udržitelné ukládání dat a zachování jejich použitelnosti. (Van Berchum & Grootveld, 2017)

## 7.2 Rozhovory s národními experty

Skupina uživatelů:

**prof. RNDr. Jan Hajič, Dr.**

(UFAL MFF UK), zástupce infrastruktury LINDAT/CLARIAH-CZ a vedoucí oborově-vědní pracovní skupiny Humanities and the Arts

**Co od EOOSC očekáváte? Proč je dobré EOOSC v ČR rozvíjet? Jaký je váš současný model práce s vědeckými daty? Na jaké problémy narážíte? Jak rozvoj EOOSC změní vaši praxi?**

Aktivita v našem oboru zpracování přirozeného jazyka je v souladu se současnými pravidly EOOSC již dávno. Všude to tak ale není. V některých oborech si výzkumníci drží data pro sebe a mají pocit, že je musí zužitkovat v nějaké publikaci, aby se jim práce vyplatila, protože za samotná data nic nedostanou. Překážky pro sdílení některých dat jsou i formální, např. v podobě autorských práv. Tyto věci je potřeba řešit i centrálně na národní úrovni a od EOOSCU očekávám, že se to bude dít.

Neočekávám, že se někdo postará o naše data, protože o ty se staráme sami a umíme to. Naopak dokážeme přinést EOOSCU spoustu technické expertizy a znalostí podložených našimi zkušenostmi.



### **Jak vnímáte současný stav implementace v ČR? Jaká jsou specifika v ČR oproti jiným státům?**

Mohu mluvit především o humanitních oborech. Tam je situace dramaticky rozdílná v jednotlivých zemích. V rozvoji ekosystému vlastních infrastruktur v tomto oboru má Česko náskok. Špatná situace je např. na Slovensku nebo ve Francii. Naopak dobrá podpora je např. ve Švédsku nebo Finsku.

V implementaci je potřeba rozlišovat mezi EOOSC a otevřenou vědou jako takovou. EOOSC je nová věc, kde bylo zatím Česko více viditelné ve třech evropských projektech (ELIXIR v biologii, EIRENE v environmentálních vědách a SSHOC v sociálních a humanitních vědách). V jiných oblastech je to složitější, nicméně spousta možností, jak uchovávat a zpřístupňovat data, u nás již v rámci otevřené vědy existuje. Jenom to zatím probíhalo trochu neorganizovaně, takže data třeba po několika letech zmizela, nebyla popsána, nebyla jasná licence apod. To jsou aspekty, které by měl EOOSC správně nastavit a posunout tak celou implementaci dopředu.

### **Jakým směrem by se měl rozvoj EOOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOOSC?**

Je otázka, kdo a proč potřebuje proniknout do evropských struktur EOOSC. Potřebujeme tam mít naše zástupce, ale nemusí tam být zastoupen každý vědec osobně. Přínosem by měla být určitá koordinace standardů služeb a přebírání dobrých praxí. Rozhodně ale nepotřebujeme nějaký centrální evropský superrepozitář.

Z projektu SSHOC máme zkušenost, že nedává smysl naházet celou vědu do jednoho pytle. Už jen sjednocení metadat napříč různými obory nedává smysl, protože i metadata jsou v různých oborech logicky různá. Takže udělat takový sjednocený centrální katalog představuje obrovskou práci, a přitom se při tak široké centralizaci vytratí potřebný specifický obsah. Např. pro jazyková data potřebujeme jiné údaje než pro chemická data. Současně chemik bude mít málokdy zájem propojovat chemická data s jazykovými. Smysl to tedy dává jen nad do jisté míry podobnými daty v rámci určitých příbuzných oborů, např. společenských a humanitních věd. Dobrá a užitečná integrace dat se tedy dá udělat spíše po oborech.

### **Jaké aktivity jsou pro rozvoj EOOSC v ČR klíčové?**

Z hlediska hardwarového zázemí by měl být český EOOSC takovou záchrannou sítí pro vědce a obory, které zatím o data postaráno nemají, ale i pro zavedené klastry, které mohou dočasně narazit na své kapacitní nebo technické možnosti. Lidské zdroje si musíme vychovat především sami v tematických klastrech, i když podpora na centrální úrovni se bude samozřejmě hodit.

Důležitá je koordinace aktivit z hlediska zákonů (např. autorské právo) a dobrých praxí. Rozvíjet je potřeba i služby národní datové infrastruktury, i když třeba federace identit v Česku je na velmi dobré úrovni. Potřebná je také koordinace směrem do Evropy. Na národní úrovni je potřeba, aby integrace nepřinášela zbytečnou práci, ale jasně definovala rozhraní a standardy – v tom je teď podle mě Česko na dobré cestě.

### **Jak hodláte dál rozvíjet své systémy v kontextu EOOSC?**

Opět je třeba rozlišovat mezi EOOSC a otevřenou vědou jako takovou. Rozvoj v oblasti otevřené vědy máme naplánovaný s partnery v rámci LINDAT/CLARIAH-CZ na sedm let v objemu cca 33 mil. Kč na hardware (z toho hodně do knihoven), ale naším hlavním úkolem bude investice do lidí, aby jednotliví partneři měli k dispozici pracovníky, kteří se orientují v otevřené vědě, datové vědě atd.

Z hlediska EOOSC se chystáme v rámci OP JAK podat projekt, který nám umožní adaptovat dnešní přístup platformy CLARIN ERIC (kterou jsme teď řízení na evropské úrovni) na přístup národního EOOSC.

Rozhodně chceme, aby naše data byla dostupná v národním prostředí. Důležitým úkolem národního EOSCu nicméně bude vyřešit problém heterogenity metadat z různých oborů, jak již byl řečeno výše. Část problému bude technické napojení a druhá část bude motivace těch, kteří ještě nejsou ochotni svá data sdílet – tu by měl OP JAK podpořit nabídkou nějaké formy financování pro zapojené organizace na oplátku za poskytnutí jejich dat.

**Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOSC)?**

V případech, kdy je infrastruktura schopná službu poskytnout, ale je to něco, co je poskytováno mimo jádro projektu, tak by bylo správné, aby zákazník za službu zaplatil. Nicméně pokud služby běží na webu a nevyžadují žádné mimořádné provozní náklady, není důvod vybírat drobné poplatky, které nakonec ani nezaplatí administrativu s jejich výběrem spojenou. Navíc by to vedlo přesně k opačnému výsledku, než jaký je sdílením dat zamýšlen - vědci by se nezapojovali a data by nepřebírali ani neposkytovali. Pokud k nám do repozitáře dává někdo data, jsou to většinou výzkumné instituce placené z veřejných peněz, takže by se jednalo jen další nesmyslné přerozdělování veřejných peněz. V Evropě se za prostý přístup k datům a služby s tím spojené neplatí. Jestliže to tedy nepřekročí nějaké kritické objemy, nedává smysl vyžadovat za tyto služby platby ani u nás.

Obecně současný přístup vlády k financování vědy nepovažuji za dobrý. Brát za dané, že vláda v současné situaci nemůže přidat vědě, je tragické. Z hlediska rozpočtu jde o malé objemy a do budoucna to zapříčiní velké potíže. V posledních dvaceti letech se úloha dat ve výzkumu významně zvyšuje a řada oborů se bez přístupu k datům vůbec nedá dělat. Aby se věda posouvala, měly by proto růst infrastruktury relativně rychleji vůči růstu celé vědy jako takové.

**RNDr. Jiří Chudoba, Ph.D.**

(FZU AV ČR), vedoucí oborově-vědní pracovní skupiny Physics

**Co od EOSC očekáváte? Proč je dobré EOSC v ČR rozvíjet?**

Grant pro klastr Physics by měl přinést především zdroje pro lidi, kteří se budou starat o data z dlouhodobého hlediska. V rámci spolupráce ve velkých mezinárodních projektech už se to daří, ale jsou i jiné projekty, kde na to nejsou zdroje ani kapacity. Pak jsou tu i lokální skupiny, které data sice produkují, ale nemají je kde uchovávat. Po skončení grantu bohužel data někdy mizí. EOSC by měl přinést na národní úrovni hardwarové zdroje pro archivaci těchto dat a zároveň experty, kteří by tato data dokázali popsat a zdokumentovat pro další využití.

**Jak v současné době pracujete s vědeckými daty? Na jaké problémy narážíte? Jak vnímáte současný stav implementace v ČR?**

Klastr Physics pokrývá různorodou oblast, kde někdo má a někdo nemá kapacity se o data starat. V rámci klastru by proto měla vzniknout spolupráce ve sdílení datových specialistů pracujících pro jednotlivé projekty třeba i jen na částečný úvazek. Zároveň mám ale obavy, kde budeme lidi pro práci s daty hledat. Máme problém najít experty na IT pro provoz a podobné to bude s datovými kurátory, které bude navíc ještě potřeba vychovat.

Zatím se inspirujeme především v zahraničí, kde je pro nás výhodné využívat kapacity Evropské gridové infrastruktury (EGI) a LHC gridu (WLCG)

**Jakým směrem by se měl rozvoj EOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOSC? Jaké aktivity jsou pro rozvoj EOSC v ČR klíčové?**

Nejdůležitějším aspektem je rozvoj lidských zdrojů. EOOSC je z našeho pohledu hodně o ukládání dat typicky s využitím centrálních katalogů, kde musí data někdo popsat a zkontrolovat jejich konzistenci. Pro vědce je to velká zátěž a nechce se jim do toho, resp. zatím často nemají důvod ani motivaci.

Zajímavý pro uživatele bude i přístup k hardwaru v rámci centrálního repozitáře, kde jim může dávat smysl mít uloženou „záložní“ kopii dat ze svých vlastních serverů.

**Jak hodláte dál rozvíjet své systémy v kontextu EOOSC? Do čeho, v jakém časovém horizontu a kolik hodláte investovat?**

Chceme podat projekt klastru Physics, do kterého budou zapojené velké infrastruktury i menší projekty. Řešení z velkých infrastruktur by nově najatí pracovníci měli přenášet k menším partnerům, pokud to bude vhodné (např. z hlediska velikosti dat). Náš odhad nákladů projektu na šest let odpovídá nákladům na 15 až 18 lidí a počítá s minimem hardwaru. Projekt prakticky neplánuje hardware pro ukládání větších objemů dat. Počítáme s využitím otevřeného software bez nákladů na licence.

**Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOOSC)?**

Ted' je třeba nabídnout a zaplatit práci datových expertů. Kontaktem výzkumníků s nimi se know-how rozšíří a výzkumníci už si pak sami dokáží plánovat kapacitu lidských zdrojů na další projekty. Spoluúčast výzkumníků bude spočívat v tom, že věnují vlastní čas na komunikaci s datovými experty.

**prof. RNDr. Jana Klánová, Ph.D.**

(RECETOX MUNI), zástupkyně infrastruktur RECETOX a EIRENE a vedoucí oborově-vědní pracovní skupiny Environmental Sciences

**Co od EOOSC očekáváte? Proč je dobré EOOSC v ČR rozvíjet?**

Očekávám větší dostupnost dat pro vědce i ostatní uživatele. Zároveň doufám, že to nevyvolá negativní jevy, jaké způsobil například tlak na publikování vědeckých výstupů v režimu Open Access: časopisy se začaly chovat nemorálně, navyšovat publikační poplatky, zavedly praxi hybridních časopisů, vznikla spousta periodik s velmi nízkou kvalitou a trh se stal nepřehledným.

Myslím tím ale skutečnou dostupnost dat. Nestačí, že je dostupný vědecký článek. Je třeba, aby byla k dispozici primární data, ze kterých vycházel. Optimálně na datovém úložišti, do kterého se data ukládají řízeně a obsahují kvalitní metadatový popis. Jedině tak je mohou efektivně využívat další uživatelé, aniž by trávili čas pátráním po zdrojích dat, komunikací s jejich původci a pokusy o jejich porozumění, vyčištění a harmonizaci.

Není přitom nutné vždy vytvářet nová centrální úložiště, na která budou všichni vědci své datové soubory odevzdávat. Proti tomu byl ve vědecké komunitě vždy odpor, protože původce dat tím nad nimi ztrácí kontrolu a možnost dalších aktualizací. Není to možné kapacitně, a u některých typů citlivých dat nejsou takové přesuny a nekontrolované manipulace ani myslitelné. Je třeba primárně pracovat se stávajícími úložišti – pokud existují – a podpořit jejich dohledatelnost, přístupnost, transparentnost a lepší využitelnost. Dnes už existují nástroje umožňující federované analýzy dat různých poskytovatelů, aniž by bylo třeba data přesunovat z původních repozitářů. Znamená to ale nemalou investici do harmonizace stávajících dat.

**Jaký je váš současný model práce s vědeckými daty? Na jaké problémy narážíte? Jak rozvoj EOOSC změní vaši praxi?**

Myslím, že v centru RECETOX máme určitý náskok, takže nás EOOSC nepřekvapil. Přes třicet let sbíráme data z dlouhodobých programů monitorujících znečištění prostředí a hodnotíme v nich dlouhodobé trendy. Proto jsme už před 15 lety vytvořili nejprve interní databázový systém GENASIS, z něž se později vyvinul národní informační systém pro sledování chemických látek v životním prostředí (viz <https://www.genasis.cz/>). Do něj se řízeně ukládají data ze všech relevantních environmentálních projektů tak, aby byla volně dostupná vědcům i dalším uživatelům.

O několik let později (v roce 2009) jsme se zapojili do Globálního monitorovacího plánu organizovaného pro zajištění hodnocení účinnosti Stockholmské úmluvy na ochranu prostředí a lidského zdraví před účinky chemických látek. Vznikla přitom obdobná potřeba informačního systému, který by umožnil celosvětová data sbírat, kontrolovat, harmonizovat, analyzovat a prezentovat. Vytvořili jsme za tím účelem klon našeho systému GENASIS (viz <https://www.pops-gmp.org/>), který je dosud jediným systémem, ve kterém se potkávají veškerá dostupná světová data o kontaminaci ovzduší a mateřského mléka. Nejedná se jen o datové úložiště, ale systém má i zobrazovací funkce. I relativně laický uživatel se tak snadno dostane k potřebným informacím.

Když jsme začali pracovat i s daty z dlouhodobých populačních studií (viz <https://www.celspac.cz/>), vytvořili jsme několik dalších interních systémů (administrativní systém pro registraci účastníků, dotazníkový systém, laboratorní systém na řízení vzorků nebo systém řízení biobanky). Největší výzvou je, aby byly všechny tyto databáze propojitelné. Abychom dokázali odpovědět na otázku, jak může zdraví konkrétního člověka s jeho genetickými vlohami ovlivňovat jeho životní styl, znečištěné prostředí, výživa, kouření nebo třeba pracovní stres. Podobné potřeby mají i jiná pracoviště, proto máme zájem rozvíjet nová řešení v rámci EOOSC a podělit se o své zkušenosti.

### **Jak vnímáte současný stav implementace v ČR? Jaká jsou specifika v ČR oproti jiným státům?**

ČR je slušně připravená díky rozvinutým výzkumným infrastrukturám, které jsou často zapojené do mezinárodních sítí, a otázku managementu dat a budování úložišť řeší dlouhodobě. Výhodou je také integrace relevantních podpůrných infrastruktur do jedné (e-INFRA) – tento proces proběhl už před několika lety. I výzkumná pracoviště, která se účastní mezinárodních projektů, mají zkušenosti se sdílením dat s mezinárodními konsorciemi. Je tu zkrátka už síť lidí, kteří mají jasnou představu, jak je třeba postupovat.

### **Jakým směrem by se měl rozvoj EOOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOOSC?**

Je třeba zamezit bezbřehé debatě a postupovat strukturovaně: definovat potřeby různých subjektů, vyjít z jejich reálných zkušeností a vytvořit základní rámec, který je možné dále rozvíjet. Nesmí se přitom zapomenout ani na data veřejné správy (na národní i evropské úrovni), která by měla být spravována na základě stejných principů. I data z veřejné správy jsou využitelná pro výzkum, stejně jako jsou výzkumná data potřebná pro podporu rozhodování. Je nutné odlišit potřebu prostého ukládání dokumentů (publikací nebo závěrečných prací) od skutečného zpřístupnění primárních dat.

### **Jaké aktivity jsou pro rozvoj EOOSC v ČR klíčové?**

Skoro vždycky skončíme u lidských zdrojů. Musíme vychovat generaci vědců, která bude správně dat věnovat potřebnou pozornost. Nejde jen o odborníky na hardware a software, ale především o uživatele samotné. V jednotlivých oborech musíme najít vědce, kteří rozumějí podstatě dat, mají s jejich analýzou a správou praktické zkušenosti a jsou ochotni se ve své kariéře soustředit víc na jejich zpracování než na generování nových. Významná část prostředků v EOOSC projektech by měla být

použita na získání a vyškolení datových specialistů pro jednotlivé oborově-vědní klastry. Měli by působit ve výzkumných infrastrukturách i velkých vědeckých centrech a umět hovořit stejným jazykem.

**Jak hodláte dál rozvíjet své systémy v kontextu EOOSC? Do čeho, v jakém časovém horizontu a kolik hodláte investovat?**

Chceme především propojovat dění na národní a mezinárodní úrovni. Centrum RECETOX je zapojeno v několika evropských projektech zaměřených na výzkum lidského exposomu (faktorů působících na zdraví člověka). European Human Exposome Network, například, je klastr devíti projektů zkoumajících nejrůznější zdravotní dopady environmentálních expozic. Sbírají data chemická a fyzikální, zdravotnická, ale i sociální a ekonomická a průřezové pracovní skupiny usilují o jejich harmonizaci a umožnění federovaných analýz. To nám umožňuje zapojit i naše data. Dalším důležitým projektem je PARC (European Partnership for the Assessment of Risks from Chemicals), nové evropské partnerství pro analýzu chemických rizik, financované od letošního roku z Horizon Europe. V obou projektech jsme spoluzodpovědní za rozvoj datových a informačních infrastruktur a vytvoření konceptu, který zajistí jejich dlouhodobou udržitelnost. Zároveň je ale RECETOX také koordinátorem nové evropské ESFRI infrastruktury pro studium lidského exposomu (EIRENE). Jelikož i pro EIRENE je (vedle budování laboratorních kapacit) integrativní analýza environmentálních, zdravotních a socioekonomických dat zcela zásadní, snažíme se maximálně využívat synergií mezi všemi projekty a vybudovat v Brně jeden z významných informačních uzlů exposomového výzkumu.

Klaster Environmental Sciences by pak měl v ČR v maximální možné míře kopírovat to, co se děje na evropské úrovni – i proto jsme se postavili do čela přípravné pracovní skupiny a usilujeme o co nejužší propojení s ostatními environmentálně zaměřenými infrastrukturami.

**Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOOSC)?**

Má-li se koncept Otevřené vědy (Open Science) skutečně implementovat, bude třeba dobré komunikace a spolupráce mezi e-INFRA a ostatními infrastrukturami a zásadních investic na obou stranách. U e-infrastruktur se jedná o investice do vytvoření architektury systému a budování potřebných kapacit, u ostatních infrastruktur pak do rozvoje oborově specifické expertízy. Takto je třeba také distribuovat veřejnou podporu na realizaci EOOSC. Nejen e-INFRA, ale i všechny ostatní infrastruktury budou potřebovat podporu na vybudování potřebných kapacit pro práci s daty, protože datové služby se stanou součástí jejich nabídky. Jejich úhrada pak může probíhat podle podobných pravidel, jako kterákoli jiná služba. Výzkumné infrastruktury, jejichž krajina je v ČR bohatě rozvinutá, mají potenciál se stát garanty implementace EOOSC a významně přispět k zapojení všech výzkumných institucí do evropské vědy.

**doc. RNDr. Jiří Vondrášek, CSc.**

(ÚOCHB AV ČR), zástupce infrastruktury ELIXIR CZ a vedoucí oborově-vědní pracovní skupiny Bio/Health/Food

**Co od EOOSC očekáváte? Proč je dobré EOOSC v ČR rozvíjet?**

Jde o novou kulturu práce s daty a vědeckými poznatky. EOOSC znamená, že se do vědeckých dat a práce s nimi začíná vnášet určitý systém. Je to i signál pro národní poskytovatele, že se data vytvořená z veřejných zdrojů budou moci archivovat a opakovaně používat a stanou se součástí většího celku v rámci principů otevřené vědy.

Mělo by to v neposlední řadě přinést i možnost dohledání vědeckých dat v případě podezření z manipulace nebo potřeby ověření publikovaných výsledků.

**Jaký je váš současný model práce s vědeckými daty? Na jaké problémy narážíte? Jak rozvoj EOOSC změni vaši praxi?**

Z hlediska konkrétní disciplíny, tj. biologie, jsou základy sdílení dat dávno nastaveny. Implementace u nás využívá zkušeností ze zahraniční spolupráce v různých typech infrastruktur.

Ted' je ale důležité vytvořit systém, který bude schopný propojit ať už na úrovni metadat nebo na sémantické úrovni různé typy oborů. To není triviální. Metadatové systémy jsou pro různé disciplíny rozdílné a metadatový adresář EOOSC by měl obsluhovat metadata napříč disciplínami.

**Jak vnímáte současný stav implementace v ČR? Jaká jsou specifika v ČR oproti jiným státům?**

Mohu mluvit za klastr Bio/Health/Food. Tam při návrhu implementace EOOSC v ČR sledujeme trendy evropských infrastruktur, jejichž nositelem je zejména zastřešující infrastruktura ELIXIR. V současné fázi jde především o informovanost potenciálních partnerů a jde o to, jestli se identifikují jako uživatelé nebo i jako tvůrci systému. Potřebujeme v rámci pracovní skupiny vytvořit síť institucí, které mají zkušenosti, a specifikovat naše nároky na repozitářovou platformu z hlediska jak investičního, tak z hlediska tzv. měkkých financí, určených jednak na vytvoření pozic pro specializované a vysoce kvalifikované pracovníky zabývající se správou dat, jednak na vzdělávání a informování širší vědecké obce.

**Jakým směrem by se měl rozvoj EOOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOOSC? Jaké aktivity jsou pro rozvoj EOOSC v ČR klíčové?**

Začal bych vzděláním. Nejtěžší část práce bude spočívat v osvětě a odborných školeních a seminářích, které každého vědce seznámí s konceptem EOOSC tak, aby se s ním dokázal identifikovat buď na základě vlastních potřeb, nebo potřeb své instituce. Ani většina zájemců o zapojení do pracovní skupiny nemá v současné době velké povědomí o kultuře práce s daty, životním cyklu dat nebo plánech na správu dat. Bude třeba zajistit profesionály, které to institucím zprostředkují, ať už vlastní, nebo nějakým způsobem sdílené s jinými institucemi.

Až na dalším místě jsou technické záležitosti, tj. zajištění dostatečného technického zázemí pro ukládání dat (úložiště, výpočetní kapacity). To mohou instituce v rámci EOOSC realizovat spoluprací především s IT infrastrukturami, nebo si mohou vytvořit vlastní repozitáře.

**Jak hodláte dál rozvíjet své systémy v kontextu EOOSC? Do čeho, v jakém časovém horizontu a kolik hodláte investovat?**

Jednotlivé infrastruktury v rámci Bio/Health/Food klustru už mají nějak rozvinutý způsob práce s daty i kapacity, které jsou k tomu určené. Některé infrastruktury se mohou stát zastřešující platformou pro jiné instituce. Hardwarové nároky by měla pokrýt infrastruktura e-INFRA, ale určité nároky zejména na lidské zdroje budou kladeny i na jednotlivé instituce. Například ELIXIR hardwarové zdroje využívá ve spojení právě s e-INFRA, ale třeba infrastruktura EATRIS vyžaduje mít řešení tzv. na místě vzhledem k citlivým datům, která uchovává. Tam budou tedy náklady větší.

Ze 14 institucí v rámci klustru jich cca 6 až 8 má v úmyslu požadovat z operačního programu prostředky pro správu a management určitých unikátních zdrojů dat. Pokud se objeví specifické požadavky, budeme se vždy snažit zjistit, jestli už tu pro ně nějaká infrastruktura je a zda je možné ji implementovat bez nároků na další náklady. Většina prostředků by tedy měla být vyžita na institucionální rozhraní a na specialisty, kteří se budou schopni o datový zdroj starat.

Zvláštní kapitolu tvoří data, která publikována nebudou, nebo to budou tzv. negativní výsledky. I data, která nepovedou k publikaci, by měla být uchována a zpřístupněna. I na to by mělo být pamatováno v rozpočtech výzkumných projektů.

Odhadem by klastr Bio/Health/Food mohl podat projekt v rozmezí 100 až 200 mil. Kč, což je nekvalifikovaný odhad na základě dosud probíhajících rozhovorů.

**Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOOSC)?**

Existují dva typy infrastruktur. V některých disciplínách se uplatňuje přístup ke službám zdarma a předpokládám, že to tak zůstane. Jiné infrastruktury vládnou velkými technologickými celky (např. infrastruktura pro strukturní biologii), kde přístup k infrastruktuře a datům obnáší náklady z hlediska jejího provozu.

Bude tedy třeba využít nějaký systémový typ klasifikace. Nepřipadá asi v úvahu vybírat peníze za data vytvořená z veřejných zdrojů od uživatelů z jiných veřejných institucí. V projektech, kde má o data zájem privátní sektor, to bude vhodné řešit na úrovni licencí a politik příslušných infrastruktur. Spíše bych tedy oddělil komerční sféru od akademické a nepovažoval bych za správné, aby výzkumníci z veřejné sféry platili za data vytvořená z veřejných zdrojů.

Skupina expertů:

**RNDr. David Antoš, Ph.D.**

(CESNET), vedoucí pracovní skupiny Architektura NDI

**Co od EOOSC očekáváte? Proč je dobré EOOSC v ČR rozvíjet?**

Očekávám evoluci. EOOSC je další krok v integraci infrastruktur pro vědu a výzkum. Prošli jsme historicky etapami, kdy takové infrastruktury vznikaly živelně a podle potřeby, v akademickém prostředí se objevovali poskytovatelé různých IT služeb. Pak přišlo období postupné integrace, nejdřív po oblastech: založte si v každém státě jednoho poskytovatele počítačové sítě pro vědu a výzkum, jednoho poskytovatele výpočetního prostředí a podobně. Pak se ukázalo, že síla infrastruktur je právě v integraci jejích složek, takže dnes máme portfolio služeb národní e-infrastruktury, které jde od počítačové sítě přes výpočetní prostředí různých typů, datová úložiště, až po prostředí pro spolupráci, a uživatel má (ideálně) jednu registraci, jejímž prostřednictvím přistupuje k celé škále služeb. EOOSC k tomu přidává jako společné téma práci s vědeckými daty a reaguje tak na myšlenky otevřené vědy a hnutí FAIR dat. To není velké překvapení, že spousta vědecké práce je o datech, jen se o tom nyní začalo velmi explicitně mluvit.

Proč je dobré EOOSC v ČR rozvíjet – pokud chceme mít rozumné evropské financování, které nám pomůže udržovat konkurenceschopnost české vědy, tak nám nic jiného nezbývá. Nároky na vědeckou práci v oblasti dat velice vzrostly. Před 20 lety jste publikovali článek, ten měl grafy, tabulky, ale nikdo se neptal: "Jak jste ta data pořídili? Mohli bychom je vidět?" To už je dneska ve spoustě oborů pro alespoň trochu solidní publikaci nepředstavitelné. Některé komunity jsou na to připraveny docela dobře, pokud si příslušné mechanismy připravily samy. Ale to je spíše menšina.

Schopnosti IT zpracovávat čím dál větší data budou mít čím dál větší vliv i na základní principy vědecké komunikace. Například ve vědách o živé přírodě se mluví o tom, že lineární vyprávění o nějaké znalosti formou publikace se začíná jevit jako zbytečnost, že znalosti jsou rozumně zachycené jen v datech. A to pak vyžaduje velmi komplexní podporu pro jejich zpracování.

### **Jak vnímáte současný stav implementace v ČR? Jaká jsou specifika v ČR oproti jiným státům?**

Jsem přiměřeně optimistický, protože máme docela silnou a vysoce integrovanou e-infrastrukturu, takže z technologického hlediska je budování infrastruktury "pro EOOSC" evolučním krokem v tom, co už dlouhá léta děláme. Ta náročná část pro nás je hledání široké multioborové shody, jak má taková infrastruktura vypadat a co přesně má dělat, aby pak výsledek byl použitelný pro velmi heterogenní obory. Máme sice dlouhodobě uživatelské komunity z různých oborů, nicméně současné pracovní skupiny se zastoupením od sociologů po fyziky, to je výzva. Řekl bych ale, že k rozumné sdílení představě postupně směřujeme.

Specifika v implementaci EOOSC jsou tématem spíše na desítky stránek než na jednoduchou odpověď, detailní informace najdete například ve zprávách projektu EOOSC SYNERGY. Obecně lze říct, že každá země je napřed v něčem jiném, někdo má vládou schválenou politiku, někdo ji teprve připravuje, někdo má existující centrum pro podporu některých vědeckých komunit, někdo ještě ne. Zásadní rozdíly jsou ve financování vědy a výzkumu, kde panuje jednota zejména v tom, že žádný ze států sledovaných v EOOSC SYNERGY nedosahuje plánovaných tří procent HDP. To jsou ale otázky na úrovni vládních politik. Na opačné straně spektra máme nějaké de facto komunitou uznávané technické standardy, kterých je rozumné se držet. Cokoli mezi tím je na nás. Když mluvíme s kolegy ze zahraničí, zjišťujeme, že řešíme velice podobné problémy. To je určitě dobrá známka.

### **Jakým směrem by se měl rozvoj EOOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOOSC?**

Je třeba dotáhnout plány a představy do podoby funkčního systému. Co se integrace týká, je hlavně potřeba opustit myšlenku, že dostaneme odněkud z Bruselu sadu standardů a seznam položek, které máme udělat. EOOSC není nějaký nasaditelný produkt, ale je to zejména základní myšlenka o práci s vědeckými daty a vedle ní způsob financování. Většina otázek je tam spíše otevřená, za běhu se vytvářejí nejen odpovědi, ale i otázky samotné. Pokud na to někdo chce mít vliv, má možnost se zapojit do řady pracovních skupin, které formují představu, jaký vlastně EOOSC bude a jaké bude mít požadavky. Zásadní je proto být s těmito skupinami v kontaktu, a tam si myslím, že již velmi dobré pokrytí komunity máme.

### **Jaké technologie navrhujete využít? Jaké nové služby implementace EOOSC přinese? Jaké požadavky budou kladeny na uživatele?**

Pro začátek je potřeba si říci, že technologická realizace EOOSC není a nemůže být o radikálně inovativních technologiích. Samozřejmě by bylo skvělé, kdyby se objevila nějaká úžasná technologie pro levné a spolehlivé ukládání dat, ale na to nelze spoléhat, když potřebujete stavět infrastrukturu. Z technologického hlediska máme datová úložiště, výpočetní prostředí různých typů, počítačovou síť, nástroje pro komunikaci, to jsou stále komponenty, se kterými počítáme a do kterých pak můžeme zasadit prostředí pro ukládání anotovaných dat, což je hlavní téma EOOSC. A k tomu potřebujeme komponenty, které už jsou také v docela pokročilém stupni rozvoje: vysokokapacitní datová úložiště, repozitáře pro agregaci metadat a pro vyhledávání, vlastní repozitářovou platformu pro ukládání dat. Dál musíme mít pomocné systémy, jako vydávání perzistentních identifikátorů pro data, nebo systém pro archivaci uložených dat formou zálohování. Není zvlášť těžké cokoli z toho uvést do provozu v malém rozsahu. Zajímavé to začne být, když očekáváte, že budete ukládat petabajty až desítky petabajtů. A velmi zajímavé, když ty komponenty potřebujete upravit a integrovat do jedné konzistentní infrastruktury; za tím je vždycky obrovská spousta práce.

Nové služby, ty hlavní a viditelné: národní repozitářová platforma bude systémem pro vytváření datových repozitářů. Jednak na ní budeme provozovat obecný repozitář pro pokrytí long tail vědy,



jednak a zejména to ale bude systém, na kterém si budete moci vytvořit vlastní repozitář s podporou specifických metadat a ve specifickém designu. I u repozitářové platformy předpokládáme, že bude heterogenní a distribuovaná, že to není jeden obrovský systém. V ní pak budou vznikat nové repozitáře odborných komunit a institucí, což pro ně bude mít zásadní výhodu v tom, že se o naprostou většinu toho systému nebudou muset starat.

Požadavky na uživatele v oblasti zacházení s daty typicky přicházejí zvenčí, od grantových agentur a donátorů, na základě postupně se vyvíjejícího mínění odborné veřejnosti, jak by se správa dat měla dělat. Návrh implementace EOOSCu se snaží toto vědcům maximálně usnadnit. Zcela zásadní aspekt celé věci jsou proto lidé. Nezbytně potřebujeme vychovávat datové specialisty, kteří budou umět vědeckým komunitám se správou dat efektivně pomáhat. To zahrnuje role, o kterých se mluví pod různými názvy, od datových knihovníků, data stewards, až po datové vědce, kteří rozumí odborné sémantice dat natolik, že jsou schopni dělat vlastní analýzy. Součástí navrhovaných projektů jsou také masivní vzdělávací programy právě v této oblasti. Pokud jde o požadavky na obecnou IT gramotnost uživatele, ideálně by měly být menší, než jaká je pro používání služeb infrastruktury potřeba dnes.

### **Jaké aktivity jsou pro rozvoj EOOSC v ČR klíčové? V jakém časovém horizontu by měly být realizovány?**

Jsem "od infrastruktury", tak bych měl říct hardware a software, ale to by byl naprostý nesmysl. Lepší odpověď je "všechny", protože bez kterékoli z vyjmenovaných celý záměr bude mít velké problémy. A v jakém časovém horizontu: idealistická odpověď by byla "včera bylo pozdě". Realistická je, že se bavíme o vybudování infrastruktury za netriviální prostředky, k jejímuž vzniku a provozu potřebujeme specialisty, kterých je nedostatek (ostatně jako kdekoli jinde v IT), a také pokrytí vědecké komunity podpůrným personálem, který se musíme teprve naučit vychovávat. To je plán na pět až sedm let.

### **Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOOSC)?**

Požadavky na zacházení s vědeckými daty vycházejí z evropské politiky v téhle oblasti, takže by mělo být úkolem EU a členských států, aby byla k dispozici infrastruktura, která je umožňuje naplňovat. Máme dlouhodobou zkušenost, že kdekoli došlo na spoluúčast uživatelů, tak to spíše používání infrastruktury poškodilo, než podpořilo. Navíc jako neziskový poskytovatel infrastruktury mohou mít hospodářskou činnost nejvýše do 20 %, to jsme v pásmu, kdy takové příjmy padnou z valné části na administrativu s nimi spojenou. Samozřejmě data dokážou růst nade všechny meze a jejich skladování rozhodně zadarmo není, takže nějaký model udržitelnosti bude určitě potřeba vymyslet, nicméně z mého hlediska tím modelem rozhodně není spoluúčast uživatelů. Spíše potřebujeme kvalifikované datové kurátory, kteří budou schopni pro každý typ dat rozhodovat o tom, jak dlouho je má smysl uchovávat.

**PhDr. Radka Římanová, Ph.D.**

(Ústřední knihovna UK), vedoucí pracovní skupiny Vzdělávání a lidské zdroje

### **Co od EOOSC očekáváte? Proč je dobré EOOSC v ČR rozvíjet?**

Od zavedení politiky EOOSC v ČR očekávám konsolidaci a standardizaci v oblasti nakládání s vědeckými daty a jejich zpřístupňováním. UK doufá, že dojde ke konsolidaci směrem k mezinárodním standardům a zvyklostem. Univerzity, které jsou nám v Evropě vzorem, jsou v oblasti datové politiky o hodně dál. Intervence by měla pomoci, ale náskok se bude těžko dohánět. Nakládání s vědeckými daty na západoevropské univerzitě zajišťují velká podpůrná centra, která vědcům pomáhají od obecných návodů a školení až po řešení konkrétních problémů. V ČR taková centra nemáme.

### **Jak vnímáte současný stav implementace v ČR? Jaká jsou specifika v ČR oproti jiným státům?**

Oproti jiným státům jsme relativně malá komunita. Specifikem je nepřipravenost organizací. Intervence MŠMT pro zavedení data management plánů byla zahájena zhruba před šesti lety, ale nebyla zaměřena na zmiňovaná datová centra. Výsledkem je, že na vysokých školách stále chybí zpracované datové politiky. Velký rozdíl je v připravenosti mezi univerzitami zejména v Praze, Brně a Ostravě a jinými univerzitami, kde nejsou k dispozici lidské zdroje, a tato oblast se zatím stále řeší spíše na bázi ad hoc případů náhlé potřeby než komplexně. Aktuální stav politiky implementace je výrazně ovlivněn příslibem možnosti čerpat prostředky z intervencí OP JAK a hledání cesty mezi vybudováním centrální infrastruktury a budováním infrastruktury na institucích.

UK by přivítala užší komunikaci mezi vedením univerzity a přípravným výborem EOSC. Z hlediska implementace EOSC nám chybí jasné strategické stanovisko k tomu, co lze a nelze v rámci intervence uplatnit a jakým směrem má instituce svou činnost napřít, aby nedocházelo k duplicitnímu financování a nezavírala se možnost implementace. Na UK působí dobře rozvinuté infrastruktury LINDAT/CLARIAH a ELIXIR, ale jiné obory zatím tápou a chybí jim jednoznačná „kuchařka“, jak věci řešit a posouvat dopředu.

### **Jakým směrem by se měl rozvoj EOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOSC? Jaké aktivity jsou pro rozvoj EOSC v ČR klíčové?**

Je třeba podpořit nejen národní infrastrukturu a oborové klustry, ale i jednotlivé instituce. Na počátku diskuse byl významně akcentován oborový prvek, ale ne všechny oblasti jsou takto jednoznačně klasifikovatelné. Rozuměli bychom modelu intervence, kde by byla jedna aktivita pro lídry v jednotlivých klastrech, druhá aktivita pro zapojení do těchto klastů a třetí aktivita pro vybudování institucionálních repozitářů tam, kde to bude smysluplné. Centrální národní řešení nemůže roli konkrétních instituce plnohodnotně nahradit, protože se jedná o individuální dlouhodobou práci s jednotlivými vědci v rámci těchto institucí. Dalším důvodem jsou citlivá data např. z medicínského výzkumu ve fakultních nemocnicích, která by neměla opustit institucionální repozitář.

### **Jaké nové požadavky budou kladeny na uživatele? Jak si představujete rozvoj v oblasti lidských zdrojů?**

Viděno pohledem univerzity: Poskytnutí hardwaru lze zajistit prostřednictvím CESNETu. Nutné je definování datové politiky v instituci. Tam musí vzniknout nové pozice datových kurátorů apod., které umožní přechod od samoobslužného přístupu vědců k profesionálnímu přístupu k zacházení s vědeckými daty. Každá VŠ má svá specifika, která národní centrum nemůže dost dobře obsáhnout. Navíc je otázka, proč bychom se měli přizpůsobovat novému centrálnímu systému metadat, když už např. data sdílená v rámci mezinárodní spolupráce vyhovují jiným mezinárodně uznaným standardům. Moje zkušenost mne vede k tvrzení, že i přes to, že bude při konverzích metadatových formátů připuštěna ztrátová konverze, mohou se objevit i specifika, která nebude jednoduché zvládnout.

Důležité je pokrýt práci s daty již ve studijních programech a pokračovat celoživotním vzděláváním vědců. Národní standardy by se měly dobře promýšlet, aby byly kompatibilní s již osvědčeným mezinárodním prostředím. Školit lze na národní úrovni, ale pro koncové vědce už by školení mělo zahrnout i specifika konkrétní instituce (tím není myšleno odklonění od standardu, ale spíše konkrétní cesta, kde a jak data ukládat, jaké jsou přístupové údaje, institucionální organizace práce apod.) nebo oborové skupiny. Z hlediska studijních programů je trendem držet se spíše na univerzálnější rovině. Systém vzdělávání je na evropské úrovni již dobře rozvinutý a lze z něj vycházet. Dobrým příkladem je VŠCHT, která v současné době testuje v rámci programu Horizont 2020 mezinárodní dvouletý studijní předmět pro doktorandy pro práci s vědeckými daty.

Finanční náročnost datové výchovy a vzdělávání lze pro UK odhadnout cca na 24 mil. Kč ročně (roční mzda cca 800 tis. Kč pro cca 30 datových specialistů). To pokrývá cca 50 000 studentů a 5 000 zaměstnanců, z nichž zhruba 2 000 se datová politika přímo dotýká.

**Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOOSC)?**

Argument povinnosti vkládat do jakéhokoli systému vlastní prostředky pomáhá obecně k zefektivnění každého systému. Na začátku je dobré poskytnout prostředky na rozjezd a pak se může zvážit samofinancování, pokud je vidět jasný přínos dané aktivity v konkrétní instituci. Zejména u malých institucí to ale může znamenat, že s koncem grantu skončí i jejich činnost v dané oblasti. Dobré by bylo, kdyby hardware dokázal udržet CESNET za přijatelných podmínek pro zapojené instituce. Takové podmínky by měly být známy ideálně už v době zapojení do projektu Národní repozitářové platformy U institucionálního repozitáře ponese náklady instituce. U projektově financovaných oborových skupin naproti tomu může vzniknout problém, pokud jim skončí projekt, ve kterém koordinují i jiné instituce v rámci oborového clusteru, kdo potom převezme tu koordinační roli? Z jakých prostředků se bude tato koordinační práce hradit?. Pokud to není jasné od samého začátku, tak dává větší smysl vytváření institucionálních repozitářů, za kterými si budou instituce stát. Koordinace by pak mohla být realizovaná v rámci dobrovolné spolupráce, i ta však něco stojí.

Skupina stakeholderů:

**doc. MUDr. Marian Hajdúch, Ph.D.**

(Ústav molekulární a translační medicíny LF UP v Olomouci), zástupce infrastruktury EATRIS-ERIC, člen Rady pro výzkum, vývoj a inovace

**Co od EOOSC očekáváte? Proč je dobré EOOSC v ČR rozvíjet?**

Očekávám lepší správu dat a zlepšení jejich využitelnosti pro výzkumné účely. Data mají velkou cenu, ale my je v ČR často dobře nevyužíváme, někdy z důvodů kapacitních nebo legislativních. Očekávám, že se tyto překážky podaří odstranit a dojde k vybudování důvěryhodných repozitářů a na ně napojeného prostoru pro práci s daty. Z mého pohledu jsou zvláště důležitá citlivá data, která souvisí např. se zdravotním stavem obyvatel nebo s jejich společenským zázemím. Bez práce s těmito daty se nedají řešit problémy obyvatel, ale jsou to data citlivá a lidé, kterých se týkají, musí mít elementární důvěru k institucím a způsobu zpracování těchto dat. I v oblasti necitlivých dat je pochopitelně potřeba se o data dobře postarat, zajistit jejich věrohodnost a zamezit jejich dezinterpretaci nebo jinému zneužití. Platí, že nelze ukládat všechna data napořád, bude tedy nutné stanovit nějaká oborově specifická pravidla, jaká data a na jak dlouho je v daném oboru potřeba schraňovat.

**Jaký je váš současný model práce s vědeckými daty? Na jaké problémy narážíte? Jak rozvoj EOOSC změní vaši praxi?**

Model práce s vědeckými daty závisí na konkrétním projektu. Často spolupracujeme na projektech, kde dostáváme již anonymizovaná data, nad nimiž nemáme přímou kontrolu. Současně také generujeme velké množství dat, které poskytujeme spolupracujícím institucím. Pro zpracování takovýchto dat jsme vyvinuli sadu nástrojů, které nabízíme komunitě v rámci infrastruktury EATRIS-ERIC pro kompletní zpracování dat v translačním medicínském výzkumu. I díky tomu jsme byli schopni poskytnout v průběhu pandemie kapacitu pro národní laboratorní informační systém, kdy se zpracovávaly statisíce až miliony laboratorních výsledků covidových testů.

Problémy, na které narážíme, se týkají především kapacit, jak z hlediska technického tak personálního. Obecně vidím závažný problém také v tom, že nejsme na EOSC plně připraveni z hlediska legislativy, ani z hlediska přípravy data management plánů a dohod o transferu dat (Data Transfer Agreement). Velkou bariérou je nemožnost data bezpečným způsobem propojovat mezi různými institucemi, které je generují. Nejsme proto např. ve zdravotnictví schopni hodnotit efektivitu konkrétní léčby, protože nemůžeme propojit data MZ a MPSV, takže nejsme schopni zjistit, kolik dnů nemoci tato konkrétní léčba ušetřila. To jsou přitom klíčové informace nejen pro vědu, ale i pro stát a jeho řízení.

### **Jak vnímáte současný stav implementace v ČR? Jaká jsou specifika v ČR oproti jiným státům?**

ČR je teď z hlediska komplexnosti plánu na implementaci EOSCu v mimořádně příznivé situaci v porovnání s dalšími evropskými zeměmi. Je to bezpochyby velkolepý plán, který představuje velkou šanci, a já doufám, že se úspěšně naplní. Současně si ale uvědomuji i velká rizika vzhledem k zatím malé připravenosti vědecké komunity.

Konkrétně v naší vědní oblasti existuje v ČR několik center, která již mají určitou míru kompetence při zpracování citlivých dat; tato centra je potřeba vhodně propojit a využít jejich zkušeností k rychlému naběhnutí systému. Obecně však připravenost vědecké komunity není velká a já vidím určitý problém v tom, že MŠMT již ve dnes připravovaných výzvách požaduje, aby se všechna výzkumná data ukládala do datových repozitářů a byla zpracována v souladu s principy FAIR. Osobně bych se přikláněl spíše k delšímu přechodnému období, protože na přípravu výzkumníků a implementaci technologií bude potřeba nějaká delší doba. A pokud nebude vše dobře připraveno, můžeme tím vědeckou komunitu od smysluplného nakládání s daty spíše odradit.

### **Jakým směrem by se měl rozvoj EOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOSC? Jaké aktivity jsou pro rozvoj EOSC v ČR klíčové?**

Účastním se řady evropských projektů se zaměřením na vědecká data a z těchto zkušeností je zcela jasné, že není možné správu dat centralizovat na jedno místo. Musí to být federalizovaná infrastruktura, a to na mezinárodní úrovni i uvnitř jednotlivých států.

Věc, kterou chci zdůraznit, je, že by měla vzniknout specifická aktivita kolem citlivých dat napříč různými obory, zaměřená na specifické technické požadavky z hlediska zabezpečení, specifické právní požadavky, a také specifické požadavky ze strany uživatelů, případně i státu, který citlivá data poskytuje. Je třeba chápat, že FAIR neznamená zcela volnou dostupnost všech dat a že citlivá data se otvírají až na základě Data Transfer Agreementu (DTA), což je právní dokument, který mimo jiné stanoví, jaká data lidé dostanou, jak s nimi mohou nakládat a kdy a jak je musí po provedení analýzy smazat. Z hlediska zpracování citlivých dat ve zdravotnictví je nyní významným evropským počinem příprava legislativního rámce pro European Health Data Space (EHDS)<sup>22</sup>, který bude určovat, jak se budou citlivá data do budoucna zpracovávat a ukládat pro účely primárního i sekundárního využití, tj. např. i pro potřeby výzkumu.

### **Jaké technologie navrhujete využít? Jaké nové služby implementace EOSC přinese? Jaké požadavky budou kladeny na uživatele?**

Primárně by to měly být technologie federalizované na jednotlivých institucích, které je potřeba dobře propojit. Je potřeba zajistit vznik metadatového adresáře, ze kterého se dozvíme, jaké typy dat jsou na konkrétních institucích k dispozici. Datové repozitáře na jednotlivých institucích je potřeba významně posílit. Z pohledu medicíny musí vzniknout síť kompetenčních center, která budou vybavena pro datovou fúzi, budou zabezpečena a budou mít technické i legislativní podmínky k této činnosti,

<sup>22</sup> [https://ec.europa.eu/health/publications/proposal-regulation-european-health-data-space\\_en](https://ec.europa.eu/health/publications/proposal-regulation-european-health-data-space_en)

včetně certifikace pro práci s citlivými daty. Myslím si, že je to velký úkol, který není realizovatelný do r. 2024.

**Jak hodláte dál rozvíjet své systémy v kontextu EOOSC? Do čeho, v jakém časovém horizontu a kolik hodláte investovat?**

Nedávno jsme si nechali zpracovat výhledovou investiční analýzu v rámci našeho ústavu a infrastruktury. V dalších letech potřebujeme investovat desítky mil. Kč do obnovy IT infrastruktury a jejího zabezpečení. V letošním roce provádíme řadu kroků, které souvisí se zajištěním bezpečnosti a získáním certifikace pro správu citlivých dat. Paralelně vyvíjíme další nástroje pro optimalizaci sběru klinických dat. V rámci jednoho experimentu jsme schopni vygenerovat i 20 TB dat, což je obrovský objem. Klíčovou podmínkou pro efektivní zapojení do mezinárodní výzkumné spolupráce, je pro nás rozvoj evropské infrastruktury, specificky navýšení přenosových rychlostí a kapacit pro přenos dat mezi datovými centry. Česká republika je na tom díky dobré práci CESNETu relativně dobře, ale narážíme na problémy s přenosem velkoobjemových dat z jiných států.

**Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOOSC)?**

To je legitimní otázka. Když uvážíme rychlost, jakou technologie zastarávají, lze předpokládat, že po skončení projektů EOOSC v rámci OP JAK bude potřeba kolem 500 až 750 mil. Kč ročně jenom na provoz, obnovu a údržbu infrastruktury. S takovými náklady je třeba počítat, ať už přijdou jakoukoli cestou ze státního rozpočtu. Ani v zahraničí není běžné, aby to bylo pro uživatele zcela zadarmo. Znamenalo by to nahromadění velkého množství datového odpadu, který blokuje místo a brání efektivnímu využití zdrojů. Určitě by se tedy do budoucna měly na financování v nějakém rozsahu z vlastních zdrojů podílet i uživatelé, což pomůže významně zvýšit kvalitu služeb a zajistit dlouhodobou udržitelnost datových repozitářů a služeb.

**RNDr. Jan Hrušák, CSc.**

(ÚFCH J. Heyrovského AV ČR), zástupce ČR v ESFRI

**Co od EOOSC očekáváte? Proč je dobré EOOSC v ČR rozvíjet?**

Očekávám, že EOOSC naplní deklarované cíle jak na evropské, tak i na národní úrovni, zároveň mám ale i určité pochybnosti. Myšlenka otevřené vědy a sdílení dat je jistě správná, ale vidím tady velký komunikační deficit. Zatím o implementaci diskutují zejména přesvědčení, ale směrem k potenciálním uživatelům diskuse mizí. Deklaruje se, že EOOSC má sloužit 2 mil. vědců v Evropě, ale valná většina z nich to zatím stále netuší. Až to zjistí, bude většina z nich v první chvíli proti, protože za tím uvidí více práce a žádné hmatatelné výhody. Zejména o tom se musí již nyní diskutovat.

Významnou výhodou (mimo těch obvykle deklarovaných) by mělo být, že EOOSC představuje bezpečný a transparentně regulovaný prostor, ve kterém je možné s daty operovat. Na rozdíl od komerčních služeb, které přinášejí nejistotu bezpečnosti a kontinuity.

**Jak vnímáte současný stav implementace v ČR?**

Myslím si, že implementace postupuje dobře v návaznosti na stávající výzkumné infrastruktury (EOOSC je vlastně v principu také výzkumná infrastruktura). Významné je, že jde o proces, který zaručuje systémové zapracování konceptu do širší politické dimenze.

S obecným vývojem jsem spokojen. Dobré je, že je zapojena širší základna potenciálních uživatelů. Nicméně je potřeba prohloubit komunikaci směrem k lidem, kteří o EOOSCu ještě nic neslyšeli. Hodně

řešíme centrální infrastruktury, ale málo vlastní úložiště, o kterých se předpokládá, že budou budovány v rámci a na náklady výzkumných a akademických institucí.

**Jakým směrem by se měl rozvoj EOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOSC? Jaké aktivity jsou pro rozvoj EOSC v ČR klíčové?**

Potřebujeme docílit kulturní změny – musíme začít vycházet z interních potřeb výzkumných institucí a vědeckých komunit. Pokud na to instituce nedostanou peníze, nebudou mít velký zájem se na EOSCu podílet. Také chybí dostatečná motivace poskytovat data – v rámci hodnocení VaVal to vědcům a institucím zatím nic nepřináší, na rozdíl např. od publikací. Toto je potřeba změnit. Samozřejmě je třeba zohlednit také specifické zvyklosti různých vědních oborů.

Rovněž je žádoucí podpořit vzdělávání a rozvoj lidských zdrojů v oblasti data managementu. Je třeba vymezit jasně definované postupy pro nakládání s daty, a toto know-how je třeba přenést k institucím a výzkumníkům. Každá instituce by měla mít vypracovaná schémata podpory na prosazení zmiňované kulturní změny, což jí mj. umožní i přístup k dalšímu financování z Evropy.

**Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOSC)?**

Jak už bylo řečeno, i toto je o kulturní a systémové změně. V ČR používáme strukturovanou metodiku hodnocení vědy a bylo by potřeba do ní zavést motivační parametr pro sdílení dat. Další motivací je přístup k budoucím finančním zdrojům navázaným na EOSC a sdílení dat, jako jsou prostředky rámcového programu Horizon Europe, ale i další. Je potřeba překonat počáteční kritickou dobu a postupně se projeví přínosy ze sdílení dat i pro disciplíny, které zatím stojí stranou tohoto dění. V principu platí, že většina vědy se dělá ve veřejném prostoru, takže případné sdílení nákladů nebo spolufinancování je jen otázka přesměrování veřejných zdrojů.

**prof. RNDr. Luděk Matyska, CSc.**  
(ISC MUNI), zástupce infrastruktury e-INFRA CZ

**Co od EOSC očekáváte? Proč je dobré EOSC v ČR rozvíjet?**

EOSC posune ČR a její výzkumnou komunitu směrem k implementaci principů Open Science v oblasti strukturované práce s výzkumnými daty. Přinese povědomí, že data jsou cenná a má smysl je udržovat, popisovat a sdílet. Na mezinárodní úrovni EOSC umožní zapojení naší vědecké komunity do mezinárodní spolupráce. Dobře popsaná a přístupná data představují kompetitivní výhodu pro tuto spolupráci.

Nechceme EOSC prezentovat jako něco, co přišlo zvenku, a čemu je třeba se přizpůsobit, ale snažíme se pochopit naše potřeby a ty promítnout do rozvoje EOSC i na mezinárodní úrovni. Evropa pak zpětně ocení náš aktivní přístup a zajímavá řešení. Pokud na národní úrovni nevznikne dostatečné obecné povědomí o důležitosti dat, určitě tu i bez ohledu na národní podporu vzniknou pozitivní ostrovy v podobě jednotlivých subjektů a oborových klastrů, ale celkově se ČR vzdálí výzkumnému prostředí v Evropě.

**Jak vnímáte současný stav implementace v ČR? Jaká jsou specifika v ČR oproti jiným státům?**

Některé české subjekty (např. členové e-INFRA CZ, dále MFF UK, FZU AV ČR) jsou v EOSC aktivitách na mezinárodní úrovni zapojeni dlouho. V ČR se o EOSC začalo více mluvit až na přelomu let 2020-2021 a celková implementace se zatím nerozvinula, s výjimkou jednotlivých dříve zmíněných subjektů. I ponětí o EOSC je zatím v ČR poměrně slabé. Prostřednictvím pracovních skupin pro rozvoj EOSC už se nicméně

podářilo aktivovat zajímavé množství lidí, kteří se chtějí podílet na systematické přípravě dalšího rozvoje. Výhodou ČR je, že na politické úrovni padlo rozhodnutí o způsobu financování rozvoje EOSC prostřednictvím OP JAK, což umožňuje plánovat další rozvoj.

Složení pracovních skupin pro rozvoj EOSC nicméně svědčí o tom, že povědomí o EOSC je více rozvinuté zatím pouze v Praze, Brně, Olomouci a Ostravě. Ostatní regiony jsou sice informovány, ale do práce se zatím nezapojují.

Je třeba si uvědomit, že rozvoj EOSC se týká především akademické sféry, protože průmysl má v oblasti dat svoje specifika a na místě je spíše určitá obezřetnost z obou stran. Rozvinutější spolupráce s aplikačním sektorem u nás existuje jen v oblasti zdravotní péče, která je na akademický výzkum úzce navázána. Infrastruktura e-INFRA CZ je nicméně v kontaktu s iniciativou Průmysl 4.0 a ČVUT se přihlásilo jako reprezentant iniciativy Průmysl 4.0 do konsorcia Gaia-X, což je určitá alternativa k EOSCu pro průmyslovou sféru.

Větší potenciál z hlediska rozvoje EOSC má přesah k veřejným datům ve veřejné správě, kde ale zatím zejména z kapacitních důvodů žádné aktivity nerozvíjíme.

**Jakým směrem by se měl rozvoj EOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOSC? Jaké aktivity jsou pro rozvoj EOSC v ČR klíčové? V jakém časovém horizontu by měly být realizovány?**

Mezi klíčové aktivity pro rozvoj EOSC patří především práce s lidskými zdroji. Je potřeba vědcům vysvětlit, o co jde, a přivést je k aktivnímu zapojení s vědomím, že se jedná o nedílnou součást běžné vědecké práce. Uživatelé by měli zjistit, že když mají data v pořádku a jsou připraveni tato data otevřít dalším zájemcům, lépe se jim bádá a spolupracuje. Znalost správných postupů práce daty jim zároveň pomůže lépe se pohybovat v evropském prostředí. Za svůj aktivní přístup by pak měli být uživatelé odměněni také v rámci svého karierního růstu.

Druhým klíčovým aspektem je vytvoření spolehlivého zázemí, které budou lidé moci pohodlně používat (metadatový adresář, národní repozitářová platforma, oborové repozitáře). Nemůže to být jediné unifikované řešení, musí být respektován současný stav a potřeby oborových clusterů. Evropa ocení aktivní přístup a zajímavá řešení.

Specifickou oblast implementace EOSC tvoří citlivá data, se kterými pracují nejen lékaři, ale třeba i ekonomové nebo sociologové. Je to tedy průřezová oblast, které je potřeba věnovat zvláštní pozornost. Je proto dobře, že se v rámci pracovních skupin pro rozvoj EOSC podařilo ustavit i pracovní skupinu zaměřenou na citlivá data.

Prostředky z intervence OP JAK by tedy měly jít primárně do lidských zdrojů, menší část prostředků by měla směřovat do rozvoje základní infrastruktury. Z hlediska časového plánu rozvoje předpokládáme, že základ metadatového adresáře by měl vzniknout do konce r. 2023 a funkční základ národní repozitářové platformy a jejího propojení s oborovými repozitáři během roku 2024. V r. 2025 by mělo dojít k uvedení do rutinního provozu a k zahájení debaty o dlouhodobé udržitelnosti infrastruktury po r. 2028.

**Jak by bylo rozumné v budoucnu vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOSC)?**

EOSC má v základních principech, že se neplatí za použití na úrovni koncového vědce. Výzkumné infrastruktury ale mají vlastní plány udržitelnosti a vlastní představy o práci s daty ve svých repozitářích. Debata o udržitelnosti po roce 2028 je zatím poněkud předčasná. Teprve po uvedení do

provozu v r. 2025 získáme přesnější představu o potřebné velikosti národní repozitářové platformy, tj. jaké množství dat půjde přímo do NRP a jaké do oborových repozitářů, a tedy i o nákladnosti dalšího provozu.

Další faktor podstatný pro udržitelnost celého zázemí EOSC je, jak se ČR postaví obecně k podpoře Open Science. To, co je dnes chápáno jako školení pro EOSC, se může stát běžnou součástí vzdělávání v rámci Open Science a mohlo by o ně být postaráno touto cestou.

### **Ing. Martin Svoboda**

ředitel NTK, zástupce projektu CARDS (IPs CARDS)

#### **Co od EOSC očekáváte? Proč je dobré EOSC v ČR rozvíjet?**

Hlavní rolí NTK ve vztahu k EOSC je metadatová podpora. Jde o zajištění naležitelnosti (Findability), dostupnosti (Accessibility) a Interoperability v rámci konceptu FAIR. EOSC v ČR je do značné míry otiskem aktivit z EU. Má zajistit, že vědecká data budou znovu využitelná. Důležité je, aby data byla ukládána v souladu s principy standardizace a kvalitního popisu. Myslím, že je účelné využít ověřených dobrých praktik (dávno již zavedených knihovnicí pro dokumenty). Jsme součástí Evropy a nebylo by výhodné zůstat stranou. Standardizace je výhodná i z ekonomického hlediska.

#### **Jak vnímáte současný stav implementace v ČR? Jaká jsou specifika v ČR oproti jiným státům?**

Současný přístup prosazovaný e-INFRA je dobrý v tom, že předpokládá vznik národní repozitářové platformy, která umožní poměrně jednoduché zakládání repozitářů každému, kdo má přiměřený IT potenciál a přitom nechce vyvíjet vlastní řešení. To je ve světě spíše nezvyklé. Tam většinou různé spolky nejprve vybudují své repozitáře a nepsaný standard se stává z těch úspěšných.

I v ČR už je ale spousta výzkumníků zapojena v projektech, které data do repozitářů ukládají, např. ELIXIR. Musíme se tedy vyrovnat i s tím, že nezačínáme od nuly, ale musíme respektovat některá pravidla stávajících repozitářů v zájmu zajištění interoperability.

#### **Jakým směrem by se měl rozvoj EOSC v ČR ubírat, aby uspokojil potřeby uživatelů a umožnil integraci v rámci evropských struktur EOSC? Jaké aktivity jsou pro rozvoj EOSC v ČR klíčové?**

Česká republika je – vzhledem ke zpoždění za světem – ve výhodném postavení a může se poučit z chyb svých předchůdců; to by ji ovšem mělo vést k pevnému odhodlání jejich chyby neopakovat. Vidím proto jako nezbytné od počátku v realitě zavádět postupy užívání standardizovaných metadatových schémat. Bez nich jsou sady vědeckých dat dostupné jen „těm, kdo se znají“ – a to asi není cílem.

Ve světě knihoven a procesů ukládání a vyhledávání textových dokumentů byl způsob jejich popisu, tj. tvorby metadat, propracován k dokonalosti, kterou všeobecný přechod k produkci, uchovávání i vyhledávání dokumentů v digitální podobě učinil téměř zbytečnou. Google ukázal, že texty i rozsáhlé dokumenty lze daleko lépe – pokud jsou dostupné v elektronické verzi – najít i bez kvalitního metadatového popisu díky rozsáhlé indexaci plných textů dokumentů.

To je ovšem situace zásadně odlišná od ukládání sad vědeckých dat. Už pro Findability sady dat identifikační údaje jako autor (včetně ORCID), název datové sady, datum publikování, název projektu, jeho akronym a číslo, licenční podmínky a trvalý identifikátor sady dat nestačí. Bez kvalitního popisu obsahu datové sady dokáže data najít jen ten, kdo zná autora nebo číslo projektu, ale zájemce nemá možnost data najít podle toho, k jakému tématu se vztahují. Tzv. věcný nebo obsahový popis je také nezbytným předpokladem pro dosažení kýžené Accessibility a zejména Interoperability ukládaných



dat. Je proto dobře, že se při návrhu implementace EOSC nezapomíná na metodickou podporu práce s metadaty výzkumných dat a perzistentními identifikátory.

Dalším naprosto klíčovým tématem je spolupráce a standardizace. Projekt je dobře nastaven v tom, že většinu energie nevěnuje na bezpochyby nesmírně významný technický návrh, ale na domluvu mezi aktéry.

Důležitá je synergie všech aktivit. Aby mohla nastat, musí všechny složky systému fungovat dohromady. Nezbytné je realistické uvažování a ochota ke spolupráci. Mj. je např. potřeba pečlivě zvažovat, jak dlouho má smysl konkrétní data uchovávat. To se liší v rámci oborů i v rámci jednotlivých projektů.

Odhad nákladů na úložné kapacity by se měl odvíjet od objemu dat a potřebné doby jejich uchování. Dobrým nástrojem by mohl být zvažovaný otevřený software Invenio, jehož vývoj vede Evropská organizace pro jaderný výzkum (CERN) a který využívá i Národní úložiště šedé literatury (NUŠL) v NTK.

### **Jak by bylo rozumné vyvážit veřejnou podporu a spoluúčast uživatelů (výzkumníků) na službách (ukládání a získávání dat v rámci EOSC)?**

Podle našich zkušeností z Czech E-LIBu je rozumným modelem dělba nákladů zhruba rovným dílem mezi státem a uživatele. Pokud nemusíte za služby platit, nebudete si jich vážit. Poskytnutí přístupu zcela zdarma pro rozjezd je rozumné pro zvýšení atraktivity služby, ale je třeba to limitovat na nějakou omezenou dobu.

Za trvale potřebnou naopak považují veřejnou podporu komunikačních a koordinačních aktivit. Může s tím být spojen i určitý nátlak na „slušné chování“ a respektování standardů na straně poskytovatelů i na straně uživatelů.

## 7.3 Slovníček pojmů implementace EOSC v ČR

Zdroj informací: e-INFRA CZ

**IPs CARDS** – Projekt (individuální projekt systémový financovaný OP JAK) reagující na systémové potřeby řízení VaVal v ČR s celonárodním dopadem, který navazuje na prioritní oblasti relevantních strategických dokumentů. Plánovaným nositelem projektu je NTK. Hlavními cíli projektu budou: pořízení a pilotní provoz Platformy nové generace (PNG) pro poskytování špičkových informačních služeb pro VaVal; zajištění odborné podpory práce s metadaty výzkumných dat; zajištění odborné podpory práce s PIDs.

**IPs EOSC-CZ** – Projekt (individuální projekt systémový financovaný OP JAK) reagující na systémové potřeby řízení VaVal v ČR s celonárodním dopadem, který navazuje na prioritní oblasti relevantních strategických dokumentů. Je plánováno, že projekt bude realizován konsorciem e-INFRA-CZ (CESNET, MU, VŠB-TUO), hlavním příjemcem za konsorcium bude MU. Projekt bude zajišťovat zázemí pro NDI, technickou realizaci národního metadatového adresáře (NMA), Sekretariát EOSC-CZ a Školící centrum EOSC-CZ.

**Národní datová infrastruktura (NDI)** – Národní datová infrastruktura je zastřešující pojem pro klíčové prvky pro datovou spolupráci na národní úrovni. Součástí NDI je Národní metadatový adresář (NMA), Národní repozitářová platforma (NRP), Sekretariát EOSC-CZ, Školící centrum EOSC-CZ, tematické a ostatní fyzické repozitáře, ale i systém politik a péče o lidské zdroje. NDI není prostým součtem uvedených prvků, ale zajišťuje jejich interoperabilitu a synergickou vzájemnou interakci.

**Národní informační portál EOSC v ČR** – Hlavní informační portál o EOSC v ČR, který bude zřízen v rámci projektu IPs EOSC-CZ. O tuto aktivitu se bude starat Sekretariát EOSC-CZ. Součástí portálu bude i komplexní nabídka školení na daná témata v celé ČR.

**Národní metadatový adresář (NMA)** – Centrální adresář a vyhledávač, který zajišťuje sběr metadat o výzkumných datech na základě základního metadatového modelu a vyhledávání v těchto metadatech. Předpokládá se, že do NMA budou indexována metadata z různých datových repozitářů a úložišť v ČR i zahraničí.

**Národní katalog repozitářů (NKR)** – Systém evidující repozitáře, metadatová schémata, řízené slovníky a ontologie v těchto repozitářích používané. NKR bude sloužit jako: a/ referenční zdroj pro získání základních informací o repozitářích a kontextu jejich obsahu; b/ adresář relevantních zdrojů k napojení do NMA; c/ evidence metadatových modelů i řízených slovníků. NKR bude přispívat ke koordinaci využívání modelů a slovníků v dalších (vznikajících) repozitářích, a tím také ke zlepšení kvality sémantické interoperability v NDI.

**Národní repozitářová platforma (NRP)** – Jádro NDI, které představuje multitenant technologickou platformu pro řešení správy a ukládání výzkumných dat pro různé druhy použití dle standardů přijatých Pracovními skupinami EOSC-CZ, schválených Koordinačním výborem EOSC-CZ a následně podporovaných a prosazovaných Sekretariátem EOSC-CZ.

**Perzistentní/trvalé identifikátory** – Identifikátory používané k trvalému a jednoznačnému určení tištěného nebo elektronického zdroje (např. ISBN = International Standard Book Number nebo DOI = Digital Object Identifier), osoby (ORCID = Open Researcher and Contributor ID) nebo instituce (ROR = Research Organization Registry).

**Platforma nové generace (PNG)** – Centrální moderní automatizovaný knihovní systém pro správu všech typů informačních zdrojů v akademických a výzkumných knihovnách. Součástí PNG bude nové moderní vyhledávací prostředí, umožňující vyhledávat jak tištěné, elektronické i digitalizované dokumenty, tak i výzkumná data.

**Pracovní skupiny EOSC-CZ** – Pracovní skupiny zřízené za účelem rozvoje spolupráce a zajištění potřebného konsensu v rámci implementace European Open Science Cloud (EOSC) v ČR mezi aktéry akademického sektoru v ČR. Koncept pracovních skupin je obsažen v Architektuře implementace iniciativy EOSC v ČR. V současné době jsou zřízeny 4 průřezové pracovní skupiny (Architektura NDI, Základní služby, Metadata, Vzdělávání a lidské zdroje), 7 tematických (oborově-vědních) pracovních skupin (Bio/Health/Food – Bio/Zdraví/Potraviny, Environmentální vědy, Humanitní vědy a umění, Sociální vědy, Datový management pro umělou inteligenci a strojové učení, Fyzikální vědy, Materiálové vědy a technologie) a jedna speciální pracovní skupina (Citlivá data).

**Sekretariát EOSC-CZ** – Bude představovat ústřední administrativní a podpůrnou součást implementace European Open Science Cloud (EOSC) v ČR. Mezi hlavní činnosti, které Sekretariát EOSC-CZ bude zajišťovat, patří: provoz a správa národního informačního portálu EOSC v ČR; logistická a administrativní podpora a odborné zázemí pro průřezové, tematické a případně další speciální pracovní skupiny pro implementaci EOSC v ČR; sledování stavu implementace EOSC v ČR a zavádění principů FAIR data; podpora koordinace, kooperace a spolupráce napříč projekty pro implementaci EOSC v ČR (publikace standardů) a podpora zapojení do mezinárodních aktivit v souvislosti s implementací EOSC v EU; spolupráce s Národním centrem pro Informační podporu VaVal (NCIP VaVal) při NTK v rámci přesahu tématu do rozvoje dalších součástí Open Science v ČR.

**Základní metadatový model (schéma)** – Model (nebo též schéma) sloužící pro základní popis výzkumných dat s ohledem na různorodost oborových metadat a dostupnost datových sad vzhledem k FAIR principům. Základní metadatový model určuje minimální povinné údaje pro indexaci záznamů do Národního metadatového adresáře (NMA), a následně může sloužit i jako východisko pro nově vznikající oborové repozitáře.

#### 7.4 Seznam zkratk

AAI	Authentication and authorization infrastructure
API	Application Programming Interface
BMBF	Bundesministerium für Bildung und Forschung (Německo)
BMBWF	Bundesministerium für Bildung, Wissenschaft und Forschung (Rakousko)
CARDS	Czech Academic and Research Discovery Services
CERIT-SC	Centrum vzdělávání, výzkumu a inovací pro ICT – Scientific Cloud
CERN	Conseil Européen pour la recherche nucléaire
CESNET	Czech Education and Scientific NETwork
CzechELib	Národní centrum pro elektronické informační zdroje
ČSDA	Český sociálněvědní datový archiv
DAU	Data Access Unit
DCC	Digital Curation Centre
DFG	Deutsche Forschungsgemeinschaft
DLTS	Dutch Techcentre for Life Sciences
DMP	Data Management Protocol
DOI	Digital Object Identifier
DTA	Data Transfer Agreement
EGI	European Grid Infrastructure
EHDS	European Health Data Space
EIF	European Interoperability Framework
EK	Evropská komise
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Reserch Infrastructures
EZK	Economische Zaken
FAIR	Findable, Accessible, Interoperable, Reusable

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

FCT	Fundação para a Ciência e a Tecnologia
FNSO	Le Fonds national pour la science ouverte
FZU AV ČR	Fyzikální ústav Akademie věd ČR
GENASIS	Global ENvironmental ASsessment Information System
H2020	Horizont 2020
HE	Horizont Evropa
IMTM	Institute of Molecular and Translation Medicine
ISBN	International Standard Book Number
ISC MUNI	Institute of Computer Science (Ústav výpočetní techniky), Masarykova univerzita
ISMN	International Standard Music Number
LPT	Long-term Preservation
MŠMT	Ministerstvo školství, mládeže a tělovýchovy
MU, MUNI	Masarykova univerzita
NCIP VaVal	Národní centrum pro informační podporu VaVal
NDCE	National Data Coordination and Expertise Centre
NDI	Národní datová infrastruktura
NFDI	Nationale Forschungsdateninfrastruktur
NKR	Národní katalog repozitářů
NMA	Národní metadatový adresář
NPOS	Nationaal Programma Open Science
NRP	Národní repozitářová platforma
NTK	Národní technická knihovna
NUŠL	Národní úložiště šedé literatury
NWO	Nederlandse Organisatie voor Wetenschappelijk Onderzoek
OCW	Onderwijs, Cultuur en Wetenschappen
OP JAK	Operační program Jan Amos Komenský
ORCID	Open Researcher and Contributor ID
PID	Persistent Identifier
PNG	Platforma nové generace
RCAAP	Repositórios Científicos de Acesso Aberto de Portugal
ROR	Research Organization Registry

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

SND	Svensk nationell datatjänst
SRIA	Strategic Research and Innovation Agenda
ÚFAL MFF UK	Ústav formální a aplikované lingvistiky Matematicko-fyzikální fakulty UK
ÚFCH AV ČR	Ústav fyzikální chemie J. Heyrovského Akademie věd ČR
UK	Univerzita Karlova
ÚOCHB AV ČR	Ústav organické chemie a biologie Akademie věd ČR
UP	Univerzita Palackého
VR	Vetenskapsrådet
VŠB-TUO	Vysoká škola báňská - Technická univerzita Ostrava
VVI	Velké výzkumné infrastruktury

## 8 Reference

Aktionsplan Forschungsdaten. [https://www.bmbf.de/bmbf/de/forschung/digitale-wirtschaft-und-gesellschaft/aktionsplan-forschungsdaten/aktionsplan-forschungsdaten\\_node.html](https://www.bmbf.de/bmbf/de/forschung/digitale-wirtschaft-und-gesellschaft/aktionsplan-forschungsdaten/aktionsplan-forschungsdaten_node.html)

Architektura implementace iniciativy EOSC v ČR. [https://www.e-infra.cz/file/8cbdb401961e0ab98fe1ac3595b63e3e/475/EOSC\\_architektura%20implementace.pdf](https://www.e-infra.cz/file/8cbdb401961e0ab98fe1ac3595b63e3e/475/EOSC_architektura%20implementace.pdf)

Bartošek, M., Růžička, M., Řepišová, Z., Hrubá, M. (2020): Průzkum vědeckých repozitářů v ČR, Zpráva a vyhodnocení. Masarykova univerzita, Ústav výpočetní techniky MU. <https://www.vyzkum.cz/FrontClanek.aspx?idsekce=876328>

Belsø, R., Hammargren, P., Ivarsson, L., Kontkanen, P., Riungu-Kalliosaari, L., Fischer, L., Garavelli S. (2022): EOSC-Nordic D2.6: Cross-border Collaboration in the Context of EOSC. <https://zenodo.org/record/6217744#.YhR9wOhBw2w>

Bianco, M., Abatecola, C., Gliozzi, C., Guerrini, M. (2021): Updated Policy Brief on Synergies between R&I partnerships and ESIF. [https://www.era-learn.eu/documents/policy\\_brief\\_synergies](https://www.era-learn.eu/documents/policy_brief_synergies)

Blanquer, I.: Identifying digital skill sets for EOSC. <https://drive.google.com/file/d/1AMvbC1ZIJXddUXatIPgnZlphbrppSSda/view?usp=sharing>

BMBF. [https://www.bmbf.de/bmbf/de/home/home\\_node.html](https://www.bmbf.de/bmbf/de/home/home_node.html)

BMBWF. <https://www.bmbwf.gv.at/>

Bodlos, A., Hönegger, L., Kaczmirek, L., Beckmann, V., Breton, V., Romier, G., van Wezel, J., Streit, A., Stevanovic, U., Galeazzi, F., Tanlongo, F., van Nieuwerburgh, I. (2020): Summary report of the EOSC-Pillar\National Initiatives" Survey'. Version 1.1, Zenodo. doi: 10.5281/zenodo.3937318. <https://zenodo.org/record/3937318#.YkHcNudBw2w>

Brown, Josh, Jones, Phill, Meadows, Alice, Murphy, Fiona, & Clayton, Paul. (2021): UK PID Consortium: Cost-Benefit Analysis (1.0). Zenodo. <https://doi.org/10.5281/zenodo.4772627>

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

CESSDA (2020): CESSDA Data Management Expert Guide <https://zenodo.org/record/3820473#.YkAsj-dBy5d>

CESSDA (2019): Adapt your Data Management Plan. A list of Data Management Questions based on the Data Management Expert Guide.

[https://www.cessda.eu/content/download/4302/48656/file/TTT\\_DO\\_DMPExpertGuide\\_v1.3.pdf](https://www.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.3.pdf)

Cestovní mapa ESFRI 2021. <https://roadmap2021.esfri.eu/>

Cordis. <https://cordis.europa.eu/>

Deuxième Plan national pour la science ouverte. <https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

DFG. <https://www.dfg.de/index.jsp>

EOSC (2021a): EOSC Interoperability Framework. Report from the EOSC Executive Board Working Groups (WG) FAIR and Architecture. <https://op.europa.eu/cs/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1>

ERC (2021b). Open Research Data and Data Management Plans. Information for ERC grantees by the ERC Scientific Council. Version 4.0. 11 August 2021.

[https://erc.europa.eu/sites/default/files/document/file/ERC\\_info\\_document-Open\\_Research\\_Data\\_and\\_Data\\_Management\\_Plans.pdf](https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf)

EOSC Glossary. <https://eosc-portal.eu/glossary>

EOSC Procurement. <https://ted.europa.eu/udl?uri=TED:NOTICE:234660-2022:TEXT:EN:HTML>

EOSC Secretariat: Identifying digital skill sets for EOSC. <https://www.eoscsecretariat.eu/news-opinion/identifying-digital-skill-sets-eosc>

EOSC Support Office Austria. <https://eosc-austria.at/>

European Commission (2017): European Interoperability Framework.

[https://ec.europa.eu/isa2/sites/default/files/eif\\_brochure\\_final.pdf](https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf)

European Commission (2018): Final Report and Action Plan from the European Commission Expert Group on FAIR Data. TURNING FAIR INTO REALITY.

[https://ec.europa.eu/info/sites/default/files/turning\\_fair\\_into\\_reality\\_0.pdf](https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_0.pdf)

European Commission, Directorate-General for Research and Innovation (2019): Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data, Publications Office,

<https://data.europa.eu/doi/10.2777/02999>

European Commission (2020a): A Persistent Identifier (PID) policy for the European Open Science Cloud. <https://doi.org/10.2777/926037>

European Commission (2020b): PID Architecture for the EOSC. <https://doi.org/10.2777/525581>

European Commission (2021): Memorandum of Understanding for the Co-programmed European Partnership for the European Open Science Cloud (EOSC).

[https://ec.europa.eu/info/sites/default/files/research\\_and\\_innovation/funding/documents/c\\_2021\\_4113\\_f1\\_annex\\_en\\_v3\\_p1\\_1213802.pdf](https://ec.europa.eu/info/sites/default/files/research_and_innovation/funding/documents/c_2021_4113_f1_annex_en_v3_p1_1213802.pdf)

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

Exploring and optimising the Dutch data landscape.

<https://www.openscience.nl/files/openscience/2020-11/NPOS%20%282020%29%20Final%20report%20Exploring%20and%20optimising%20the%20Dutch%20data%20landscape.pdf>

FCT. <https://www.fct.pt/>

Funding and Tender portál. <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/home>

Heikkurinen, M. (2020): National/International engagement plan with policy makers and funders. EOSC-SYNERGY-D5.1. <http://dx.doi.org/10.20350/digitalCSIC/12611>

Hellström, Maggie, Heughebaert, André, Kotarski, Rachael, Manghi, Paolo, Matthews, Brian, Ritz, Raphael, Conrad, Anders Sparre, Weigel, Tobias, & Wittenburg, Peter. (2019): Initial Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC) (1.0). Zenodo. <https://doi.org/10.5281/zenodo.3574203>

Hnátková, E., Dibuszová, E., Svoboda, M. (2022): Otevená věda: Analýza mezinárodního prostředí. NTK, projekt Národní centrum pro informační podporu výzkumu, vývoje a inovací. <https://doi.org/10.48813/80s4-bv02>

Horizont Evropa. <https://www.horizontevropa.cz/cs>

Horizon Europe (2021): Data Management Plan Template [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan-template\\_he\\_en.docx](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan-template_he_en.docx)

Implementation Plan Investments Digital Research Infrastructure.

<https://www.nwo.nl/en/researchprogrammes/implementation-plan-investments-digital-research-infrastructure>

Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, <https://www.enseignementsup-recherche.gouv.fr/fr>

NFDI. <https://www.nfdi.de/>

NPOS. <https://www.openscience.nl/en/national-programme-open-science>

NWO. <https://www.nwo.nl/>

OCW. <https://www.rijksoverheid.nl/ministeries/ministerie-van-onderwijs-cultuur-en-wetenschap>

OECD (2007): OECD Principles and Guidelines for Access to Research Data from Public Funding. <https://www.oecd.org/sti/inno/38500813.pdf>

Open Science 2030 in the Netherlands – NPOS2030 Ambition Document.

<https://survey.surf.nl/upload/surveys/493148/files/NPOS2030%20Ambition%20Document.pdf>

Open Science Policy Austria. Österreichische Policy zu Open Science und der European Open Science Cloud. <https://www.bmbwf.gv.at/Themen/HS-Uni/Hochschulgovernance/Leitthemen/Digitalisierung/Open-Science/Open-Science-Policy-Austria.html>

Ouvrir la science. <https://www.ouvrirlascience.fr/accueil/>

## Analýza aktuálního stavu kapacit pro implementaci EOSC v ČR

Pracovní program „Evropské výzkumné infrastruktury – včetně e-infrastruktur“, 2016-2017.

<https://www.h2020.cz/cs/storage/f753c31d4ec41c302d7be26021b0573816d9721b?uid=f753c31d4ec41c302d7be26021b0573816d9721b>

Pracovní program „Evropské výzkumné infrastruktury – včetně e-infrastruktur“, 2018-2019.

<https://www.h2020.cz/cs/storage/4b13baa81e92f9d455741c29f3bda8a4e04f2a5f?uid=4b13baa81e92f9d455741c29f3bda8a4e04f2a5f>

PwC EU Services (2018). Cost of not having FAIR research data. Cost-Benefit analysis for FAIR research data. <https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en>

RCAAP. <https://www.rcaap.pt/>

Recherche Data Gouv. <https://www.ouvrirelascience.fr/recherche-data-gouv-plateforme-nationale-federee-des-donnees-de-la-recherche/>

SND. <https://snd.gu.se/sv>

Strategic plan HE 2021-2024.

[https://www.eeas.europa.eu/sites/default/files/horizon\\_europe\\_strategic\\_plan\\_2021-2024.pdf](https://www.eeas.europa.eu/sites/default/files/horizon_europe_strategic_plan_2021-2024.pdf)

Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC). 2021.

[https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0\\_15Feb2021.pdf](https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf)

Swedish National Roadmap for the European Research Area 2019–2020.

<https://www.government.se/information-material/2019/05/swedish-national-roadmap-for-the-european-research-area-20192020/>

Swedish Research Council. <https://www.vr.se/>

Tuuli (2017). DMPTuuli - Data Management Planning Tool. <https://www.dmptuuli.fi/>

Van Berchum, M. & Grootveld, M. (2017). Research data management. An overview of recent developments in the Netherlands. <https://pure.knaw.nl/portal/en/publications/research-data-management-an-overview-of-recent-developments-in-th>

Votavová, B., Hnátková, E., Dibuszová, E., Svoboda, M. (2022): Analýza účelnosti a zájmu o centralizované zajišťování perzistentního identifikátoru ORCID iD. NTK, projekt Národní centrum pro informační podporu výzkumu, vývoje a inovací. <https://doi.org/10.48813/6erb-z009>

Wilkinson, Mark D. at al (2016): The FAIR Guiding Principles for scientific data management and stewardship. <https://www.nature.com/articles/sdata201618>