

# ARCHITEKTURA IMPLEMENTACE INICIATIVY EOSC V ČR

## A. VÝCHODISKA A RÁMCOVÉ PRINCIPY

1. **Monitoring datových úložišť** existujících velkých výzkumných infrastruktur, resp. institucionálních datových úložišť výzkumných organizací ČR a **příkladů dobré praxe nakládání s vědeckými daty na úrovni ČR, resp. propojení datových repositářů ČR na existující mezinárodní iniciativy** – vytvoření základu pro dlouhodobý monitoring a vyhodnocování postupu implementace EOSC na úrovni ČR
2. **Identifikace optimálních vědně-oborových EOSC komunit (clusterů) v rámci ČR a indikace potřeb takto identifikovaných odborných komunit ČR** ve střednědobém výhledu 3–5 let a dlouhodobém výhledu až 7 let, co se týče nezbytných investic do hardware, software a rozvoje lidských zdrojů
3. **Definice technického rámce a sady společných parametrů a standardů**, umožňujících budování konsistentního systému federované datové infrastruktury v ČR, vycházející z národní repositářové platformy, jakožto páteře komplexní datové infrastruktury ČR, a z vědně-oborových clusterů, které tuto páteř doplňují o specifické vědní oblasti
4. **Podpora aktivit rozvoje datové infrastruktury ČR** centrální podpůrnou jednotkou, jako ústředním referenčním hubem a zprostředkovatelem agendy a standardů rozvoje datové infrastruktury ČR; centrální podpůrná jednotka bude rovněž zprostředkovávat **konceptní propojení implementace EOSC v ČR na evropskou úroveň**. Dalšími aspekty jsou zajištění *data sovereignty* (dlouhodobá záruka manipulace s daty v souladu s legislativním prostředím a potřebami ČR) a příprava dlouhodobého uchovávání dat (dlouhodobé udržitelnosti vědeckých dat a jejich managementu)
5. **Adresování otázek spojených s přenosem FAIR data principů do legislativního prostředí ČR, resp. institucionálního prostředí výzkumných organizací ČR** za účelem optimalizace federované datové infrastruktury ČR, budované a operované v návaznosti na evropskou EOSC datovou infrastrukturu; napojení datové infrastruktury ČR na existující národní iniciativy (otevřené pro evropskou úroveň)
6. **Propagace a podpora pozitivního přijetí konceptu *Open Science* v ČR**, tj. podpora kultury ukládání a sdílení vědeckých dat a prosazování této kultury, jako jednoho ze stěžejních principů při rozvoji lidských zdrojů ve vědecké sféře a tvorbě a implementaci veřejných politik v oblasti výzkumu v ČR

## B. SEKRETARIÁT EOSC-CZ: PODPŮRNÁ JEDNOTKA DATOVÉ INFRASTRUKTURY ČR

1. **Podporuje přenos technických standardů evropské EOSC datové infrastruktury během budování a provozu datové infrastruktury ČR a napomáhá k napojení datové infrastruktury ČR na síť EOSC**, co do přenosu evropské dobré praxe *data policy* a *data managementu* do prostředí ČR
2. **Zprostředkovává vzájemnou interakci mezi vědně-oborovými clustery a národní repositářovou platformou datové infrastruktury ČR**, co do rozvoje a kontroly dodržování technických specifikací při budování datových repositářů a zajištění jejich interoperabilního provozu na bázi federativního modelu datové infrastruktury ČR, primárně formou aplikace technických standardů napříč národní datovou infrastrukturou ČR
3. **Ustavuje odborné pracovní skupiny k adresování otázek souvisejících s implementací EOSC v ČR** a spolupracuje na jejich adresování s vědně-oborovými EOSC komunitami, a to **na principu tzv. *co-creation* a *co-ownership*** přístupu

4. **Etabluje se při NTK ve vzájemné spolupráci s vysokými školami, AVČR, jako i dalšími výzkumnými organizacemi ČR (Koordinační rada Sekretariátu EOSC-CZ) a dále i koordinátory vědně-oborových repositářů datové infrastruktury ČR**
5. **Ve spolupráci s e-INFRA CZ tvoří, komunikuje a prosazuje společnou pozici federované národní datové infrastruktury ČR v rámci EOSC Association, dále koordinuje realizaci národní *multi-tenant* repositářové platformy ČR, zabezpečuje centrální metadatový interoperabilní adresář a hájí zájmy definované vědně-oborovými clustery a národní repositářovou platformou datové infrastruktury ČR při budování a provozu evropské EOSC datové infrastruktury**
6. **Podporuje CESNET v roli tzv. *Mandated Organisation*, ve které CESNET na základě pověření od MŠMT zajišťuje zastoupení ČR v EOSC Association, tím, že zprostředkovává koordinaci klíčových stakeholderů na národní úrovni ČR a umožňuje jejich propojení na evropské výzkumné komunity při budování a provozu datové infrastruktury EU na federativním principu**
7. **Uskutečňuje informační, diseminační, školicí a vzdělávací aktivity pro implementaci EOSC v ČR, vhodně doplňující související oborové a institucionální aktivity**

#### **C. VĚDNĚ-OBOROVÉ REPOSITÁŘE DATOVÉ INFRASTRUKTURY ČR**

1. **Sdružují, budují a rozvíjí kapacity datových center výzkumných organizací ČR na základě vědně-oborového klíče a vznikají okolo relevantních vědeckých disciplín a jejich institucionálních nositelů a velkých výzkumných infrastruktur s již existujícími robustními *data policy* a *data management* plány. Vědně-oborové repositáře jsou napojeny na evropské a další mezinárodní iniciativy ukládání a sdílení vědeckých dat, pokud takové existují**
2. **Působí jako huby datové infrastruktury ČR pro vzájemně příbuzné skupiny vědních oborů, u nichž jsou identifikovány obdobné očekávané nároky na datovou infrastrukturu a na způsob nakládání s daty, a sdílením dobré praxe zprostředkovávají principy EOSC příbuzným vědním oborům s méně rozvinutou datovou politikou**
3. **Naplňují principy inkluze a vyvážené institucionální a geografické prezence napříč celou ČR, avšak současně vždy za předpokladu naplnění klíčových minimálních požadavků ze strany asociujících se partnerů, co se týče stavu rozvoje a implementace jejich *data policy* a *data management* plánů a standardů pro technickou interoperabilitu**
4. **Vznikají jako centralizované anebo distribuované huby datové infrastruktury ČR na základě sebeorganizace vědecké komunity ČR a naplnění vstupních předpokladů pro federaci datových center ze strany asociujících se partnerů**
5. **Budovány a provozovány jsou v gesční příslušnosti a ve vzájemné spolupráci veřejných vysokých škol, veřejných výzkumných institucí a dalších výzkumných organizací ČR v rámci konsorciálních projektů koordinovaných se Sekretariátem EOSC-CZ**
6. **Účastní se odborných pracovních skupin k adresování otázek implementace EOSC v ČR, přičemž úzce spolupracují na jejich adresování se Sekretariátem EOSC-CZ na principu tzv. *co-creation* a *co-ownership* přístupu**
7. **Institucionální repositáře výzkumných organizací budou za využití zdrojů Operačního programu Jan Amos Komenský (OP JAK) soutěžených prostřednictvím výzev *Open Science* rozvíjeny v jejich kapacitě, kterou budou zapojeny do vědně-oborových clusterů datové infrastruktury ČR anebo**

**do národní repositářové platformy** (rozvoj nad rámec tohoto vymezení budou muset výzkumné organizace v případě svého zájmu financovat už za využití jiných rozpočtových zdrojů, než OP JAK)

**8. Prvotní návrh vědně-oborových clusterů, viz níže, vyplývá z číselníku skupin oborů podle Frascati manuálu, přičemž definitivní výčet vědně-oborových clusterů bude dále debatován na platformách odborných pracovních skupin pro adresování otázek souvisejících s implementací EOSC v ČR, které budou ustaveny při Sekretariátu EOSC-CZ**

- Natural Sciences
- Engineering and Technology
- Medical and Health Sciences
- Agricultural and Veterinary Sciences
- Social Sciences
- Humanities and the Arts

#### **D. NÁRODNÍ REPOSITÁŘOVÁ PLATFORMA DATOVÉ INFRASTRUKTURY ČR**

- 1. Představuje páteř národní datové infrastruktury ČR a jejích služeb, a to včetně jejího napojení na evropskou EOSC datovou infrastrukturu, s primárním použitím pro vědní obory a výzkumné skupiny nezapojené do struktury vědně-oborových repositářů, čímž je zajištěna potřeba naplnění principů *Open Science* a FAIR nakládání s vědeckými daty v plném rozsahu vědních disciplín**
- 2. Etabluje se jako distribuovaná *multi-tenant* platforma, v níž je možné vytvářet konkrétní datové repositáře, asociované se všemi ostatními složkami datové infrastruktury ČR v rámci systému vzájemně interoperabilních a kompatibilních datových center ČR, a naplňuje také všechna kritéria technického rámce a sady společných parametrů datové infrastruktury ČR; současně dále zajišťuje základní společné služby, jako je tvorba základních metadat a řízení přístupu k datům**
- 3. Národní repositářová platforma je budována a provozována v gesční příslušnosti NTK, e-INFRA CZ a AVČR se zapojením vysokých škol, veřejných výzkumných institucí zřízených AVČR a dalších výzkumných organizací v ČR, a slouží pro zajištění ad hoc potřeb nakládání s vědeckými daty, která nevyžadují specifickou podporu vědně-oborových repositářů datové infrastruktury ČR**
- 4. Součástí základní páteře datové infrastruktury ČR je mj. základní metadatový adresář, budovaný v gesční příslušnosti NTK, e-INFRA CZ, AVČR a představitelů vědně-oborových repositářů datové infrastruktury ČR**
- 5. V rámci národního datového repositáře datové infrastruktury ČR jsou výzkumná data schraňována, ale jejich původcům, tzn., autorským výzkumným organizacím a výzkumným týmům i samotným výzkumníkům, zůstává nadále plná kontrola nad takto schraňovanými vědeckými daty**

## PODPORA IMPLEMENTACE INICIATIVY EOSC V ČR

Finanční podpora implementace iniciativy EOSC v ČR by měla být MŠMT poskytována z **Operačního programu Jan Amos Komenský (OP JAK)**<sup>1</sup> prostřednictvím vyhlášení dvou výzev *Open Science* a výzvy na podporu realizace **Individuálního projektu systémového**<sup>2</sup> (IPs), jehož nositelem by se měla stát **NTK ve spolupráci s e-INFRA CZ**. První výzva *Open Science* a výzva na IPs by měly být vyhlášeny v roce 2022.

### Individuální projekt systémový

V rámci IPs OP JAK budou realizovány aktivity směřující k tvorbě jednotného a vzájemně propojeného prostředí pro implementaci politiky *Open Science* v ČR.<sup>3</sup> Tyto aktivity budou zahrnovat:

- **Zavádění a ověření platformy nové generace vyhledávání a správy informačních zdrojů společně pro informační centra výzkumných organizací a pro knihovny s celostátní působností** a vytvoření nadstavbového podpůrného software na bázi umělé inteligence
- **Podpora budování jádra/základu národní repositářové platformy datové infrastruktury ČR a také standardizace přístupu k vědeckým datům ustavením a provozem Sekretariátu EOSC-CZ**, včetně školicích a vzdělávacích aktivit v oblasti *data management a stewardship* s přesahem do principů *Open Science*
- **Ustavení centra podpory zavádění perzistentních identifikátorů ve výzkumu, vývoji a inovacích** pro osoby, digitální objekty, instituce, publikace apod.
- **Posilování interoperability stávajících anebo zaváděných datových, informačních, knihovnických a ostatních systémů a repositářů** ve výzkumu, vývoji a inovacích ve spolupráci s Radou pro výzkum, vývoj a inovace, jako správcem Informačního systému výzkumu, vývoje a inovací ČR

Dílní aktivity realizované v IPs OP JAK ze strany NTK budou komplementární k aktivitám **Národního centra pro informační podporu výzkumu, vývoje a inovací (NCIP VaVal)**, a to zejména v oblastech otevřeného přístupu k vědeckým publikacím a národnímu repositáři pro ně, k centru pro přidělování a správu identifikátoru ORCID a vybraným podpůrným službám v oblasti vzdělávání a k dalším aktivitám v oblastech uchování vědeckých dat a nakládání s nimi.

### Výzvy Open Science

Výzvy *Open Science* by měly podporovat implementaci iniciativy EOSC na úrovni výzkumných institucí, resp. konsorcií těchto výzkumných organizací. Ve výzvách by takto měly být podpořeny **vědně-oborové repositáře datové infrastruktury ČR a rozvoj národní repositářové platformy datové infrastruktury ČR**.

Řídicí výbor OP JAK předpokládá, že dojde k **vyhlášení 2 na sebe navazujících výzev**. Vyhlášení 1. výzvy je předpokládáno v roce 2022; vyhlášení 2. výzvy poté cca o 2 roky později. Předpokládaná rozpočtová alokace na obě dvě výzvy dohromady by měla činit cca **2,5 mld. Kč**.

<sup>1</sup> *Finální podoba OP JAK závisí na výsledcích vyjednávání s Evropskou komisí.*

<sup>2</sup> *Projekt bude podléhat hodnoticímu procesu věcně příslušné výzvy OP JAK.*

<sup>3</sup> *IPs OP JAK bude spojovat aktivity plánované v ideovém záměru projektu IPs CARDS a aktivity sekretariátu EOSC-CZ. Název projektu nově vzniklého tímto spojením bude upřesněn.*

Předpokládané podporované aktivity výzev *Open Science* budou:

- **Podpora investičních výdajů na datové repositáře, rozvoj národní repositářové platformy a s tím související technologie;**
- **Podpora vzniku metodik/standardů pro práci s daty** v koordinaci se Sekretariátem EOSC-CZ;
- **Vznik a další podpora pracovních pozic pro odborníky na práci s daty** a další relevantní odborníky;
- **Vzdělávání lidských zdrojů pro oblast *Open Science*** – pozn.: vzdělávání může být koncipováno tak, že nemusí směřovat výlučně dovnitř institucí příjemců/partnerů, avšak také do vnějšího prostředí, zejména v rámci vědně-oborově clusterů datové infrastruktury ČR (tzn., „zkušenější“ výzkumné instituce mohou sehrát roli vzdělavatelů pro ty instituce, které dosud s *data management* takové zkušenosti nemají);
- **Povinná spolupráce s podpůrnou jednotkou** (tzn., se Sekretariátem EOSC-CZ); povinné dodržování standardů definovaných odbornými pracovními skupinami Sekretariátu EOSC-CZ.

**Výzvy OP JAK a podporované aktivity, a to především vzdělávací aktivity, budou připraveny tak, aby nedocházelo k překryvu s aktivitami, jež by měly být podpořeny z IPs OP JAK a projektu NCIP VaVal.**