

ZÁKLADNÍ OBSAH PROJEKTU NRP - VÝZVA OPEN SCIENCE I-



Obsah

1.	ZKRATKY A VYSVĚTLIVKY	4
2.	ZÁKLADNÍ ÚDAJE	6
3.	ÚVOD	6
4.	VAZBA NA JINÉ PROJEKTY.....	6
5.	CÍLE PROJEKTU.....	8
6.	ORGANIZAČNÍ STRUKTURA, ŘÍZENÍ PROJEKTU A POPIS ROLÍ REALIZAČNÍHO TÝMU.....	10
6.1	Organizační struktura a řízení projektu.....	10
7.	KLÍČOVÉ AKTIVITY.....	11
7.1	KA1 – Povinná aktivita „řízení projektu“	11
7.2	KA 2 – „Vývoj a provoz Národní repozitářové platformy“	11
7.2.1	KA 2.1 Vývoj a provoz technického základu repozitářové platformy.....	14
7.2.2	KA 2.2 Vývoj a provoz systému Invenio.....	14
7.2.3	KA 2.3 Vývoj a provoz systému CLARIN-DSpace.....	15
7.2.4	KA 2.4 Vývoj a provoz systému ASEP/ARL.....	15
7.3	KA 3 – „Pilotní repozitáře“	15
7.3.1	KA 3.1 Catch-all repozitář	16
7.3.2	KA 3.2 Pilotní repozitář pro biologická zobrazovací data	17
7.3.3	KA 3.3 Pilotní repozitář národní úložiště pro biodiverzitní data	18
7.3.4	KA 3.4 Pilotní repozitář Molecular BIOPHYSICS DATABASE (MBDB).....	19
7.3.5	KA 3.5 Pilotní repozitář HUME Lab.....	20
7.4	KA 4 – „Integrace služeb NRP“	21
7.4.1	KA 4.1 Správa metadatových profilů v NRP.....	22
7.4.2	KA 4.2 Podpora práce s licencemi	22
7.4.3	KA 4.3 Řízení přístupu k datům	23
7.5	KA 5 – „Vývoj a provoz nových služeb NRP“	24
7.5.1	KA 5.1 Podpora FAIRifikace výzkumných dat	25
7.5.2	KA 5.2 Podpora plánování správy dat.....	26
7.5.3	KA 5.3 Automatizace sběru dat a metadat.....	27
7.5.4	KA 5.4 Integrace výpočetních workflows	29
7.5.5	KA 5.5 Roadmapa pokročilých služeb NRP/NDI	30
7.6	KA 6 – Kyberbezpečnost a podpůrné služby	30
7.6.1	KA 6.1 Kyberbezpečnostní dohled a zabezpečení NRP	31
7.6.2	KA 6.2 Compliance a řízení ISMS i s péčí o citlivá data.....	32
7.6.3	KA 6.3 Pro-uživatelský přístup implementace služeb NRP a NDI	33

7.6.4	KA 6.4 Uživatelská podpora a Service Desk.....	35
7.7	KA 7 – „Školení a osvěta v oblasti funkcionalit a služeb NRP“	35
7.7.1	KA7.1 Metodické materiály pro software/služby NRP	37
7.7.2	KA7.2 Komunita data stewardů.....	38
7.7.3	KA7.3 Konzultační služby a školicí služby pokročilé podpory NRP	38

1. ZKRATKY A VYSVĚTLIVKY

Zkratka	Popis
CARDS	Czech Academic and Research Discovery Services; IPs projekt, jehož řešitelem je Národní technická knihovna, viz cards.techlib.cz
Centrum PID	Národní centrum pro perzistentní identifikátory; metodické pracoviště NTK financované z IPs CARDS; https://identifikatory.cz/cs/
CERIT-SC	Pracoviště Ústavu výpočetní techniky MU; národní centrum poskytující vysoce flexibilní úložné a výpočetní zdroje a související služby
DMP	Data Management Planning, plánování správy dat
EOSC	European Open Science Cloud
EOSC-CZ	European Open Science Cloud Czech Republic; IPs projekt, jehož řešitelem je Masarykova univerzita
FM	Finanční manažer
FTE	Full Time Equivalent; pojem, kterým se označuje jednotka vyjadřující míru zapojení či kapacitu vytížení pracovníka přepočtená na 100% úvazek
HMP	Hlavní manažer Projektu
HW	Hardware
IČO	Identifikační číslo osoby
IS KP21+	Informační systém konečného žadatele/ příjemce
IT4I	IT4Innovations; národní superpočítačové centrum VŠB-TUO
Mxy	Měsíc realizace projektu
NDI	Národní datová infrastruktura
NKR	Národní katalog repozitářů, plánovaný výstup IPs CARDS
NMA	Národní metadatový adresář, plánovaný výstup IPs EOSC-CZ
NRP	Národní repozitářová platforma
OT	Odborný tým
PID	Persistent Identifier, perzistentní identifikátor
PM	Projektový manažer
xyPM	Person months; pojem vyjadřující práci jednoho člověka za měsíc
ŘVP	Řídicí výbor Projektu
SPpŽP	Pravidla pro žadatele a příjemce – specifická část
SW	Software

VaV	Výzkum a vývoj
VVP	Výkonný výbor projektu
ZPP	Základní parametry projektu (příloha prvního aktu)

2. ZÁKLADNÍ ÚDAJE

Název projektu	Národní repozitářová platforma pro výzkumná data
Název žadatele	CESNET, zájmové sdružení právnických osob
Název partnera/partnerů	Biotechnologický ústav AV ČR, v. v. i. České vysoké učení technické v Praze Knihovna AV ČR, v. v. i. Masarykova univerzita Národní technická knihovna Středisko společných činností AV ČR, v. v. i. Univerzita Karlova Univerzita Palackého v Olomouci Ústav fyzikální chemie J. Heyrovského AV ČR, v. v. i. Vysoká škola báňská – Technická univerzita Ostrava

3. ÚVOD

Projekt *NRP pro výzkumná data* se soustředí na realizaci klíčových komponent, které systémově přispějí k výraznému zlepšení současného stavu v oblasti správy výzkumných dat. Hlavní výstupy se tak soustředí na rozvoj architektury Národní datové infrastruktury (NDI) a specificky NRP jako její ústřední součásti (a to včetně náležitého zabezpečení a technické podpory), zprovoznění konkrétních platform pro tvorbu repozitářů NRP, pilotní provoz vybraných repozitářů a „catch-all“ repozitáře, implementované v rámci NRP, zajištění funkčních vazeb mezi konkrétními repozitáři a systémem užívání PID, zajištění komponent pro napojení repozitářů na Národní metadatový adresář a zajištění podpory pro práci s licencemi, vývoj integrovaného řešení pro plánování správy dat, stejně tak jako vytvoření koordinované vzdělávací podpory pro efektivní využití NRP.

4. VAZBA NA JINÉ PROJEKTY

Návaznost předkládaného projektu s projekty IPs CARDS a IPs EOSC-CZ vychází ze strategického dokumentu *Architektura implementace EOSC v ČR*¹. Společným cílem předkládaného i obou zmíněných projektů je vybudování národní datové infrastruktury (NDI) jako základu ekosystému služeb a nástrojů zajišťujících implementaci EOSC v ČR. Realizací těchto projektů vznikne komplexní ekosystém, který vědecké komunitě poskytne prostředí pro kvalitní správu FAIR výzkumných dat. Cíle projektu a jejich vzájemné provázanosti jsou uvedeny dále v této kapitole, případně jsou dále rozvedeny v kapitole 7 KLÍČOVÉ AKTIVITY.

Hlavními prvky Národní datové infrastruktury (NDI) jsou:

- **Národní repozitářová platforma (NRP)** — distribuované multiplatformní prostředí pro vytváření instancí datových repozitářů, tedy datových úložišť ukládajících data opatřená bohatými metadaty pro naplnění FAIR principů,
- **Národní metadatový adresář (NMA)** — agregátor metadat a centrální vyhledávací bod pro datové sady z vědeckého prostředí v ČR,

¹ https://www.eosc.cz/media/3517655/eosc_architektura-implementace.pdf

- **Národní katalog repozitářů (NKR)** — nástroj pro evidenci repozitářů, metadatových schémat a řízených slovníků užitých v NDI,
- **soustava datových úložišť pro zpracovávaná data**, která zahrnuje jednak operativní úložiště MetaCentra a IT4Innovations, jednak vysokokapacitní úložiště pro rozpracovaná data velkých objemů,
- **pomocné systémy pro integraci**, jako např. systémy pro přidělování perzistentních identifikátorů digitálním objektům, integrace AAI, služby pro datové přenosy z a do prostředí datových repozitářů a podobně.

Následující schéma představuje celý ekosystém, s jasným vyznačením odpovědnost projektu NRP a klíčovými prvky implementace v rámci projektu EOSC-CZ (Národní metadatový adresář) a CARDS (Národní katalog repozitářů a Centrální vyhledávací portál; včetně Platformy nové generace pro bibliografické záznamy).

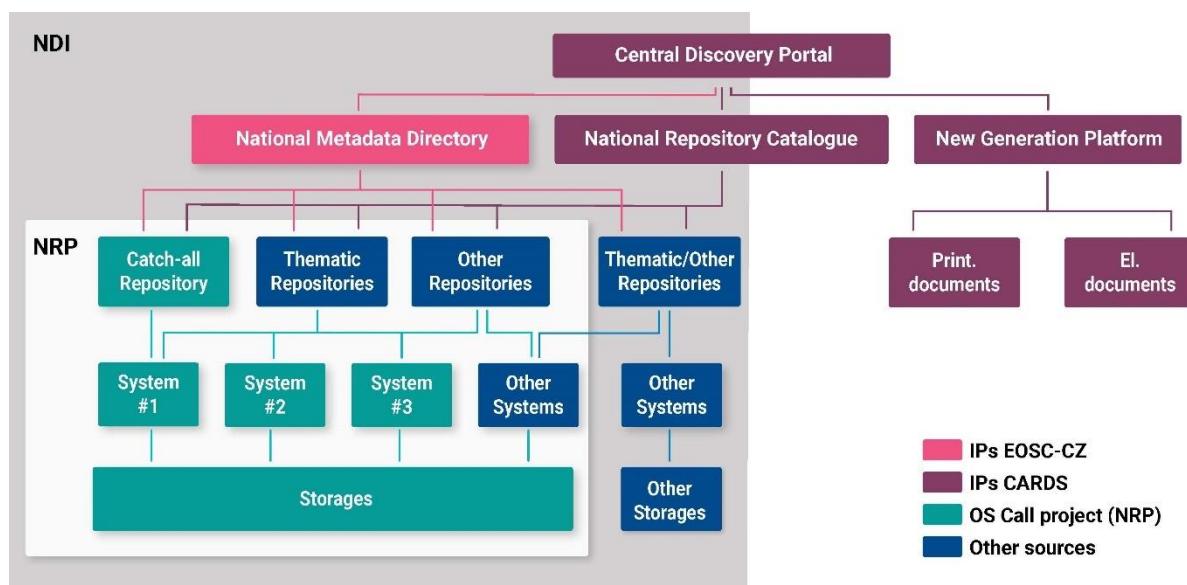


Schéma 1 Ekosystém služeb e-infrastruktury

Další součástí Architektury implementace EOSC v ČR jsou následující „netechnické“ komponenty:

- Sekretariát EOSC, zajišťující administrativní, podpůrné a monitorovací součásti implementace iniciativy EOSC v ČR,
- Virtuální Školicí centrum, budované v rámci IPs EOSC-CZ.

Projekt Národní repozitářová platforma pro výzkumná data dále staví na dosavadních aktivitách členů dalších projektů.

5. CÍLE PROJEKTU

Cílem projektu *Národní repozitářová platforma pro výzkumná data* je návrh, implementace a pilotní provoz Národní repozitářové platformy a systému bezprostředně souvisejících klíčových služeb, společně se zajištěním vzdělávání uživatelů, datových odborníků a dalších relevantních pracovníků výzkumných organizací. Nezbytným souvisejícím cílem je vytvoření kyberbezpečnostního zázemí včetně garance plné compliance s relevantními zákonnými i dalšími předpisy a rovněž zajištění metodické i provozní podpory uživatelů vytvořené Národní repozitářové platformy.

Základních cílů projektu bude dosaženo postupnou realizací následujících dílčích cílů:

Technické a softwarové zázemí národní repozitářové platformy: Pořízení, instalace, zprovoznění a další rozvoj základní úložné kapacity, včetně vrstvy objektového obecného úložiště, které umožní vytváření a provoz repozitářových systémů i samostatných repozitářů (viz další bod). Celkově bude k dispozici cca 50 PB úložné kapacity pro výzkumná data (uvádíme využitelný objem pro výzkumná data, nikoliv raw kapacitu úložišť) s vysokou úrovní zabezpečení proti ztrátě i výpadku přístupu. Plné kapacity bude dosaženo v roce 2028 (podcíl KA2.1). Již v roce 2025 (Q3/2025) bude k dispozici 18 PB ve dvou geograficky odlišných lokacích. Cíle bude dosaženo v rámci aktivity KA2, koordinované sdružením CESNET, další přímo zapojení partneři jsou UK, SSČ AV ČR, MU, VŠB-TUO. Realizací tohoto cíle vznikne v ČR integrované zázemí pro ukládání FAIR výzkumných dat.

Nasazení a pilotní provoz jednotlivých repozitářových systémů: V rámci projektu (aktivita KA2) budou od roku Q4/2025 (po dosažení prvního podcíle KA2.1) v pilotním provozu repozitářové systémy CESNET Invenio (podcíl KA2.2, odpovídá CESNET), CLARIN-Dspace (podcíl KA2.3, odpovídá UK) a ASEP/ARL (podcíl 2.4, odpovídá SSČ a KNAV AV ČR). Do realizace jsou dále zapojeni partneři MU a VŠB-TUO. Realizací tohoto cíle bude mít výzkumná komunita k dispozici tři repozitářové systémy, v nichž bude možné vytvářet tematické, specializované i další repozitáře podle dlouhodobých potřeb výzkumných komunit i institucí. Všechny repozitářové systémy budou sdílet celou úložnou kapacitu. Výzkumná komunita tak nebude nucena sama přímo investovat do technického zázemí ani do specializovaného know-how, nezbytného pro provoz repozitářových systémů, a bude se moci plně soustředit na odborné kompetence a péči o repozitáře a data v nich uložená.

Catch-all repozitář (podcíl KA3.1, odpovídá CESNET): Bude vytvořen a provozován catch-all repozitář pro FAIR data, pro něž v daném okamžiku ještě nebude k dispozici vhodný oborově specifický či podobný specializovaný repozitář. Repozitář bude k dispozici od Q4/2025 a bude vytvořen v rámci KA3 s přispěním partnerů NTK, UK a MU zejména v metodické oblasti. Repozitář bude dále vyvíjen do konce podaktivity. Jedním z cílů catch-all repozitáře je poskytnout i výzkumníkům bez existujících oborových či jinak specializovaných repozitářů (zpravidla patřícím k tzv. long-tail vědeckých pracovníků) odborně spravované prostředí pro jejich data.

Systém řízení přístupu k repozitářům a uloženým FAIR datům (podcíl KA4.3, odpovídá MU ve spolupráci se sdružením CESNET): Návrh, implementace a pilotní provoz systému řízení přístupu k repozitářům, službám i datům uloženým v rámci NRP. Systém bude propojen na společné autentizační zázemí e-INFRA CZ a umožní podrobné řízení přístupu tak, aby jak služby, tak uživatelská data mohly být plně pod kontrolou odpovědných správců. Realizace proběhne v rámci KA4. Podobně jako repozitářové platformy, i zde je hlavním cílem zbavit výzkumné instituce, týmy i výzkumníky zátěže v podobě správy vlastních systémů autentizace (např. registrace uživatelů) a zejména řízení přístupu (včetně potřeby dodržovat nezbytné kyberbezpečnostní i GDPR standardy v kontextu péče o data registrovaných uživatelů). Výzkumná komunita (správci repozitářů i datových sad) tak dostane nástroje

k detailnímu nastavení přístupových práv, přitom ale ponese i nadále plnou odpovědnost za rozhodnutí o tom komu udělit přístup.

Systém návazných služeb a nástrojů repozitářové platformy (podcíl KA5, odpovídá MU, CESNET, UK, ÚFCH JH, ve spolupráci s dalšími partnery projektu): konkrétní nástroje, služby a další komponenty jsou podrobně včetně časového harmonogramu popsány v klíčových aktivitách KA4 a KA5. Cílem je poskytnout kvalitativní zlepšení výzkumného prostředí z hlediska možností práce s daty a plnění FAIR principů. Mezi klíčové služby patří nástroje pro práci s metadaty, a to včetně práce s oborově specifickými metadatovými formáty a práce s licencemi, včetně vazby na řízení přístupu (obé realizováno v rámci KA4). Rozšířené služby jsou pak zacílené zejména na nadstandardní podporu dobré praxe a plnění FAIR principů pomocí řádného a efektivního plánování správy dat, podpory automatizace a strojového zpracování dat, metadat s integrací do služeb technologického zázemí NRP a jejích repozitářových systémů. Neméně důležitým cílem je umožnit přímé napojení výzkumných přístrojů pro sběr dat i metadat, stejně jako zpřístupnění dat pro jejich výpočetní zpracování v národní i mezinárodní výzkumné infrastruktuře (všechny tyto cíle budou dosaženy v rámci KA5). První verze jednotlivých nástrojů a služeb budou k dispozici do konce roku 2025, další verze pak budou podle potřeby i zpětné vazby uživatelů. Všechny návazné služby a nástroje budou plně funkční a integrované ve finálních verzích v roce 2028.

Mapování potřeb a míry uspokojení vědecké komunity v souvislosti s péčí o výzkumná data (podcíl KA6.3, odpovídá MU): V rámci KA6 proběhnou postupně tři velké plošné průzkumy vztahu výzkumné komunity k péči o výzkumná data a způsobech, jakými tuto péči zajišťují. Jednotlivá šetření proběhnou v roce 2024, 2026 a poslední na přelomu 2027/2028 tak, aby výsledky zpracování byly k dispozici ještě v průběhu realizace projektu. Bude oslovena co nejširší vědecká komunita (externí spolupráce s Národním centrem SYRI), výsledky šetření budou zveřejněny a současně využity pro hodnocení dopadů realizace projektu NRP pro výzkumná data i případných dalších projektů podpory péče o výzkumná data na způsob a kvalitu práce výzkumných komunit v ČR.

Zajištění kyberbezpečnosti NRP (podcíl KA6.1, odpovídá CESNET ve spolupráci s MU): Jedná se o komplexní cíl, jehož složkami bude centrální logování a zpracování dat z kyberbezpečnostní telemetrie všech komponent NRP, dále procesní a metodická úprava podmínek provozní bezpečnosti za využití proaktivních a reaktivních opatření, pravidelné penetrační testování NRP a jejích služeb. Základní funkční kyberbezpečnostní zázemí bude ustaveno do instalace technického zázemí NRP, tedy nejpozději Q4/2025. Zajištění centrálního log managementu bude k dispozici od Q4/2025 (a dále se bude průběžně a komplexně doplňovat). Tento cíl je klíčový pro dosažení nezbytné úrovně důvěry v celé prostředí NRP a s tím spojené ochoty skutečně vkládat do tohoto prostředí výzkumná data, která mohou často představovat významnou součást intelektuálního majetku institucí či výzkumných týmů.

Realizace školení a osvěta, podpora pro získávání a prohlubování znalostí a dovedností nezbytných pro optimální využívání infrastruktury NRP (podcíl KA7, odpovídá UK ve spolupráci s MU a se zapojením všech partnerů, kteří vytváří služby, nástroje či další komponenty NRP): V rámci KA7 bude ke každé komponentě, nástroji a službě, která bude vytvořena a provozována v rámci projektu a bude dostupná uživatelům, vytvořen odpovídající metodický a školicí materiál a rovněž systém metodické podpory využití komponent, nástrojů a služeb. Metodické postupy, materiály a školicí materiály budou vytvářeny souběžně s implementací a prvním nasazením jednotlivých nástrojů tak, aby byly k dispozici v době jejich zpřístupnění uživatelům. Každý produkt NRP, repozitář i softwarová služba bude mít zázemí v návodech a návrzích workflow pro jejich užití. Tyto návody budou uživatelsky srozumitelné a budou využitelné jednotlivými institucemi přímo pro výzkumné pracovníky nebo jiné relevantní odborníky. Kvalita a relevance těchto materiálů a návodů bude komunikována v komunitě Data Stewardů. Data Stewardi (KA7.2) mají kritickou roli v rámci komplexního stabilního zázemí péče o výzkumná data jako odborníci, kteří pomohou vytvořit prostředí pro péči o výzkumná data na

jednotlivých institucích, budou úzce spolupracovat s výzkumnými týmy a podpoří využití výsledků projektu NRP na institucích. Komunita Data Stewardů, sdružující osoby na odpovídajících odborných pozicích v rámci vysokých škol, ústavů AV ČR i dalších výzkumných pracovišť bude kontaktním místem i pro další pracovníky datové podpory, kterým bude poskytována podpora pro realizaci datové politiky jejich institucí. Metodické materiály (KA7.1) nebudou pouze zaměřené na technické aspekty ovládní, ale budou reagovat i na potřebu legislativní a právní problematiky. V této oblasti budou poskytovány i právní a etické konzultace, které budou řešit otázky, které výzkumné instituce řeší v souvislosti s principy FAIR data. Cílem projektu NRP v oblasti vzdělávání a osvěty je tedy to, aby žádný výzkumný pracovník neřešil oblast výzkumných dat jako novou, nechtěnou a zatěžující, ale jako integrální součást své výzkumné práce a zároveň měl k dispozici technické i metodické řešení, která bude moci efektivně využít.

6. ORGANIZAČNÍ STRUKTURA, ŘÍZENÍ PROJEKTU A POPIS ROLÍ REALIZAČNÍHO TÝMU

6.1 ORGANIZAČNÍ STRUKTURA A ŘÍZENÍ PROJEKTU

Organizační struktura projektu je uvedena níže, přičemž efektivní řízení projektu zajišťuje splnění úkolů projektu, řešení případných problémů a dosažení stanovených cílů. Důraz je kladen na bezproblémový průběh, který bude plně zajišťovat níže zobrazená struktura.

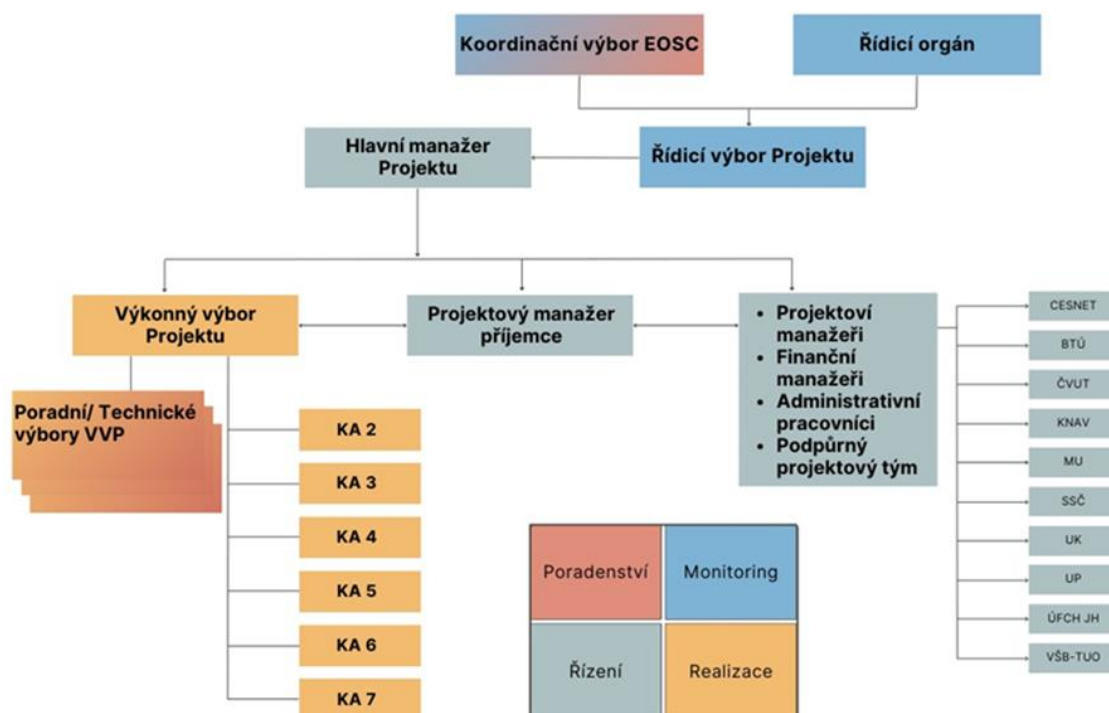


Schéma 2 Organizační struktura projektu

7. KLÍČOVÉ AKTIVITY

7.1 KA1 – POVINNÁ AKTIVITA „ŘÍZENÍ PROJEKTU“

7.2 KA 2 – „VÝVOJ A PROVOZ NÁRODNÍ REPOZITÁŘOVÉ PLATFORMY“

Cílem klíčové aktivity je vybudování a provoz Národní repozitářové platformy. To zahrnuje zejména všechny její vrstvy:

- hardware (servery pro provoz aplikací a pro ukládání dat),
- včetně operačních systémů (Linux),
- úložiště pro vlastní uživatelská data ukládaná do repozitářů a data z nezbytných pomocných systémů (Ceph),
- prostředí pro provoz systémů repozitářů, tedy servery pro běh softwarových řešení repozitářů (Kubernetes),
- a vlastní implementace softwarových repozitářových systémů (zejména CESNET Invenio, CLARIN DSpace a ASEP ARL), které umožní uživatelským skupinám vytvářet repozitáře jako službu.

Integrální součástí této klíčové aktivity je i vývoj ve všech softwarových vrstvách platformy, a to jak za účelem efektivní správy a provozu instalovaných systémů, tak i vývoje vlastních repozitářových systémů. U repozitářových systémů se vývoj v této klíčové aktivitě věnuje zejména jejich základním funkcionalitám, které vyžadují zásahy do jejich jádra, a tedy detailní znalost jejich vnitřního fungování. Přitom aktivita postupuje v souladu a koordinaci s hlavními komunitami, které na vývoji těchto nástrojů pracují. Vývoj na těchto vrstvách zahrnuje standardní inovaci těchto softwarových balíčků včetně integrace jejich nových mainstreamových verzí, dále vývoj funkcí, které jsou nezbytné pro činnost platformy jako celku nebo umožňují budování nadstavbových služeb, které jsou předmětem klíčových aktivit 4 a 5, a v neposlední řadě rovněž pokrytí potřeb uživatelských skupin provozujících repozitáře.

Výstupy této aktivity mají samozřejmě **dopad na širokou uživatelskou komunitu**, která používá repozitáře jako službu. Z procesního a organizačního pohledu nicméně výstupy KA2 primárně slouží následujícím skupinám:

1. správcům repozitářů provozovaných v platformě, zejména
 - a. repozitářů vznikajících na žádost odborných pracovních skupin pro implementaci EOSC v ČR,
 - b. libovolných dalších repozitářů, které budou v souladu s politikou pro jejich vytváření, požadovaných externími uživatelskými skupinami,
 - c. repozitářů provozovaných jako pilotní projekty (KA3), které z pohledu budování infrastruktury slouží jako ověřovací prostředí,
2. správcům a uživatelům testovacích a demonstračních repozitářů vytvořených pro potřeby odborných školení,
3. skupinám podílejícím se na vývoji komponent NRP zejména v rámci klíčových aktivit 4 a 5.

Všem výše uvedeným skupinám je také poskytována pokročilá uživatelská podpora (Level 3), zahrnující oblasti jako pomoc se zakládáním nových instancí repozitářů v platformě a řešení problémů přesahujících běžnou konfiguraci (zejména týkajících se rozvíjených vlastností, případně rovněž oprav chyb v systémech repozitářové platformy). Aktivita rovněž vytváří technickou dokumentaci

budovaných a provozovaných služeb jak pro vlastní potřebu (tedy pokrytí procesů pro provoz systémů), tak pro potřebu správců repozitářů.

Pro KA2 jakožto poskytovatele infrastrukturní služby je zásadním partnerem správce repozitáře, který reprezentuje uživatelskou komunitu a typicky přichází z prostředí mimo projekt. Správce repozitáře rozhoduje o politikách fungování své instance repozitáře, postupech pro ukládání dat a řízení jejich životního cyklu v repozitáři, ustanovuje kurátory dat a rozhoduje o pravidlech přístupu uživatelů a jejich přiřazení k oprávněním v repozitářích. Správce repozitáře zajišťuje uživatelskou podporu pro koncové uživatele svého repozitáře a provádí základní nastavení instance repozitáře.

Pravidla pro poskytování instancí repozitářů, jejich životního cyklu, role správce repozitáře s vymezením odpovědnosti, určení rozsahu dokumentace, kterou správce repozitáře má k příslušné instanci repozitáře vytvořit, a další nezbytné podmínky pro jeho provozování v platformě budou součástí dokumentace pro správce repozitáře, která vznikne v rámci této klíčové aktivity.

Základní hardwarová architektura repozitářové platformy je postavena na distribuovaném systému datových úložišť a clusterů pro běh vlastních repozitářových systémů v prostředí Kubernetes. Úložiště budou postavena zejména na softwarově definovaném úložném systému Ceph a primárním rozhraním bude protokol S3. Úložná kapacita bude sloužit jednak pro vlastní data v repozitářích (datasety), jednak pro pomocná data nezbytná k bezpečnému provozu NRP, jako jsou například zálohy databází repozitářových nástrojů. Do datového úložiště bude implementována funkcionality periodických kontrol fixity (binární neměnnosti) dat a konfigurovatelných geografických replikací ukládaných dat pro dosažení vysoké úrovně odolnosti proti ztrátě nebo binárnímu poškození dat.

Pro běh repozitářů (jejich rozhraní a aplikační logiky) bude vybudována soustava clusterů pro provozování kontejnerizovaných aplikací v prostředí Kubernetes, které budou kolokovány s datovými úložišti v místech s kvalitním připojením do pátevní počítačové sítě CESNET². Prostor Kubernetes bude kapacitně připraveno i pro provozování integračních nástrojů z KA4 (nástroje pro podporu práce s licencemi, nástroje pro řízení přístupu k datům) a nových služeb vytvářených v rámci KA5 (integrace workflows, podpora FAIRifikace výzkumných dat).

Technická realizace systému Kubernetes bude stát na nástrojích s otevřeným zdrojovým kódem, jako je např. RKE2. Systém bude integrovat komponenty pro jeho snadné použití, jako je automatické zakládání doménových jmen, vydávání SSL certifikátů, integrace s AAI, ale zároveň budou k dispozici operátory usnadňující provozování některých hotových aplikací, jako například Postgres databáze. Systémy budou nastaveny a provozovány v režimu vysoké dostupnosti tak, aby výpadek jedné lokality nezpůsobil nedostupnost provozovaných služeb, zde bude využito dynamické routování.

O umístění zařízení pojednává kap. 12 Studie proveditelnosti. Část serveroven bude připravena i pro ukládání citlivých dat (do úrovně medicínských dokumentací včetně).

Citlivá data ukládaná systémem Ceph budou již na úrovni vlastního repozitáře silně šifrována, což pak umožní šifrovaná citlivá data geograficky distribuovat na úložiště v celé repozitářové platformě. Přístup k nim bude nicméně možný pouze z aplikačního clusteru provozovaného v silně zabezpečeném prostředí. S ohledem na vývoj, jak v oblasti nástrojů pro práci s daty tohoto typu, tak i legislativy, se bude prostředí pro citlivá data vyvíjet. Řešení implementačních detailů bude nezbytnou součástí projektu a vznikající technické architektury systému.

² <https://www.cesnet.cz/sit-cesnet3>

Aktivita KA2 zajišťuje **přípravu technické dokumentace pro veřejné zakázky** na pořizování hardware a podílí se na jejich vyhodnocování a uvádění do provozu, což zahrnuje činnosti od koordinace lokálních podmínek umístění zařízení, přes vlastní dodávku, provedení akceptačních testů, instalaci systémů a zahoření až po převedení do plného provozu.

Při zprovoznění Národní repozitářové platformy budou k dispozici tři **softwarové repozitářové systémy**, které byly identifikovány v rámci pracovních skupin EOSC CZ. Pokrývají širokou oblast zázemí již provozovaných repozitářů a řešitelé projektu s nimi již mají dlouhodobé zkušenosti. Konkrétně pro projekt zprovozní a poskytne CESNET InvenioRDM, CLARIN-DSpace a ASEP ARL. Každý zvolený repozitářový systém má mezinárodní nasazení a podporu. Každý využívá odlišný přístup k modelování a uchování metadat, je vytvořen nad jiným technologickým stackem a udržován primárně svou vlastní uživatelskou a vývojářskou komunitou a má své dominantní silné stránky z hlediska použití. Souběžně nasazení tří systémů považujeme za optimální z pohledu odstranění rizika monopolizace, garance realizovatelnosti požadavků různých odborných komunit a skupin na vytvoření repozitářů v rámci NRP, robustnosti vůči ztrátě mezinárodní podpory vývoje konkrétního repozitářového systému, resp. zájmu uživatelů o ni. Současně si realizace tří souběžných repozitářových systémů vynutí dostatečnou flexibilitu na nižších úrovních NRP stacku, což otevírá prostor pro případné zavedení dalšího, dnes ještě neidentifikovaného (případně dokonce neexistujícího) repozitářového systému (v rámci KA3.5 je podpořena instalace systému Islandora s cílem ověřit flexibilitu NRP a možnost rozšíření o další repozitářové systémy). Vybrané systémy jsou v projektu zaštitěny skupinami, které je dlouhodobě rozvíjejí a používají.

Invenio Framework/InvenioRDM je repozitářová technologie vyvíjená v CERNu (Evropská organizace pro jaderný výzkum). Technologie je vysoce škálovatelná, je na ní provozován např. repozitář Zenodo³ s více než 3 miliony záznamů. Popisy datasetů jsou založeny na hierarchickém modelu, umožňují tak popsat širokou množinu metadat. V rámci České republiky kromě institucionálních repozitářů publikací systém Invenio využívá například Národní technická knihovna v rámci vývoje Národního repozitáře pro výstupy textové povahy.

CLARIN-DSpace je ucelené softwarové řešení vyvíjené jako rozšířená a vylepšená verze DSpace, který je světově nejpoužívanějším softwarovým repozitářovým systémem. CLARIN-DSpace tvoří propojený celek vč. databáze, metadat, frontendu, AAI řešení, podpory OAI-PMH aj., se základní podporou Amazon S3 a s integrovaným licenčním frameworkem, který podporuje široké spektrum licencí od zcela otevřených po omezené vyžadující pro přístup k datasetu podpis licence.

Softwarové řešení ASEP-ARL slouží jako repozitář pro evidenci vědeckých výsledků (publikace, aplikované výsledky, ...) a pro ukládání vědeckých dat založený na mezinárodních knihovnických standardech z hlediska struktury popisu (MARC, MODS) i z hlediska implementovaných komunikačních protokolů (Z39.50 a OAI-PMH). Systém vychází z knihovního systému ARL, který je vyvíjen lokální firmou s působností v Česku a na Slovensku. Postupný vývoj, ve spolupráci s Knihovnou AV ČR, vedl nejprve k funkčnosti pro evidenci vědeckých výsledků (publikací, patentů, aplikovaných výsledků,...) kompatibilní s formáty národního CRIS systému RIV, následně k zajištění funkcí repozitáře pro ukládání těchto výsledků a posledním významným rozšířením byla implementace funkcí pro ukládání dat. Nově vytvořené řešení je pod označením ASEP-ARL uvolněno jako open source řešení a je vhodné např. v těch případech, kdy je zájem provozovatele o užití propojení vědeckých dat a publikačních výsledků a o možnost vykazování v rámci národního systému pro evidenci a hodnocení vědy.

³ <https://zenodo.org/>

KA2 bude podporovat provoz i jiných systémů schopných ukládat data do prostředí S3 a běžet v prostředí Kubernetes. Tyto systémy musí koncepčně naplňovat znaky repozitářů (tedy systémů s evidencí metadat, které jsou schopny ukládat citovatelné záznamy), případně pomocných komponent pro běh takových systémů (např. databázových strojů nebo obrazových serverů). Konkrétní příklad takového systému je podpořen v rámci KA3.3.

Provozované prostředí bude sloužit rovněž pro vývoj, testování a provoz komponent vyvíjených v klíčových aktivitách KA4 a KA5, které budou často využívat API poskytované repozitářovými nástroji. Vývojový tým jádra repozitářových systémů bude technologicky podporovat vývoj v KA4 a KA5, zejména tvorbou komponent jádra systémů, které budou nezbytné pro podporu funkcí vyvíjených v KA4 a KA5, dále pak konzultacemi jak v otázkách architektury vyvíjených nástrojů, tak i při řešení konkrétních problémů při vývoji.

Aktivita využívá výstupů komplementárního projektu IPs EOSC-CZ. Z prostředí Národní repozitářové platformy budou zejména exportována metadata do Národního metadatového adresáře, repozitáře budou napojeny na prostředí přenosu velkých objemů dat (využitých mimo jiné i v rámci KA5.3), budou použity komponenty pro přidělování perzistentních identifikátorů a dodržena doporučení a standardy pro monitoring infrastruktury a pro autentizační nástroje.

7.2.1 KA 2.1 VÝVOJ A PROVOZ TECHNICKÉHO ZÁKLADU REPOZITÁŘOVÉ PLATFORMY

Aktivita zajišťuje pořízení hardwaru pro ukládání dat a provoz repozitářové platformy, instalaci a provoz základních systémů platformy na úrovních od správy hardware, přes provozované operační systémy po provoz úložného systému Ceph včetně služby pro zajištění fixity dat, a dále běh kontejnerových aplikací Kubernetes. Součástí zajištění provozu je rovněž sledování telemetrie provozovaných systémů, jejich provozního monitoringu a bezpečnosti (ve spolupráci s KA6 a shodě s požadavky na bezpečnost prostředí i dat).

Aktivita poskytuje infrastrukturu S3 úložišť a Kubernetes jako službu jak následujícím aktivitám, tak i přímo dalším (i externím) implementátorům repozitářových systémů v platformě (příkladem je pilotní provoz repozitářového systému Islandora v rámci KA3.5).

7.2.2 KA 2.2 VÝVOJ A PROVOZ SYSTÉMU INVENIO

Pro systém CESNET Invenio proběhne integrace do základního prostředí NRP (prostředí pro běh systémů a datového úložiště Ceph), vyvíjeného v rámci KA2.1, a současně napojení na průřezové systémy, vyvíjené v rámci dalších aktivit projektu (zejména KA4 a KA5) i v rámci projektu EOSC-CZ (např. export dat do NMA či velkoobjemové přenosy dat). Bude vyvinuta a nasazena služba pro ukládání velkých dat, služby pro transformaci dat (jejich šifrování/dešifrování pro ukládání citlivých dat, verifikace formátů, transformace pro účely diseminace). V oblasti metadat budou nasazeny služby pro jejich pokročilé zadávání a vizualizaci. Ve spolupráci s KA5.3 bude precizován model automatického vytváření záznamů s možností přístupu komunity k ještě nefinalizovaným záznamům. Pro zjednodušení strojového přístupu k záznamům dojde k rozšíření možností importu/exportu metadat a datasetů.

V rámci aktivity budou podporovány také pilotní repozitáře (v testovací, staging i produkční fázi jejich životního cyklu), včetně nezbytných aktivit k udržení kvality jejich provozu (např. exportu monitorovacích informací) a zálohování. Součástí bude rovněž sběr auditních záznamů a statistických údajů z provozu systémů.

7.2.3 KA 2.3 VÝVOJ A PROVOZ SYSTÉMU CLARIN-DSPACE

Rozvoj systému CLARIN-Dspace se v rámci NRP zaměří hlavně na integraci NRP S3 úložiště (v systému Ceph) vč. podpory jeho specifických možností, zlepšení nabídky licencí, napojení repozitářů na e-INFRA CZ/EOSC AAI, integraci s DMP, zlepšení statistických a analytických funkcí jak pro správce repozitáře i pro uživatele nahrávajícího do repozitáře datasety, konfiguraci metadatových schémat a na workflow pro publikaci datasetů z dat streamovaných přímo z přístrojů. Současně budou integrovány nové funkce, které se objeví v hlavní komunitní vývojové verzi projektu DSpace.

7.2.4 KA 2.4 VÝVOJ A PROVOZ SYSTÉMU ASEP/ARL

Rozvoj ASEP-ARL bude cílen na implementaci S3 pro připojení fyzického datového úložiště, provozování na platformě Kubernetes, integraci stávajícího AAI systému s e-INFRA CZ/EOSC AAI, automatizaci propojení s ORCID a dalšími PID a na plnění požadavků plynoucích z klíčových aktivit KA4 a KA5 a KA6 EOSC-CZ/NRP zaměřených na poskytování služeb. Součástí rozvoje budou statistický a analytický modul, integrace s DMP systémem, exportní funkce do NMA, konfigurace metadatových schémat a implementace workflow pro publikaci datasetů z různých zdrojů.

Implementace nových funkcí systému ASEP-ARL bude kromě testovacího prostředí ověřována také průběžným nasazením v institucionálním repozitáři Akademie věd České republiky, kde je možné získat zpětnou vazbu od jejích 53 vědeckých ústavů.

7.3 KA 3 – „PILOTNÍ REPOZITÁŘE“

Klíčová aktivita Pilotní repozitáře má dva souběžné cíle: (i) Vytvoření catch-all repozitáře, který slouží pro ukládání dat a datových sad, pro které není k dispozici vhodnější tematický či jiný specializovaný repozitář. (ii) Přinést do projektu přímou a včasnou zpětnou vazbu od prvních uživatelských skupin Národní repozitářové platformy.

KA3 zajistí interakci se správci a vývojáři vlastní infrastruktury (interakce s KA2) i služeb (interakce s KA4 a KA5), ale rovněž s KA6 ohledně kyberbezpečnostních a uživatelských aspektů poskytovaného prostředí NRP, kterým přináší bezprostřední a nefiltrované zkušenosti z vytváření a přenosu repozitářů do prostředí NRP. Od pracovníků v KA3 to také vyžaduje “snášet nepohodlí” spojené s používáním infrastruktury v rané fázi jejího vzniku, kdy všechny funkcionality a závislosti ještě nejsou plně vyladěny. Vstup z KA3 pak má zejména směřovat ke stabilizaci NRP jako celku, finalizaci procesů jejího fungování jak organizačního, tak technického. Tím bude významně přispívat ke zvýšení kvality platformy jako celku.

Takové partnerství je výhodné oboustranně. Reprezentanti repozitářů zapojených v KA3, vyjma příspěvku na přenos repozitáře do prostředí NRP, získávají přímý přístup ke skupinám pracujícím na vývoji funkcí platformy a mají tak přímočarou možnost ovlivňovat priority jejího rozvoje v přímém kontaktu s těmi, kdo pracují na realizaci.

Pro výběr pilotních repozitářů byla hlavním kritériem skutečnost, že se jedná o repozitáře s jasnou podporou uživatelských komunit, které samy o sobě nejsou závislé na projektu NRP, mají tedy samostatné financování a jejich uživatelská komunita má dlouhodobou představu o jejich potřebnosti a stabilním provozu, který by zajišťovala nezávisle na existenci projektu NRP. Zároveň se jedná o repozitáře ve vědecky excelentních oblastech. V případě KA3.5 je pak součástí pilotního repozitáře i přenos celého nového repozitářového systému Islandora do prostředí NRP.

7.3.1 KA 3.1 CATCH-ALL REPOZITÁŘ

Catch-all datový repozitář bude sloužit jako instance poslední volby pro ukládání datových sad. Předpokládáme, že bude využíván zejména méně organizovanými odbornými komunitami, které nemají vlastní oborové repozitáře ("long-tail věda"). Standardní doporučení pro volbu repozitářů samozřejmě zásadně preferují použití etablovaných oborových repozitářů, které jsou navíc obvykle nastaveny pro specifická oborová metadata. Mezi dalšími možnostmi se v různém pořadí (dle konkrétního autora doporučení) uvádějí institucionální repozitáře (zejména mají-li formulovanou dlouhodobou politiku ukládání dat) a catch-all repozitáře (jako je například Zenodo provozované CERN).

Účelem catch-all repozitáře je poskytnout systém pro ukládání dat v prostředí národní e-infrastruktury, v jurisdikci České republiky, a vybavený dostatečnou kapacitou pro datové sady. Sekundárním účelem catch-all repozitáře je pokrytí přechodného období, kdy se jednotlivé odborné komunity teprve formují a vyhovující oborový repozitář ještě nemají k dispozici, přesto však potřebují ukládat datové sady do repozitářových systémů.

Motivace vytvořit specializované oborové repozitáře nicméně zůstane zachována. Zásadní nevýhodou catch-all repozitářů je generické metadatové schéma, které musí být dostatečně univerzální pro pokrytí prakticky libovolného vědeckého oboru. To poněkud limituje nalezitelnost datových sad v takovém repozitáři, protože základní metadatový model nedokáže plně reflektovat potřeby daných oborů, a tudíž není možné hledat podle oborově specifických polí. Oborově zaměřené repozitáře mají oproti obecným větší možnosti ve škálovatelnosti vyhledávání podle parametrů nastavených na míru potřebám dané vědní komunity.

KA3.1 tedy vybuduje obecný repozitář pro použití vědeckou komunitou pro obecné použití. Z technického hlediska se bude jednat o standardní instanci repozitáře na repozitářovém softwaru Invenio, kde správci repozitáře budou součástí řešitelského týmu projektu. Pro repozitář bude užit základní metadatový model, který je plánovaným výstupem IPs CARDS při NTK. Rovněž se řešitelský tým bude koordinovat s Centrem repozitářů a správy metadat NTK, které v rámci komplementárního projektu IPs CARDS poskytuje metodickou podporou práce s metadaty výzkumných dat.

Catch-all datový repozitář bude vybaven jednoduchým řízením přístupu uživatelů. Předpokládáme, že bude obsahovat datové záznamy ve stavu rozpracováno a zveřejněno, a že vkladatel záznamu sám rozhoduje o zveřejnění záznamu, v důsledku tedy odpovídá za správnost zadaných údajů. Vkladatelem budou moci být standardní uživatelé e-infrastruktury. Z důvodu ochrany kapacit NRP předpokládáme, že běžná oprávnění uživatele po registraci budou umožňovat vkládat záznamy relativně malých datových objemů. Právo vkládat větší datové sady bude vyžadovat schválení administrátorem systému na žádost uživatele.

Metadata catch-all repozitáře budou standardně propagována do Národního metadatového adresáře budovaného v projektu IPs EOSC-CZ. Repozitář bude plně integrován do prostředí Národní datové infrastruktury.

Pro podporu catch-all repozitáře jsou plánovány pracovní pozice správce systému, který bude primárně odpovědný za konfiguraci, provoz a rozvoj této instance repozitáře, a dále pozice pro uživatelskou podporu. Zde se jedná o přímou podporu koncových uživatelů a libovolných problémů s používáním systému.

7.3.2 KA 3.2 PILOTNÍ REPOZITÁŘ PRO BIOLOGICKÁ ZOBRAZOVACÍ DATA

Biologické zobrazování je nedílnou součástí nejenom biomedicínského výzkumu, ale i dalších příbuzných oborů, jako je vývoj bio- a bionano-materiálů nebo environmentální vědy. V České republice je dostupnost špičkových zobrazovacích technologií primárně zabezpečena velkou výzkumnou infrastrukturou CzechBioImaging (CzBI). Ve spolupráci s CESNETem se CzBI již několik let snaží zlepšovat svou infrastrukturu pro ukládání a přenos dat. Na podnět International Advisory Board, která provádí pravidelné hodnocení infrastruktury, se CzBI rozhodlo vybudovat vlastní licencovaný repozitář pro biologická zobrazovací data. V současnosti existuje mezinárodní archiv biologických zobrazovacích dat "BioImage Archive", který se zaměřuje výhradně na data k významným publikacím. Ty se do repozitáře BioImage Archive posílají manuálně cestou přes webové rozhraní. Repozitář CzBI – Národní úložiště pro biologická zobrazovací data – využije možnosti NRP a poskytovaných služeb pro práci nejen s vybranými daty, ale pro komplexní práci s širokým spektrem výstupů práce v infrastruktuře (včetně tzv. negativních výsledků) a jejich sdílení (viz níže). Tento repozitář také umožní automatický sběr dat a metadat z mikroskopů (ve spolupráci s KA 5.3). Existence CzBI repozitáře v NRP přispěje k naplnění záměrů definovaných v dokončované politice CzBI pro správu dat. Tato politika definuje doporučené postupy pro uživatele CzBI. Repozitář bude primárním místem pro práci s FAIR daty v prostředí infrastruktury CzBI.

CzBI repozitář bude pro své fungování sdílet s repozitářem BioImage Archive standardy a metadatová schémata: OME-TIFF a OME-NGFF/ZARR formáty souborů⁴, metadatová schémata REMBI a navazující NBO-Q⁵ nebo OME-OWL ontologii pro světelnou mikroskopii⁶. Ty byly definovány v rámci globálních aktivit QUAREP-LiMi a Open Microscopy Environment, které se také podílejí na jejich rozvoji. Členové CzBI jsou zapojeni do těchto aktivit.

CzBI repozitář pro biologické zobrazování představuje excelentní prostředí pro testování práce s různými typy dat, jako i komplexních funkcionalit NRP. Očekáváme celkem kolem 0,5 PB FAIR dat do konce roku 2025, 4 PB pak do konce 2028. Typickým výstupem měření v oblasti biologického zobrazování může být velké množství malých souborů, které je potřeba přenášet do úložiště napojeného na repozitář v krátkých časových úsecích. Například: vznik >1000 souborů, každý s velikostí 2–50 MB a frekvencí 5–200 souborů za sekundu. Budeme testovat vytěžování dat v různých frekvencích. Pro zachování neměnnosti dat je potřeba tyto procesy maximálně urychlit. Moderní mikroskopy však stále častěji generují velké soubory (> 50 GB) v kontejnerových formátech (OME-ZARR). Dostupnost takovýchto dat pro, kupříkladu, následné zpracování (viz KA5.4) vyžaduje postupy, které zatím v české infrastruktuře nebyly testovány. Experimentální postupy v této doméně také často propojují několik měření na různých přístrojích (KA5.3), které vyžadují zaznamenávání jednotlivých procesů pomocí kombinací několika metadatových schémat a ontologií. Typickým příkladem je kombinování dat ze světelné mikroskopie s daty z průtokové cytometrie nebo elektronové mikroskopie. Otestujeme možnosti vkládání sémanticky kombinovaných dat do jednoho záznamu nebo prolinkovaných záznamů. V některých případech může také dojít k simultánnímu toku dat z více přístrojů do jednoho záznamu. Pro tyto případy budeme sledovat správnost přidělování a harmonizaci metadat záznamů. Dále zde ověříme automatické přidělování vybraných identifikátorů (IPs EOSC-CZ), a přidělení licence (KA4.2). V neposlední řadě bude prověřena funkcionálnost FAIRifikace záznamů (KA5.1) a možnosti napojení správy dat na DMP (KA5.2). Pilotní CzBI repozitář umožní ověření

⁴ doi.org/10.25504/FAIRsharing.cq8tg2, doi.org/10.25504/FAIRsharing.9af712

⁵ <https://doi.org/10.1038/s41592-021-01166-8>, fairsharing.org/4747

⁶ <https://fairsharing.org/350>

robustnosti NRP pro víceúrovňový vstup FAIR dat a metadat od koncových uživatelů, ale i pro kontrolu vstupů a výstupů větším počtem uživatelů.

Výstupem této aktivity bude plně funkční repozitář pro biologická zobrazovací data, který bude maximálně využívat komponenty a funkce NRP pro komplexní správu dat a metadat a to včetně automatických procesů, z kterých některé budou v daném oboru unikátní. Vedlejším výstupem bude i návod s doporučeními pro nastavení budoucích repozitářů pro práci s komplexními experimentálními postupy.

7.3.3 KA 3.3 PILOTNÍ REPOZITÁŘ NÁRODNÍ ÚLOŽIŠTĚ PRO BIODIVERZITNÍ DATA

Uchovávané vědecké doklady rostlin, hub a živočichů jsou neocenitelným základem pro výzkum biodiverzity. Dokladované exempláře organismů jsou nezbytné pro vědecký popis nových druhů, pro určení zeměpisného rozšíření taxonů a změn tohoto rozšíření v čase, ale jsou nenahraditelné také pro související genetický, ekologický a environmentální výzkum, který může vést k novým biotechnologickým aplikacím nebo efektivním řešením v oblasti ochrany přírody. Uchováváním dokladů a vzorků ve stálých a přístupných depozitářích, poskytují herbáře, muzea a genové banky ověřitelný referenční materiál pro budoucí vědecký výzkum, nové studie a publikace a zajišťují reprodukovatelnost vědy.

Herbáře jsou sbírky sušených, vylisovaných nebo jinak konzervovaných rostlin a hub – dokladové položky, které obsahují informace o taxonomické identitě sbíraného organismu, místě a době sběru a často i typu stanoviště a informace o živém organismu. Celkem je v České republice 64 institucionálních herbářových sbírek (z nichž 47 je evidovaných v mezinárodním Index herbariorum), které obsahují asi 10 milionů herbářových položek (2,5 % světových herbářových položek), a to nejen z Česka, ale z celého světa. České herbáře jsou proto významným a mezinárodně uznávaným zdrojem informací o rostlinné rozmanitosti obecně. Téměř 75 % exemplářů je v České republice uchováno v pěti největších sbírkách: Univerzity Karlovy (PRC, oficiální akronym herbářové sbírky dle mezinárodního Index herbariorum⁷), Národního muzea (PR a PRM), Moravského zemského muzea (BRNM), Masarykovy univerzity (BRNU) a Botanického ústavu AV ČR (PRA). V současné době však pouze BRNU a PRC poskytují online přístupná data, tj. metadata a související skeny a fotografie exemplářů ve vysokém rozlišení, prostřednictvím mezinárodního databázového systému JACQ⁸, z něhož je možné data stáhnout, a jsou také předávána do Global Biodiversity Information Facility (GBIF)⁹, největší světové on-line platformy pro data o biodiverzitě. Databázový systém JACQ je založen na používání metadatových standardů pro biodiverzitní data (Darwin Core¹⁰). Aktuálně je takto přístupných téměř 230,000 výskytových dat založených na herbářových položkách (ze stabilními identifikátory) a cca 15,000 skenů nebo fotek těchto položek ve vysokém rozlišení. Oba datasety zpřístupněných položek (BRNU a PRC) jsou hodně citována (spolu přes 700 citací za poslední 4 roky) ve vědecké literatuře, včetně prestižních časopisů typu PNAS nebo Nature.

Odhadovaný roční přírůstek všech uchovávaných sbírek v ČR je cca 20,000–30,000 dokladů. Zhruba tři čtvrtiny těchto položek jsou uloženy ve výše zmíněných největších institucionálních sbírkách. Herbářové instituce v ČR tak představují distribuovaný systém znalostí o biodiverzitě organismů, který však není napojen (s výjimkou BRNU a PRC) na formalizovanou národní síť a národní úložiště dat. Navíc tento systém zatím není propojen s národním a globálním systémem genetických bank, které pracují

⁷ <https://sweetgum.nybg.org/science/ih/>

⁸ <https://www.jacq.org/#database>

⁹ <https://www.gbif.org/>

¹⁰ <https://dwc.tdwg.org/>

s genomickým materiálem spojeným s herbářovými doklady. Cílem tohoto pilotního projektu je proto vytvoření národního úložiště pro biodiverzitní data, který bude vycházet a využívat funkce Národní repozitářové platformy a uchovávat metadata exemplářů/fyzických objektů a souvisejících snímků a genetických vzorků ze zúčastněných institucí v České republice. Takový uzel poskytne nejen snadný a efektivní způsob propojení institucí na národní úrovni, ale také usnadní jejich integraci do mezinárodních iniciativ jako je GBIF, GGBN (Global Genome Biodiversity Network¹¹) nebo EUROPEANA, vše v souladu s principy FAIR přístupů k datům. Tento uzel zajistí také propojení s dalšími herbářovými a genovými sbírkami, především ze střední Evropy, které jsou již zapojeny v JACQu nebo jsou propojeny jinými databázovými systémy. Několik nejvýznamnějších českých institucí (zmiňovaných výše) se zapojí do realizace pilotního projektu. Od samotného počátku to bude koordinátor pilotní části, Univerzita Karlova (PRC), který již v současnosti využívá datová úložiště CESNETu pro své digitalizované doklady. V průběhu řešení se budou připojovat další spolupracující instituce zmíněné výše (BRNM, BRNU, PR, PRA, PRM, BRNU a BRNM), ale také Jihočeské muzeum v Českých Budějovicích (CB). V rámci národní iniciativy EOSC-CZ budou herbáře a související a příbuzné sbírky sloužit jako modelový uzel národního úložiště biologické rozmanitosti, který vytyčí cestu pro systematické uchovávání dat a metadata o biologické rozmanitosti na národní úrovni.

Tento pilotní projekt repozitáře rovněž ukáže schopnost Národní repozitářové platformy efektivně hostit podobné národní uzly stávajících výzkumných datových služeb implementovaných pomocí dalších nástrojů, než jsou tradiční repozitáře, bezpečně uchovávat metadata národních zdrojů biologické rozmanitosti a případně i rozšiřovat jejich funkčnost a pomáhat řešit problémy s procesy implementace nových služeb a integrace s jinými databázemi.

Výstupem tohto pilotního repozitáře bude plně funkční národní repozitář pro herbářové sbírky a genetickou banku, který bude propojen s mezinárodními databázemi, zejména pak s GBIF a GGBN, což umožní vzájemnou výměnu a obohacení znalostí o biodiverzitě na národní i mezinárodní úrovni.

7.3.4 KA 3.4 PILOTNÍ REPOZITÁŘ MOLECULAR BIOPHYSICS DATABASE (MBDB)

Databáze molekulární biofyziky (MBDB) je repozitář, který umožňuje shromažďovat bohatě anotovaná primární měření z biofyzikálních technik. Zpočátku budou shromažďována nezpracovaná měření ze tří různých technik, konkrétně z mikroskopické termoforézy (MST), interferometrie na biologické vrstvě (BLI) a rezonance povrchových plazmonů (SPR). V přípravě jsou další dvě techniky, konkrétně izotermální titrační kalorimetrie (ITC) a hmotnostní fotometrie (MP), cílem je implementovat postupně další techniky v závislosti na tom, jak se bude rozšiřovat potřeba vědecké komunity.

Vývoj byl zahájen ke konci roku 2022 a převážně probíhá ve spolupráci IBT AV ČR a CESNET, nedávno se však připojili i místní partneři z ČVUT v Praze. Na vývoji se podílí i evropští spolupracovníci z akademické i průmyslové sféry, a to prostřednictvím konsorcia MOSBRI, které je projektem standardizace biofyziky financovaným EU a je tvořeno 13 výzkumnými centry excelence a 2 průmyslovými partnery. MBDB je aktuálně vyvíjena jako lokální repozitář; správci mají zájem tento lokální repozitář přesunout do prostředí NRP, jakmile bude k dispozici. MBDB je implementována pomocí frameworku Invenio, a to na základě dlouhodobé spolupráce mezi BTÚ a CESNETem.

Předpokládanými uživateli budou převážně akademičtí výzkumní pracovníci, kteří budou chtít ukládat data například v souvislosti s publikováním vědeckých článků, ale i výrobci přístrojů, kteří mohou chtít ukládat referenční soubory dat pro své techniky, přičemž budou povoleny i negativní výsledky. Uživatelé by v dlouhodobém horizontu mohli pocházet z celého světa, ale zpočátku je budou pravděpodobně využívat převážně evropská centra výzkumných infrastruktur. U většiny technik bude

¹¹ https://data.ggbn.org/ggbn_portal/

velikost jednotlivých záznamů jen zřídka přesahovat 10 MB, avšak u některých technik, včetně hmotnostní fotometrie, bude velikost záznamů 1–10 GB. V letech 2025–2028 je očekávaný počet 100–1000 záznamů ročně (tj. cca 1 TB ročně).

V rámci MBDB byly implementovány větší a složitější datové modely, než které byly dostupné v nativním systému Invenio. To činí indexování a vyhledávání v metadatech větší výzvou, protože vyžaduje rozšíření možností vyhledávání ve vnořených strukturách a také efektivní zpracování vyhledávání polymorfních datových objektů. V projektu podpořený vývoj obecných nástrojů pro zvládnutí složitosti MBDB výrazně sníží bariéru pro budoucí skupiny uživatelů se zájmem o zřízení instancí repozitářů v prostředí Invenio, kteří budou mít podobné požadavky jako MBDB. Takové požadavky se dají očekávat od uživatelů z oblasti věd o živé přírodě. Dalším aspektem se zobecnitelným rozsahem je vývoj schopnosti udržovat místní informace v synchronizaci s externími zdroji. Konkrétně MBDB bude používat řadu externích zdrojů pro anotaci dat. V současné době se používá ROR pro afilience, taxonomie NCBI pro anotaci biologického organismu, projekty OpenAire pro informace o financování, ORCID pro autentizaci (později také obohacení o metadata) a později Pubchem pro chemické druhy. To bylo implementováno pomocí slovníků Invenio. Tyto zdroje však obsahují velké a stále rostoucí množství záznamů (3,2 milionu u největšího z nich), takže je nezbytný vývoj nástrojů pro prohledávání online zdrojů a lokální ukládání pouze těch záznamů, které se používají. Navíc, i když se identifikátory spojené s externími záznamy nezmění, údaje s nimi spojené se mohou v průběhu času měnit (např. přejmenování organismu, sloučení institucí pod novým názvem). Bude třeba implementovat obecnou strategii pro udržování aktuálnosti místních informací a také možnost zjistit, jak se informace jeví v době původního uložení, a pro slovníky Invenio bude třeba vyvinout nástroje, jak toho dosáhnout.

Kromě implementace samotného repozitáře MBDB jako pilotního repozitáře v NRP bude KA3.4 poskytovat pomoc se začleněním takto nabytých zkušeností a expertizy do dokumentace a procesu podpory uživatelů. Konkrétně například zkušenosti ohledně modelování dat, konfigurace stávajících modulů Invenio, začlenění vlastních modulů Invenio, vývoje frontendových komponent a rozvoje workflows. To výrazně napomůže snížit časovou i technickou bariéru při vývoji budoucích instancí repozitářů.

7.3.5 KA 3.5 PILOTNÍ REPOZITÁŘ HUME LAB

HUME Lab – Laboratoř pro experimentální humanitní vědy je infrastruktura provozovaná Filozofickou fakultou Masarykovy univerzity¹², disponující špičkovým přístrojovým vybavením používaným badateli při výzkumu kognice (např. percepce, myšlení), chování a interakcí lidí. Laboratoř a její vybavení je používáno vědci z řady oborů, nejčastěji psychology, kognitivními vědci a religionisty, kteří se organizují ve výzkumných skupinách, či individuálními badateli z antropologie, archeologie, ekonomie, vzdělávání, historie, lingvistiky nebo informační vědy. Řada projektů je interdisciplinární povahy, spolupráce vzniká zvláště s přírodními (např. geografie) a počítačovými vědami. Přístrojové vybavení pro výzkum hojně využívají také studenti při zpracovávání bakalářských, magisterských i dizertačních prací.

Výzkumná komunita si osvojila praxi ukládání výzkumných dat na vlastní záznamová média. V badatelských skupinách postupně sílilo uvědomění si nevyhovujícího stavu této praxe, která znesnadňuje kumulaci výzkumných dat, jejich sdílení a využití pro publikační výstupy v Q1 impaktovaných časopisech, stejně jako spolupráci se spřátelenými badatelskými skupinami

¹² humelab.cz

operujícími na Palackého univerzitě v Olomouci, na Karlově univerzitě v Praze, na spřátelených zahraničních univerzitách i na zahraničních institucích, které využívají ke generování dat stejnou technologii. Výsledkem je započítí prací na přípravě repozitářové platformy s názvem HUME Lab ve fakultní infrastruktuře Digitalia MUNI Arts¹³ s cílem zajistit dlouhodobé a kurátorsky ošetřené uložení výzkumných dat, společně se standardizací workflows a nastavením fází zpracování dat, ve který dochází k transformaci dat z přístrojových vstupů po závěrečné analýzy. V jednotlivých fázích zpracování je nutné ukládat také paradata, která jsou podmínkou důvěryhodnosti měření a opakované využitelnosti dat.

Chystaná repozitářová platforma bude sloužit komunitě výzkumníků se zájmem o doménu měření pohybu očí a rovněž fyziologických reakcí (např. dilatace zornice) pomocí technologie eye-trackingu. Komunita publikující posledních osm let v prestižních zahraničních časopisech má dlouhodobou podporu Filozofické fakulty Masarykovy univerzity, která ze svého rozpočtu finančně zajišťuje fungování infrastruktury HUME Lab. Infrastruktura Digitalia MUNI ARTS je součástí velké výzkumné infrastruktury LINDAT/CLARIAH-CZ, a je také zahrnuta do strategického záměru fakulty. Filozofická fakulta MU se zaručuje udržet její provoz v dalším desetiletí i v případě ukončení projektového financování. Zástupci výzkumné komunity vyjádřili zájem o převedení vznikajícího řešení do prostředí NRP formou pilotního repozitáře.

Shromažďovaná přístrojová a doprovodná data budou v souladu s definovanými pravidly NRP uložena do Ceph úložiště NRP. Data budou rozdělena do kategorií dle typu přístroje, typu podnětů, typu měřených pohybů, tematického zaměření studie. Samotná repozitářová platforma bude vytvořena v systému Islandora – na tomto systému je založena infrastruktura Digitalia MUNI ARTS. V případě pilotního repozitáře HUME Lab bude vytvořena nová samostatná instance systému v prostředí NRP.

7.4 KA 4 – „INTEGRACE SLUŽEB NRP“

Cílem této klíčové aktivity je integrace a provoz základních služeb podporujících práci s výzkumnými daty v NRP. Hlavní účelem je poskytnout nástroje pro plnění esenciálních požadavků na práci s tzv. FAIR výzkumnými daty, bez kterých by celkový provoz NRP nedokázal naplnit svůj účel. Konkrétní služby předkládané k implementaci v této klíčové aktivitě byly identifikovány na základě dlouhodobé práce více než 80 odborníků napříč vědeckými obory sdruženými v otevřené průřezové české pracovní skupině EOSC Základní služby¹⁴. Sběr požadavků na základní služby pro vědeckou komunitu pro práci s výzkumnými daty začal na kick-off meetingu této pracovní skupiny 12. 11. 2021. Od té doby byl seznam postupně rozšiřován a upřesňován jednotlivými členy pracovní skupiny a hromadně revidován na 11 širokých setkáních pracovní skupiny. Výsledkem společné práce a oponentury v rámci pracovní skupiny byl prioritizovaný seznam 35 potřebných služeb rozdělených do 11 kategorií. Každá ze služeb byla hodnocena s ohledem na náročnost její implementace, přínos pro koncové uživatele i důležitost z hlediska národních a mezinárodních trendů a standardů práce s daty. Pro každou službu bylo rovněž posuzováno, kým a jak by měla být v rámci ekosystému EOSC aktivit implementována.

V rámci této klíčové aktivity byly k implementaci vybrány prioritní služby potřebné pro zajištění esenciální práce s výzkumnými daty v NRP. Tj. služby, jejichž cílem je umožnit práci s metadaty dle požadavků jednotlivých repozitářů a oborů a v souladu se základním metadatovým modelem a doporučeními definovanými v komplementárním projektu CARDS, tedy pomoci efektivnímu provázání repozitářů pracujících s více metadatovými standardy; poskytovat správcům repozitářů a

¹³ digitalia.phil.muni.cz

¹⁴ <https://www.eosc.cz/pracovni-skupiny/zakladni-sluzby>

výzkumníkům nástroje pro určování licencí pro výzkumná data v NRP, tedy nepominutelnou podporu pro uživatele, ale i správce repozitářů; a zejména poskytnout nástroje pro precizní řízení přístupu k datům a metadatům v NRP dle požadavků vlastníků dat, a to i s ohledem na potřeby související s problematikou řízení přístupů k citlivým datům.

Tato klíčová aktivita využívá primárně rozhraní, funkcí a služeb jádra NRP, které jsou implementovány v klíčové aktivitě KA2. Implementované služby jsou pak dostupné pro instance repozitářů v NRP, včetně těch implementovaných v klíčové aktivitě KA3.

Popis jednotlivých služeb je podrobněji popsán v následujících podsekcích.

7.4.1 KA 4.1 SPRÁVA METADATOVÝCH PROFILŮ V NRP

Aktivita bude koordinovat práci s metadaty v NRP a je komplementární k projektům IPs CARDS a IPs EOSC-CZ. Aktivita bude vyvíjet a provozovat nástroj pro správu metadatových profilů. Nástroj umožní efektivní správu metadatových schémat (obecně datových specifikací) určených pro popis dat v jednotlivých repozitářích. Jelikož jedno schéma může rozšiřovat jiné, například pro konkrétní podobor, je mezi nimi třeba zajistit konzistentní propagaci změn. Současné přístupy umožňují správu jednotlivých datových specifikací a jejich technických artefaktů. Pro správu sad specifikací a jejich vazeb ve formě souvisejících a navzájem propojených metadatových profilů včetně propagace změn v jejich hierarchiích však budou muset být současné přístupy dále rozvinuty a v průběhu projektu implementovány při současném zohlednění požadavků uživatelů v projektu. Jelikož budou vyvíjené funkcionality relevantní v celé mezinárodní komunitě zabývající se tvorbou webových datových specifikací, jejich vztahů a propagací změn mezi nimi, budou dosažené výsledky prezentovány na relevantních mezinárodních konferencích a také popsány v žurnálových publikacích. Výstupem nástroje budou definice metadatových schémat, které budou registrovány v Národním katalogu repozitářů (výstup IPs CARDS) a které budou podkladem pro implementaci metadatových schémat a rozhraní do konkrétních repozitářů NRP. Aktivita dále ve spolupráci s IPs CARDS vytvoří metodiku pro správu metadatových aplikačních profilů právě pomocí vyvinutého nástroje. Tato aktivita nezahrnuje následnou implementaci takto vytvořených a spravovaných metadatových profilů v jednotlivých konkrétních repozitářích NRP.

Tato aktivita je v projektu NRP financována pouze po dobu prvních tří let realizace, tj. do konce roku 2026. Další podpora této aktivity by již překročila meze univerzálnosti tohoto projektu. Další rozvoj nástroje je třeba plánovat a realizovat s ohledem na oborovou specifičnost, a proto další rozvoj a podpora této aktivity patří spíše do projektu v připravované výzvě Open Science II. Jednání s týmem připravujícím projekt do této výzvy započne bez zbytečného odkladu po zahájení jeho přípravy.

Hlavní výstupy

- Nástroj pro správu metadatových profilů
- Metodika pro správu metadatových profilů

7.4.2 KA 4.2 PODPORA PRÁCE S LICENCEMI

Repozitáře v NRP obecně mohou umožňovat používání různých licencí pro data uživatelů. Tvůrci repozitářů by proto měli mít k dispozici nástroje pro definici licencí, které „jejich“ repozitáře podporují, případně by bylo vhodné nabídnout uživatelům, kteří budou do repozitářů nahrávat data, vodítka ve výběru licence. Na druhé straně je pak třeba zajistit, aby podmínky zvolené licence skutečně byly při

přístupu k repozitářům a v nich uloženým datasetům dodrženy. Vhodné licencování dat a využívání datasetů ve shodě s jejich licencí je důležitým prvkem plánování správy dat, tato aktivita má proto úzkou vazbu také na činnosti KA5.2, kde budou tamní nástroje navázány na podpůrné nástroje práce s licencemi vyvinuté a nasazené v rámci této aktivity.

Tato aktivita zajistí komplexní podporu uživatelům při výběru vhodné licence a při řízení přístupu k datasetům v souladu s jejich licencí. V rámci aktivity vznikne analýza licenčního prostředí od datasetů se zcela individuálním přístupem, přes uživatelem definované licenční smlouvy, které je možno vyplnit a odeslat online, až po datasety s veřejnými licencemi.

Na základě právní analýzy proběhne jednak vytvoření *Průvodce výběrem licence*, který tvůrce datasetu (nebo software) provede výběrem vhodné licence. Následně bude tento průvodce implementován a integrován do modulu pro nahrávání datasetů pro použití v jednotlivých repozitářových platformách.

Další aktivitou je implementace řešení pro strojově zpracovatelnou definici licencí v repozitářové platformě, vč. explicitních atributů pro každou takto zpracovanou licenci, které umožní automatizaci procesů (např. pro plánování správy dat, viz KA5.2) a řízení přístupu k datům, např. příznakem, že licence vyžaduje podpis smlouvy o přístupu k datům apod. Tento licenční modul bude napojen na AAI řešení NRP (viz KA4.3). Pro licence, kde přístupovou smlouvu nelze jednoduše podepsat online při autentizaci uživatele, bude implementován proces podporující individuální žádost o přístup k datasetům.

7.4.3 KA 4.3 ŘÍZENÍ PŘÍSTUPU K DATŮM

Tato aktivita vyvíjí a provozuje komponenty NRP, které souvisí se zajištěním řízení přístupu k repozitářům i datům v nich uloženým. Aktivita vyvine nástroje pro autorizaci a komplexní řízení přístupu k datům ukládaným v jednotlivých repozitářích (včetně řízení přístupu ke správě samotných instancí repozitářů). Řešení bude provázáno s KA4.2 a podporou kontroly licencí a z nich vyplývajících požadavků na řízení přístupu. Řešení bude připravené zprostředkovávat autorizační rozhodnutí poskytované různými systémy a zajistit unifikovaný způsob distribuce autorizačních pravidel do napojených systémů a repozitářů. Pro tento účel budou využita standardizovaná řešení (např. *GA4GH passports* a postupně i eWallets a analogická řešení eIDAS) a v případě potřeby bude přispěno k rozvoji standardů podle konkrétních potřeb uživatelských komunit.

Specifickým způsobem bude ošetřen přístup k citlivým datům. U tohoto typu dat budou aplikovány další kontrolní mechanismy, například vyšší úroveň autentizace nebo vyšší úroveň ověření identity uživatele, a bude zajištěno využití dat v souladu s uděleným informovaným souhlasem. U parametrických citlivých dat bude navržen a implementován systém kontroly přístupu až na úrovni jednotlivého záznamu. Příkladem takového řízeného přístupu k citlivými datům s vysokou granularitou jsou nástroje vyvinuté v rámci infrastruktury EATRIS-CZ, které byly využité i v rámci pandemie COVID-19 (ClinData, CovIT, PreClinData, MedChemBio portál¹⁵).

Kromě spolupráce v rámci KA4 (s KA4.2) je samozřejmě nezbytná spolupráce s KA2 a KA5, ale i v rámci KA6 a zajištění plné kyberbezpečnostní compliance vytvářených a provozovaných řešení. Ve spolupráci s KA6.3, které nám pomůže identifikovat typy uživatelů, nejdůležitější workflow a zapojené komponenty, budeme mapovat i specifické potřeby jednotlivých uživatelských skupin. Po celou dobu realizace předpokládáme rovněž spolupráci na mezinárodní úrovni, zejména v kontextu aktivit souvisejících s celoevropským přístupem k řešení AAI pro EOSC a specificky problematiky řízení

¹⁵ <https://portal.imtm.cz/>

přístupu k datům. Podstatná bude rovněž kontinuální spolupráce s KA6.3 na zvyšování uživatelské přívětivosti vytvářených řešení, protože AAI komponenty jsou zpravidla první, s nimiž uživatel přichází do styku. Aktivním zapojením do národních i mezinárodních aktivit souvisejících s AAI zajistíme kompatibilitu s rozvíjejícími se standardy v akademickém prostředí a případnou adaptaci v mezinárodní komunitě standardů a postupů vzniklých v rámci projektu. Nezbytnou součástí této klíčové aktivity je proto i aktivní účast na mezinárodních akcích a konferencích, kde budeme rovněž prezentovat naše řešení.

Činnost v rámci KA4.3 úzce souvisí a navazuje na aktivity v rámci projektu IPs EOSC-CZ i velké výzkumné infrastruktury e-INFRA CZ. V rámci projektu IPs EOSC-CZ je vytvářena a dále rozvíjena celková koncepce a architektura autentizační a autorizační infrastruktury implementace EOSC v ČR a je podpořena integrace a specifické požadavky na autentizaci a kontrolu identity (identity vetting) uživatelů, včetně propojení na eIDAS. e-INFRA CZ se pak zabývá implementací obecných principů AAI, zejména opět se zaměřením na autentizační mechanismy. Naproti tomu projekt NRP a aktivita KA4.3 se věnuje oblasti autorizace a řízení přístupu, se specifickými řízení přístupu k datovým sadám, resp. jejich součástí (fine grained autorizace). Tyto tři projekty se tak vzájemně doplňují – IPs EOSC-CZ zajišťuje koncepci a architekturu řešení AAI pro EOSC v ČR, e-INFRA CZ zajišťuje implementaci obecných autentizačních mechanismů a obecné zázemí pro jednoduchou autorizaci, zatímco NRP se věnuje specifickým problémům a tvorbě komponent pro fine grained autorizaci v prostředí péče o data. Tímto způsobem je zajištěna jak vysoká synergie, tak i odstranění zbytečných duplicít vývoje a provozu; jednotlivé projekty přímo využívají jinde dosažených výsledků.

7.5 KA 5 – „VÝVOJ A PROVOZ NOVÝCH SLUŽEB NRP“

Tato klíčová aktivita je komplementární ke klíčové aktivitě KA4, aby doplnila portfolio služeb pro podporu FAIR v NRP. Zatímco KA4 se zabývá poskytováním esenciálních služeb, bez kterých by NRP nebyl schopen efektivně plnit svůj účel (viz KA4), tak v této klíčové aktivitě se věnujeme službám, které jsou potřebné pro efektivní podporu FAIR principů na úrovni NRP. V rámci této klíčové aktivity tedy pokrýváme i specifické oblasti, jako je napojení softwarových workflows pro zpracování dat a vědeckých přístrojů, které data produkují nebo zpracovávají a generují, ukládají a publikují doprovodné provenance informace apod. Nedílnou součástí je také podpora pro plánování práce s výzkumnými daty, která se stává soudobým standardním požadavkem *research funding agencies* v ČR i v Evropě, neboť je důležitou pomůckou pro kvalitní zvládnutí managementu dat v současném výzkumu, který stojí na masivním zpracování digitálních dat, čemuž má implementace NRP napomáhat.

Služby předkládané k implementaci v této klíčové aktivitě byly identifikovány a vybrány obdobným způsobem jako služby KA4 (viz KA4).

Tato klíčová aktivita používá primárně rozhraní, funkcí a služeb jádra NRP, které jsou implementovány v klíčové aktivitě KA2, a funkcí a služeb poskytovaných klíčovou aktivitou KA4. Implementované služby jsou pak dostupné pro instance repozitářů v NRP, včetně těch implementovaných v KA3. Při vývoji a provozu se opíráme o procesní podporu poskytovanou klíčovou aktivitou 6. Služby této klíčové aktivity jsou zahrnuty ve vzdělávacích aktivitách klíčové aktivity KA7. Integrální součástí této klíčové aktivity je i vývoj ve všech softwarových vrstvách pro jednotlivé služby, které s jádrem platformy komunikují přes jim poskytovaná rozhraní, a následný provoz těchto služeb. Využito bude i úložných (S3) a výpočetních (Kubernetes) služeb poskytovaných klíčovou aktivitou KA2.

Popis jednotlivých služeb je podrobněji popsán v následujících podsekcích.

7.5.1 KA 5.1 PODPORA FAIRIFIKACE VÝZKUMNÝCH DAT

Cílem podpory FAIRifikace je předat uživatelům informaci o míře plnění principů FAIR pro jednotlivé datasety ukládané v rámci NRP a poskytnout jim zpětnou vazbu a související doporučení. Služba bude ověřovat míru naplňování jednotlivých aspektů FAIR vůči zvoleným kritériím a jejich plnění kontrolovaným datasetem. Bude vytvořen technický rámec, v němž budou k dispozici vybrané nástroje pro kontrolu FAIRness. Způsob integrace bude odpovídat možnostem konkrétního nástroje, zejména z hlediska jeho rozšiřitelnosti. K vyhodnocení naplňování FAIR principů bude docházet jak podle standardních metrik odvozených z externích nástrojů, tak podle interních metrik definovaných specificky pro potřeby NRP. Na základě výsledku analýzy budou uživatelé automaticky poskytnuta doporučení, jak zlepšit kvalitu datasetů a vyřešit nedostatky. Současně bude sloužit pro správce repozitářů jako zdroj informací o naplňování FAIR principů u ukládaných datových sad. Podpůrné nástroje a materiály budou průběžně aktualizovány a přizpůsobovány uživatelským potřebám, jak na základě vývoje v této oblasti, tak na základě metod UX designu. Předpokládá se úzká spolupráce s oborově zaměřenými pracovními skupinami zapojených do projektu Open Science II, ve kterém budou řešit způsob plnění FAIR principů v konkrétních vědních oblastech.

Bude aktivně sledován vývoj a praktiky v oblastech FAIR principů, služeb EOSC a otevřené vědy v mezinárodním prostředí tak, aby připravované řešení reflektovalo nejnovější poznatky a zkušenosti. Připravované řešení bude prezentováno a konzultováno na mezinárodních a národních odborných fórech.

Naplňování FAIR principů má závislost i na standardech a doporučení jednotlivých vědních oborů. V rámci open science se předpokládá významnější mezioborová spolupráce. Z toho důvodu bude vytvořena pilotní FAIRifikační nástavba na vybrané nástroje pro sběr (pre)klinických dat v oblasti laboratorních analýz, které mohou kombinovat data z různých oborů, a tím se může stát modelovým příkladem pro další obory, které pracují s daty jednotlivých osob. FAIR principy též předpokládají vznik komunitních standardů pro jejich naplňování, v rámci této aktivity bude též provedena technická příprava pro pilotní vytváření těchto standardů v podobě tzv. FAIR Implementation Profiles (FIP) v souladu s metodikou iniciativy GO FAIR a provázání FIPs na plánování správy dat. Obě tyto činnosti budou základem pro jejich další rozvoj v rámci vědních oborů v navazujícím projektu Open Science II.

Služba bude propojena se systémem pro podporu tvorby DMP (KA5.2). Jednak bude využívat informace z DMP jako vstupy do kontrol a dále výsledky kontrol budou zpětně zobrazitelné v rámci DMP dashboardu. Plánování správy dat též bude využívat informace a znalosti z FIP. Všechny tyto i výše zmíněné navzájem sdílené informace budou využívány pro zkvalitňování datasetů (jejich FAIRness) a tím i repozitářů, které tuto službu využijí.

Hlavní výstupy

- Analýza a doporučení k naplňování FAIR principů v rámci NRP.
- Služba pro podporu zvyšování FAIRness datasetů ukládaných do repozitářů v rámci NRP.
- Technická příprava pro komunitní plánování naplňování FAIR principů v OS II (FAIR Implementation Profile Wizard).
- Návod na implementaci nástroje pro ověřování naplňování FAIR principů do repozitářů NRP, materiály a doporučení pro uživatele.

7.5.2 KA 5.2 PODPORA PLÁNOVÁNÍ SPRÁVY DAT

Cílem této aktivity je připravit podporu plánování správy dat jako přímou součást NRP, která nejen umožní vytváření tzv. plánů správy dat (DMP, *data management plan*), ale bude na základě strukturovaného a spravovaného znalostního modelu interaktivně a efektivně provázet uživatele (především vědce) možnostmi správy dat s přímou návazností na služby NRP, data v ní uložená, ale i s ohledem na nastavení instituce, která je vlastníkem a/nebo správcem dat.

Plánování správy dat bude uživatelům NRP nabízeno prostřednictvím nástroje DSW¹⁶, který je již využívaný českými institucemi, vyvíjený v ČR a mj. doporučený Evropskou komisí, OP JAK a TAČR. Nástroj bude integrován s ostatními komponenty NRP, např. bude nabízet uživatelům vhodné repozitáře, dostupné a podporované standardy, doporučené licence, zohledňovat úroveň citlivosti dat a související pravidla. Zde bude probíhat intenzivní spolupráce s týmem NTK, který v rámci IPs CARDS vytvoří Národní katalog repozitářů (NKR), který bude evidovat repozitáře a standardy a který poskytuje metodickou podporu práce s metadaty, což náplní činnosti zahrnuje mj. specificky poskytovat součinnost při hledání vhodného repozitáře či metadatového schématu. Implementace tedy bude koordinována s Centrem repozitářů a správy metadat NTK.

Výzkumníci budou mimo správného využívání funkcionalit a služeb NRP také vedeni k aplikování tzv. *best practices* (např. FAIR, Open Science, *reproducibility*) na základě obecně dostupných doporučení i specifických v rámci projektu (KA5.1). K tomu budou využívány stávající technické možnosti DSW, které budou dále dle potřeb KA5.1 dovyvinuty. Jedná se zejména o kódování a vyhodnocování metrik DMP, poskytování interaktivní zpětné vazby, provazby a reference na edukativní a metodické materiály.

Podpora plánování správy dat se zaměří nejen na vytváření DMP dokumentů, ale bude zejména podporovat proces správy výzkumných dat, tedy průběžné aktualizace plánu v průběhu projektu a pokrytí celého životního cyklu datových sad. Tento životní cyklus bude vhodnou formou vizualizován a plány správy dat budou mít vybrané role (autor a kurátor dat, správce repozitáře apod.) a budou průběžně automatizovaně vyhodnocovat např. v návaznosti na vyhodnocování plnění principů FAIR.

Pro usnadnění plánování správy dat, zejména s ohledem na rozhodování o způsobech naplnění FAIR principů bude systém pro DMP též připraven na využívání komunitních standardů FIP (viz KA5.1).

Výstupem plánování nebudou jen dokumenty (např. PDF nebo MS Word) sloužící jako příloha ke grantovým žádostem pro české prostředí (v češtině i angličtině), ale také strojově čitelná podoba DMP v souladu s RDA DMP Common Standard for maDMPs¹⁷ uzpůsobeného o specifické údaje klíčové pro národní prostředí a automatizaci v rámci NRP. Skrze tuto strojově čitelnou podobu s důrazem na využití PID bude zajištěna interoperabilita za účelem dalších automatizací a integrací v jiných nástrojích i mimo NRP. Důležitou funkcionalitou bude i přiřazování PID samotným DMP jakožto klíčovými dokumentům a výstupům projektů. Instancím repozitářů v NRP bude umožněno přizpůsobení plánování uživatele dané instance a cílem poskytnout jim detailnější vedení specifické pro danou instituci či typy projektů. Bude také vybudováno aplikační rozhraní (API) pro možnost integrace plánovací služby pro třetí strany (do informačních systémů uživatelů a jejich institucí, např. pro napojení na CRIS systémy, notifikační aplikace, portálové služby a tzv. dlaždicové systémy). Toto rozhraní je zamýšleno zejména s ohledem na budoucí integrace, ale bude využito i v rámci ostatních vyvíjených služeb NRP (např. KA5.4).

¹⁶ <https://ds-wizard.org>

¹⁷ <https://www.rd-alliance.org/group/dmp-common-standards-wg/outcomes/rda-dmp-common-standard-machine-actionable-data-management>

Návrhy architektury, integrací, machine-actionable DMP, “lessons learned” z implementace a další postupy zefektivňování plánování správy dat budou prezentovány a konzultovány s odbornou veřejností v rámci mezinárodních konferencí zaměřených na toto téma a sloužících k výměně zkušeností z národních prostředí (např. RDA Plenary, IDW / SciDataCon, IDCC, RSECon apod.). Dílčí výsledky budou rovněž prezentovány formou OpenAccess (např. Data Science Journal, Scientific Data (Nature) či International Journal of Digital Curation).

Hlavní výstupy

- Funkcionalita pro sandboxing a bootstrapping pro testování a zpětnou vazbu (v návaznosti na potřeby technické podpory pro školení v KA7).
- Umožnění práce s lokalizovanými strojově-zpracovatelnými plány správy dat a vizualizace procesu plánování správy dat v prostředí NRP.
- Finální verze nástroje pro plánování správy dat integrovaný v prostředí NRP umožňující efektivní vedení uživatelů a úpravu obsahu dle organizačních a národních postupů i jejich budoucích změn [V4].

7.5.3 KA 5.3 AUTOMATIZACE SBĚRU DAT A METADAT

Tato aktivita se zaměřuje na možnosti přímého ukládání dat produkovaných vědeckými přístroji do datových úložišť NRP a následné další využití a organizaci takových dat v rámci NDI.

Aktivita zmapuje potřeby pro přímé napojení vědeckých přístrojů vědeckých infrastruktur a institucí do NRP a zajistí následný vývoj kritických komponent. Návrh fungování napojení ilustruje Obrázek 1. Komponenty pro napojení přístrojů budou přizpůsobeny konečným použitým technickým řešením, která budou finalizována v průběhu řešení tohoto projektu.

V rámci aktivity budou navrženy postupy, včetně vývoje potřebných komponent a aplikací, pro přímé napojení přístrojů na repozitáře NRP, umožňující vytěžování dat a metadat z ovládacích zařízení přístrojů, a identifikaci uživatelů, kteří měření provádí. Aktivita bude pracovat s dostupnými identifikátory, aby byla zajištěna jednoznačnost údajů, eliminovalo se manuální zadávání a tím snížilo i riziko chybovosti. Kromě identifikátorů osob budou užívány také např. identifikátory institucí, projektů/aktivit, nebo použitých materiálů (IPs EOSC-CZ). Následně budou pro zajištění interoperability metadata generovaná přístroji harmonizována se základním metadatovým modelem pro NMA a NRP definovaným v projektu IPs CARDS. Vytěžování neměnných dat a metadat z přístrojů s přímým navázáním na uživatele, instituci a projekt umožní jednoznačnou identifikaci vzniku dat a jejich původce. Pro potřeby jednoznačné identifikace v rámci KA5.3 budou také vytvořeny zjednodušené katalogy, které umožní evidenci a přidělování PID instancím vědeckých přístrojů a fyzickým vzorkům měřeným na daných přístrojích. Katalogy budou navrženy tak, aby umožnily jejich rozšíření /škálování pro potřeby vědeckých komunit v programu Open Science II.

Ve spolupráci s KA2 zajistíme podporu tvorby záznamů v delších časových horizontech na straně softwarových platforem instancí repozitářů v NRP, podporu agregace dat z více měření do jednoho datasetu a možnost kombinování různých částí záznamů pro výstup do archivní části repozitáře. Systém bude také podporovat komplexní kontrolu přístupů k datům (*multiuser access*), případně možnost navazující práce s daty (*workflows*), a/nebo ověření plnění principů FAIR pro automatický sběr dat – zde bude využíváno služeb a rozhraní implementovaných v KA2, KA4 a ostatních aktivitách KA5. Klíčovou funkcionalitou automatizace sběru dat do repozitářů, kterou v rámci projektu implementujeme, bude efektivnější ukládání a práce s generovanými daty a jejich snadnější zveřejňování, což podporuje soudobou dobrou Open Science praxi.

V přípravné fázi projektu jsme identifikovali 10 typů přístrojů, nebo jejich virtuálních reprezentací, u kterých bude v rámci této aktivity implementováno přímé propojení přístroje nebo přístrojových komplexů s datovými úložišti v NRP, s pilotními repozitáři a nastavení napojení na služby NRP tak, jak to popisuje Obrázek 1. Tyto případy užití (*use-cases*) byly vybrány tak, aby pokryly různé typologie přístrojů: komerční i nekomerční (*home-built*) přístroje, s jedním operátorem i víceuživatelské, s manuálním, poloautomatickým i plně automatickým ovládním, umístěné v individuálních laboratořích nebo ve velkých výzkumných infrastrukturách. Přístroje se liší velikostí jednotlivých souborů (<1 GB, >1 GB), které jsou výstupem přístroje, počtem souborů generovaných v průběhu typického měření (<100 souborů, >1000 souborů) a komplexitou záznamů (jednoduchý *versus* vícesložkový záznam). Jedná se převážně o měřicí přístroje, ale seznam zahrnuje i přístroj pro simulace (klastr) a virtuální systém (web crawler). Dva případy reprezentují sestavu s více moduly, které generují data. Tyto sestavy budou využity pro ověření synchronizace sběru dat a metadat z více jednotek do jednoho záznamu, jeden případ simultánně generuje data a metadata z více měření (> 50) najednou. Jeden případ byl vybrán pro ověření implementace pro náročná dlouhodobá měření s velmi velkými objemy dat (> 100 GB). Pro hlubší ověření funkcionality agregace dat z různých měření a víceuživatelský přístup k rozpracovanému záznamu otestujeme automatický a paralelní sběr dat z různých instancí přístrojů a od různých tvůrců (uživatelů) do jednoho záznamu. Náročnost implementace se bude lišit z důvodu různé komplexity případů užití. V případě jednodušších přístrojů navrhujeme lidskou kapacitu pro vývoj a implementaci 3–4 PM, u složitějších systémů 8–10 PM. Výstupem těchto případových studií budou doporučení, metodické a technické návody a softwarové komponenty pro implementaci automatizace sběru dat pro různé instance přístrojů, různé typy repozitářů v NRP a pro různé uživatelské prostředí.

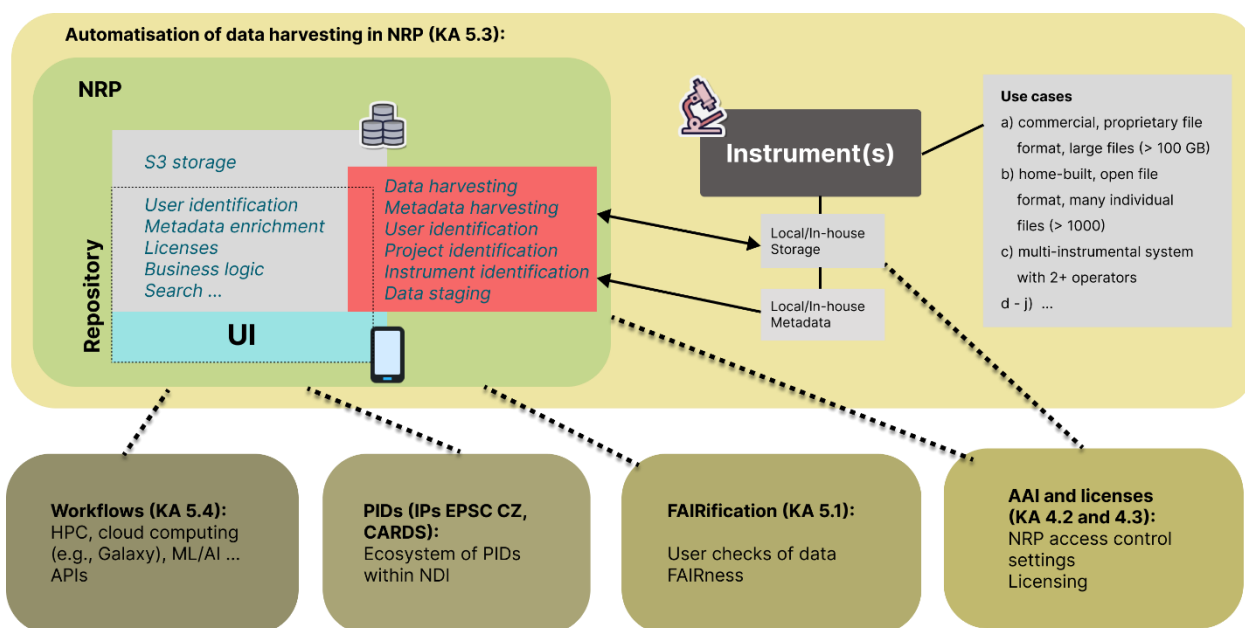


Schéma 3 Schéma zapojení vědeckých přístrojů do NRP a návaznost na služby

Návrhy řešení pro automatizaci procesů sběru dat a metadat a jejich implementace budou prezentovány na mezinárodních konferencích zaměřených na toto téma (např. RDA Plenary, International Data Week, EOSC Symposia, EMMC Conference apod.), případně na hlavních oborových konferencích nebo v časopisech dle oboru (např. AI pro automaticky sbíraná multimodální webová data). Postupy a implementace budou dále konzultovány se zahraničními partnery v rámci etablovaných spoluprací a výměnných pobytů, nebo v průběhu lokálních workshopů (např. NFDI Workshops).

Hlavní výstupy, milníky a činnosti

Poskytnutí potřebných podpůrných komponent, které bude možné (po případné úpravě pro konkrétní případ) zjednodušeně využít pro napojení ovládacích jednotek přístrojů do repozitářů NRP, včetně podrobných návodů (v češtině i angličtině) pro implementaci v RI, institucích nebo na jiných pracovištích.

7.5.4 KA 5.4 INTEGRACE VÝPOČETNÍCH WORKFLOWS

Cílem této klíčové aktivity je umožnit efektivní a přímočaré propojení NRP s výpočetními kapacitami. Konkrétním výstupem bude napojení NRP na výpočetní infrastruktury e-INFRA CZ a podpora a pilotní ověření napojení externích workflow nástrojů pro zpracování dat. Obecně se jedná se o širokou problematiku, kterou je třeba rozdělit na dílčí části a která se z podstaty věci pohybuje na rozhraní několika projektů a aktivit. Ambicí této aktivity není poskytnout komplexní řešení pro všechny scénáře výpočetních vědeckých workflow zahrnujících data, ale vhodným způsobem přispět k doplnění existujících komponent a ověřit vybrané případy užití.

Základním scénářem bude využití nástrojů pro přenos dat vzniklých v rámci projektu EOSC-CZ k realizaci výpočetních úloh využívajících NRP jako zdroj dat. Zde se jedná především o vyhledání datový sad, jejich přenos (stage-in) na výpočetní prostředí a podpora konkrétních případů užití, případně pilotních uživatelských skupin, ve spolupráci s projektem EOSC-CZ. Specifickým případem je pak umožnění provádění výpočtů v případě vybraných externích výpočetních prostředí, která podporují zpracování citlivých dat (návaznost na projekt EOSC-ENTRUST, projekt SensitiveCloud a pilotní prostředí v rámci IT4I). Následným krokem je ověření a podpora komponent pro publikaci dat do repozitáře přímo z výpočetních prostředí.

Pokročilejším scénářem je napojení existujících nástrojů pro řízení workflow a vybraného prostředí pro GUI přístup uživatelů k výpočetním kapacitám. Zde je cílem nabídnout ověřená rozhraní a poskytnout odpovídající podporu pro správce a vývojáře pilotních workflow ve vybraných nástrojích. Konkrétně se jedná o workflow systémy Galaxy (CESNET) a Lexis (IT4I) a prostředí pro GUI přístup založená na technologii Jupyter notebooků. Pro zajištění koordinace s vývojáři workflow systému se předpokládají cesty na zahraniční konference. Cílem je vzájemná prezentace navrhovaných a realizovaných modifikací. Příkladem takové konference je např. konference European Galaxy Days, kde se každoročně prezentují novinky ve vývoji a je možné se setkat i s vývojáři.

Velmi důležitým aspektem výpočetních úloh a workflow z hlediska NRP jsou metadata. Vlastní práce s metadatami není cílem této aktivity, zvolená řešení však nesmí bránit tomu, aby ve všech popsáných případech byla metadata k dispozici (hledání dat, přenos dat, externí workflow nástroje) a aby bylo možné s nimi odpovídajícím způsobem pracovat. Předpokládáme intenzivní spolupráci s ostatními aktivitami projektu i s komplementárním projektem IPs CARDS a jeho Centrem repozitářů a správy metadat, přičemž tato aktivita především poskytne přístup formou API k metadatům a událostem popisujícím vlastní zpracování dat (průběh zpracování a popis prostředí ve kterém proběhlo). Specifická pozornost bude věnována koordinaci s aktivitou KA5.2 Plánování správy dat (možnost vyhodnocovat DMP s ohledem na skutečné zpracování dat) a KA5.3 Příprava pro automatizaci sběru dat. Výsledky práce budou využity ve spolupráci s tvůrci repozitářů v rámci NRP, a to včetně potřeb pilotních repozitářů z KA3.

7.5.5 KA 5.5 ROADMAPA POKROČILÝCH SLUŽEB NRP/NDI

Odpovědností této aktivity je dohled nad implementací a provozem služeb definovaných v KA 4 a 5. Ve vazbě na vývoj architektury NDI/NRP a výstupy pracovních skupiny EOSC-CZ, zejména skupiny *Základní služby*, bude také udržován aktuální seznam požadovaných služeb NDI/NRP a u vybraných služeb bude navržena jejich implementace a provoz. Tato aktivita tedy zajistí průběžné sledování požadavků a postupný rozvoj služeb v souladu s postupně se vyvíjenými nároky uživatelů.

V aktivitě bude sledován vývoj nových technologií a architektur pro práci s daty ve vědeckém prostředí – v případě potřeby bude navržen způsob adaptace služeb NRP/NDI. Pro zajištění přehledu o *state-of-the-art* v oboru se předpokládají cesty na zahraniční konference v oblasti datových infrastruktur, kde může být tato problematika konzultována jak se zahraničními poskytovateli služeb vědecké komunitě, tak s jejich uživateli. V rámci roadmapy budou vznikat doporučení, *best-practices* a ukázkové nasazení uživatelských služeb, které budou udržitelně provozované nad daty v NRP. U málo využívaných služeb bude identifikována příčina nízkého využití (např. nevhodná implementace, nevhodné API, nepotřebná služba, ...) a navržena úprava služby, náhrada jinou službou nebo zrušení služby bez náhrady, kdy je pak součástí vždy řešení procesu migrace uživatelů dané služby.

Jedním ze sledovaných témat bude např. napojení NRP na prostředí pro správu a analýzu citlivých dat tak, aby byly splněny všechny administrativní, technologické a legislativní požadavky na správu tohoto specifického typu dat. Z pohledu automatizace *workflows* a poskytování dat z NRP bude vhodné analyzovat možnosti pro budování a poskytování služeb typu anonymizace, parametrizace dat umožňující využití modelů strojového učení apod.

7.6 KA 6 – KYBERBEZPEČNOST A PODPŮRNÉ SLUŽBY

Cílem této aktivity je zajištění průřezových činností a služeb jak v rámci projektu, tak i jako jeho rozhraní směrem k uživatelům a veřejnosti. Uvedené zahrnuje zejména o následující činnosti:

- Zabezpečení NRP a zajištění kyberbezpečnostního dohledu
- Compliance a řízení ISMS i s péčí o citlivá data
- Pro-uživatelský přístup návrhu a řízení implementace služeb NRP, UI a UX
- Uživatelská podpora a ServiceDesk

Vybudování a provoz Národní repozitářové platformy, jakožto multitenantní technologické platformy sloužící k správě, ukládání a následnému zpřístupnění dat oprávněným uživatelům představuje unikátní prostředí, jehož cílem je sloužit celému výzkumnému prostoru ČR.

Toto řešení krom nepopíratelných pozitiv spočívajících v jednotné správě dat, jejich dostupnosti, replikaci aj. s sebou nese i určitá rizika, mezi která je možné zařadit kybernetické útoky, selhání lidského faktoru, ne zcela vhodné nastavení uživatelského prostředí aj.

Cílem klíčové aktivity KA6 je nastavení standardů a pravidel v oblasti kybernetické bezpečnosti a compliance, dále pak začlenění uživatelské podpory v podobě service a help desku do stávajících struktur e-INFRA CZ a zajištění fungování jednotlivých součástí NRP skrze pro-uživatelský přístup a řízení implementace služeb NRP, UI a UX.

Výstupy této aktivity mají dopad na celý projekt NRP a taktéž na všechny uživatele a administrátory, neboť mimo jiné nastavuje a dále specifikuje standardy a pravidla užívání NRP z pohledu bezpečnosti, pravidla pro přístup k jednotlivým typům dat, jakož i legální limity.

7.6.1 KA 6.1 KYBERBEZPEČNOSTNÍ DOHLED A ZABEZPEČENÍ NRP

Vzhledem ke vzniku Národní repozitářové platformy, jakožto multitenantní technologické platformy sloužící ke správě, ukládání a následnému zpřístupnění dat oprávněným uživatelům nelze otázku kybernetické bezpečnosti ani podceňovat ani bagatelizovat. Je to oblast, která je pro NRP a uložená data klíčová, a proto je třeba ji řešit od počátku projektu.

Kybernetická bezpečnost v projektu bude představovat souhrn právních, organizačních, technických a vzdělávacích prostředků, které směřují k zajištění ochrany počítačových systémů a dalších prvků ICT, aplikací, dat a uživatelů. Kybernetická bezpečnost bude realizována jak v rámci kyberprostoru, tak mimo něj.

Cílem kybernetické bezpečnosti bude zajistit důvěrnost, dostupnost a integritu veškerých dat a informací v projektu NRP. Zajištění integrity dat a informací představuje nezbytnou součást návrhu architektury, služeb, jakož i provozu NRP. Cílem kybernetické bezpečnosti je také vhodně nastavit standardy pro uživatele, technologie a procesy. Specifická pravidla a standardy pak bude třeba nastavit pro zvláštní kategorie dat (např. citlivá data, duševní vlastnictví aj.)

V rámci projektu budou pravidla kybernetické bezpečnosti aplikována v průběhu celého jejich životního cyklu projektu.

Hlavní výstupy

- **Zajištění fyzické bezpečnosti.** V NRP budou přijata bezpečnostní pravidla a opatření pro:
 - prostředí, ve kterém se nachází data a informace (data),
 - režimová pracoviště, kde bude umístěn specifický hardware umožňující např. administrativní přístup do systému či přístup k citlivým či jinak chráněným datům,
 - řízení přístupu, kontroly vstupu, monitoringu určených prostor,
 - zajištění bezpečnosti „lidských zdrojů“ [zejm. zabezpečení fyzického perimetru (kanceláře, trezory aj.)].
- **Návrh architektury kybernetické bezpečnosti** informačních a komunikačních systémů NRP a systémů s NRP souvisejících.
- **Zajištění sběru bezpečnostní telemetrie z NRP za pomoci vybudování systému centrálního log managementu.** V rámci systému dojde ke sběru:
 - provozních a bezpečnostních dat,
 - auditních záznamů,
 - požadavků uživatelů,
 - informací o incidentech, hrozbách a zranitelnostech.
- **Zajištění provozní bezpečnosti.** Tato činnost v rámci spouštění a následném provozu NRP v sobě bude zahrnovat zejména:
 - vytvoření, nastavení a přijetí pravidel pro správce a uživatele,
 - bezpečnostní a provozní dohled (monitoring provozní; monitoring bezpečnostní; sledování zranitelnosti; uložení a zpracování provozních logů na základě zákonných požadavků; zajištění nepopiratelnosti; identifikace provozních problémů; identifikace kybernetických bezpečnostních událostí či incidentů; ex-post analýza při řešení provozních nebo bezpečnostních incidentů),
 - reaktivní bezpečnost spočívající v zapojení bezpečnostních týmů CESNET-CERTS a CSIRT-MU do řešení kybernetických bezpečnostních událostí a incidentů souvisejících s NRP a uživateli,
 - proaktivní bezpečnost v rámci které budou aplikovány penetrační testy (FLAB, CSIRT-MU) sloužící k ověření nastavení bezpečnosti systémů, služeb a aplikovaných pravidel.

Penetrační testování bude prováděno periodicky (1× za 12 měsíců) a ad hoc při významných změnách v rámci NRP.

- **Zajištění organizační bezpečnosti.** V rámci organizační bezpečnosti dojde ke vzniku dokumentace, včetně asset managementu, nastavení pravidel a procesů vyplývajících zejména z ISMS, GDPR aj.
- **Zajištění bezpečnosti AAI.** Správci repozitářů i konkrétních datových sad budou definovat pravidla, kdo a za jakých podmínek je oprávněn data do NRP vkládat či k nim přistupovat (viz i KA4.2, věnovaná licencím). Bude proto třeba zajistit nezpochybnitelnou auditní stopu využití AA infrastruktury (KA4.3) a zajistit, aby bylo možné doložit dodržení požadavků a podmínek, vzešlých ze zvoleného nastavení přístupových oprávnění. K tomuto účelu bude v NRP vybudován centrální logovací a analytický systém.
- **Vzdělávání v oblasti kybernetické bezpečnosti.** Ve spolupráci s KA7 dojde k vytvoření certifikovaného e-learningového kurzu NRP zaměřeného na oblast kybernetické bezpečnosti, ochrany dat a jejich možnému zpracování.

7.6.2 KA 6.2 COMPLIANCE A ŘÍZENÍ ISMS I S PÉČÍ O CITLIVÁ DATA

Vznik jednotné architektury pro ukládání vědeckých dat a propojení jednotlivých repozitářů má řadu nesporných výhod pro ukládání FAIR dat. Zároveň však s sebou nese i nutnost vypořádat se s právními a organizačními aspekty tohoto přístupu. Proto musí být nedílnou součástí projektu NRP také činnosti v oblasti compliance a ochrany informací (s důrazem na péči o citlivá data). Tyto činnosti se zaměří zejména na:

- soulad projektových činností a výstupů projektu s právním předpisy, podmínkami projektu a závaznými standardy pro jednotlivé činnosti,
- ochranu dat a informací před neoprávněným přístupem (citlivá data, duševní vlastnictví, chráněné informace apod.),
- ochranu osobních údajů a soukromí uživatelů,
- transparentní komunikace služeb směrem k uživatelům.

Činnosti v oblasti compliance a ochrany informací, s ohledem na výše uvedené, budou zahrnovat:

- soustavný monitoring plánovaných aktivit a jejich soulad s právními předpisy a dalšími závaznými pravidly,
- koordinované řešení sporných právních otázek a příp. zajištění sjednocení právních výkladů,
- zajištění poradenství členům jednotlivých týmů v oblasti compliance a ochrany informací,
- v případě potřeby formulace závazných pravidel a postupů pro zapojené osoby,
- sledování relevantní legislativy a smluvních závazků a formulace opatření v případě zjištěných neshod,
- sběr a koordinace řešení podnětů a upozornění na non-compliance stavy.

Cílem činností v oblasti compliance a ochrany informací je zejména zajistit úspěšnou realizaci projektu NRP i z pohledu jeho souladu s právními předpisy, předcházet případným sankcím za porušení předpisů či smluv a přispět tak i k udržitelnosti projektu.

Vznikem jednoho koordinačního místa pro všechny výše uvedené činnosti bude docíleno určité jistoty, že nebude docházet k tříštění informací, plýtvání zdrojů na duplicitní činnosti a bude usilováno o jednotný postup zejména ohledně řešení právních otázek kyberbezpečnosti, ochrany osobních údajů i ochrany informací obecně.

Činnosti v oblasti compliance a ochrany informací bude koordinovat compliance manažer, který bude úzce spolupracovat s právníky a dalšími odborníky na jednotlivé oblasti compliance a ochrany informací. Bude úzce spolupracovat s právníkem KA7/Data Lawyer. Mezi hlavní aktivity compliance manažera bude patřit:

- se v rámci prevence non-compliance stavů účastnit plánování/revizí/hodnocení projektových a aktivit s cílem včas zajistit hodnocení rizik a stanovení případných opatření,
- zajišťovat zpracování právních analýz a expertních posudků za účelem hodnocení rizik a stanovení opatření v oblasti compliance,
- napomáhat při integraci projektových aktivit do činností a úkolů koordinátorů a partnerů všech dotčených projektů,
- zajištění poradenství v oblasti GDPR/kybernetické bezpečnosti/compliance.

Zapojení compliance manažera do strategického plánování hlavních činností od samého začátku projektu pak umožní předcházet vzniku non-compliance stavů, které by mohly zvyšovat náklady na jeho realizaci či dokonce ohrozit úspěšnost celého projektu. Důraz na bezpečnost informací a ochranu osobních údajů uživatelů a transparentní řešení podnětů od realizačních týmů pak přispěje k důvěryhodnosti výsledků projektu a poskytovaných služeb.

7.6.3 KA 6.3 PRO-UŽIVATELSKÝ PŘÍSTUP IMPLEMENTACE SLUŽEB NRP A NDI

Ekosystém NRP bude tvořen řadou samostatně vyvíjených, ale vzájemně propojených nástrojů a služeb, s očekávaným využitím jak koncovými uživateli (vědecké komunity, datová podpora, veřejnost), tak správci služeb a repozitářů v NRP/NDI prostředí (řízení přístupů, řízení nahrávání datasetů, poskytování licencí, navigace a informační architektura atd.). Z tohoto důvodu je nezbytné, aby interakce uživatelů s těmito systémy a službami byla relativně jednoduchá (*easy to use*), intuitivní, kompatibilní s FAIR principy a s přesahem na mezinárodní úroveň. Navrhované řešení musí současně respektovat rovnováhu mezi uživatelskou použitelností (*user experience*), technickými možnostmi systému a bezpečnostními požadavky stanovenými jak legislativními normami, tak architekturou ekosystému NRP/NDI.

Pro-uživatelský přístup neboli design zaměřený na uživatele (*user-centered design*) umožňuje vytvářet pozitivní a relevantní uživatelskou zkušenost, neboť aktivně a iterativně reflektuje potřeby uživatelů, a to během celého procesu řešení problému. Tento přístup zapojuje uživatele do samotného procesu navrhování a tvorby NRP/NDI a jejich služeb skrz uživatelský výzkum (*user experience research*), testování (*user testing*) a monitoring kvality navrhnutých řešení.

V rámci pro-uživatelského přístupu bude prováděno kontinuální mapování potřeb a preferencí uživatelů v kontextu používání NRP a analýza samotného ekosystému NRP. Výsledky budou podkladem pro další rozhodování (*data-driven decision making*). Cílem těchto aktivit je objevování slabých míst a problémů (*usability issues*), a také vytvoření základního přehledu pro zlepšení a rozvoj poskytovaných služeb NRP/NDI, související uživatelské podpory (KA6.4 Uživatelská podpora a Service Desk), vzdělávacích aktivit (KA7 NRP a Školící centrum EOSC-CZ) a pro plánování pokročilých aktivit a služeb (KA5.5 Roadmapa pokročilých služeb NRP/NDI).

Výsledná zjištění, budou prioritizována dle jejich dopadu na primární cíle projektu NRP a představují východisko pro tvorbu vhodných návrhů a prototypů možného řešení, které reflektují technické a bezpečnostní aspekty budovaného ekosystému NRP/NDI. Zvolené řešení bude UX designový tým koordinovat s technickým a kyberbezpečnostním. Uvedeným postupem bude podpořeno i budování uživatelsky orientovaného designu napříč vývojovými týmy, a tedy i v celém ekosystému NRP/NDI.

Pro naplnění cíle aktivity KA6.3 bude UX designový tým operovat na průřezové úrovni napříč klíčovými aktivitami projektu NRP. Dle zjištěných výsledků a následné prioritizace bude následně úzce spolupracovat s vybranými týmy v rámci konkrétně definovaných činností. Kontinuální práce týmu bude prováděna v úzké součinnosti s klíčovou aktivitou zaměřenou na vzdělávání (KA7) a jejich metodiky pro jednotlivé platformy, zejména dojde k zapojení komunity Data Stewardů (KA7.2). V rámci identifikace slabých míst bude nastavena intenzivní kooperace s klíčovou aktivitou Uživatelská podpora a Service Desk (KA6.4). Pro zajištění bezpečnosti celého ekosystému NRP/NDI bude nastavena i úzká komunikace s aktivitou zaměřující se na kyberbezpečnost (KA6.1.).

Pro zajištění souladu s potřebami vědecké komunity bude extenzivně využívána i expertíza Pracovních skupin pro implementaci EOSC v ČR (PS EOSC CZ) a analytického týmu IPs EOSC-CZ. Dále bude nastavena úzká spolupráce se Sociologickým ústavem AV ČR v rámci plošného mapování vědecké komunity ČR a dalších cílových skupin. Kvantitativní šetření budou probíhat v koordinaci i s dalšími projekty, zejména PS EOSC CZ a projektem CARDS. V případě nutnosti mohou být konzultačně přizvána další expertní pracoviště.

Veškerá relevantní zjištění budou průběžně prezentována vhodnou formou vedoucím dalších klíčových aktivit, aby s nimi bylo možné efektivně pracovat.

Hlavní výstupy:

Hlavním výstupem aktivity KA6.3 bude kontinuální zlepšování kvality prostředí a služeb NRP zejména z uživatelského pohledu, skrz aplikaci designu zaměřeného na uživatele. Hlavní výstup je možné rozdělit do dalších klíčových výstupů a aktivit:

- **Metodické zajištění, realizace, analýza a prezentace výsledků z kvalitativních a kvantitativních šetření, například:**
 - **Vytvoření uživatelských person (user personas) pro cílové skupiny uživatelů NRP/NDI.** Persony neboli profily fiktivních uživatelů reprezentují objevené vzorce chování a jejich potřeby a poskytují podklad pro rozhodování v rámci dalších aktivit NRP.
 - **Vytvoření user journey map.** Vytvoření vizualizace průchodu persony službou/ekosystém NRP a znázornění jednotlivých akcí uživatele k tomu, aby dosáhl svého cíle.
 - **Uživatelské scénáře (user scenarios).** Uživatelské scénáře představují základ pro popis situací, se kterými se uživatel při interakci se systémem setkává a slouží jako podklad pro tvorbu prototypů.
- **Zmapování vědecké komunity ČR.** V průběhu projektu budou provedeny tři komplexní kvantitativní výzkumy, které poskytnou vhled do vnímání implementace klíčových oblastí a služeb NRP/NDI, ale i iniciativy EOSC CZ.
- **Metodické zajištění/popis monitoringu prostředí a služeb NRP a pravidelný reporting relevantním zúčastněným stranám** (úzká spolupráce s dalšími KA v rámci projektu NRP, EOSC-CZ a CARDS).
- **Participace na vývoji a implementaci konkrétních služeb/komponent NRP a NDI** ve formě designových aktivit a konzultací. Jde například o tvorbu a testování UX designů a prototypů pro služby NRP/NDI s cílem získat zpětnou vazbu pro zlepšení uživatelské zkušenosti a zvýšení efektivity daných služeb.

Další konkrétní výstupy a aktivity budou definované v průběhu projektu na základě aktuálních zjištění a potřeb jak uživatelů, tak projektového týmu a dalších stakeholderů. Definování případných dalších

výstupů proběhne po uplynutí 12 měsíců od začátku projektu skrz Uživatelský poradní výbor Výkonného výboru NRP, který bude ustanoven při zahájení projektu. Tento poradní výbor bude zahrnovat vybrané pracovníky napříč KA a vybrané externí odborníky/odbornice.

7.6.4 KA 6.4 UŽIVATELSKÁ PODPORA A SERVICE DESK

Uživatelská podpora a Service Desk představují nezbytnou součást projektu NRP zejména s ohledem na zajištění nepřetržitého (24/7/365) využívání monitoringu funkčnosti technologických součástí NRP (viz KA 2 a 3) a dále pak jako jednotné místo sloužící pro příjem a řešení požadavků uživatelů.

Service Desk přijímá, zpracovává a řeší (v rozsahu L1) uživatelské požadavky, vyhodnocuje problémy či incidenty a eskaluje je na podporu vyšší úrovně či vlastníka služby. Taktéž zajišťuje evidenci požadavků, jejich kategorizaci, vyplnění potřebných metadat pro dlouhodobou statistiku a informuje vlastníka o využívání podpory pro jeho službu.

Vlastní činnost uživatelské podpory a service desku bude plně navázána na service desk e-INFRA CZ. V rámci NRP tedy dojde k integraci a posílení již existujícího a funkčního systému. Dílčí rozšíření stávající služby service desk je však nezbytné zejména vzhledem k nezbytnosti monitoringu nově vzniklých repozitářových platforem a služeb na ně navázaných, jakož i vzhledem k předpokládanému nárůstu uživatelů a požadavků spojených s NRP.

Hlavní výstupy

- Garance, že každé službě NRP odpovídá alespoň jedna RT fronta s definovanými OLA.
- Příjem a zpracování požadavků na služby a hlášení o incidentech. V rámci činnosti bude prováděna klasická L1 podpora, tj. příjem požadavků elektronicky a telefonicky; jejich zpracování do lístku v systému pro správu požadavků; odpověď v rámci dohodnutých pravidel se správci služeb a předání lístku na vyšší úroveň podpory v případě potřeby.
- Monitoring stavu služeb jak směrem k uživateli, tak k vlastníkovi služby. Service Desk bude používat monitorovací nástroje pro zjištění stavu software, hardware, síťových linek aj. V případě zaznamenání výpadku vytvoří v rámci stanovených pravidel a postupů v nástroji pro správu požadavků lístek a tento incident budou aktivně řešit. Evidován bude taktéž stav řešení, plánované výpadky, odstávky aj.
- Reporting dostupnosti služeb. Na základě monitoringu stavu služeb a z toho vznikajících lístků bude Service Desk provádět reporting dostupnosti jednotlivých služeb. Evidování budou dotčení uživatelé, délky výpadků, typy výpadků (plánovaný/neplánovaný), ovlivnění dohodnutých specifických pravidel služby aj.
- Uživatelská podpora bude poskytována v součinnosti s KA 7.3. Vybudován bude „helpdesk“ pokročilé podpory (level 2) pro software a služby NRP. Vyškolení metodici budou schopni poskytovat podporu data stewardům a současně budou schopni koordinovat řešení události se Service Deskem.

7.7 KA 7 – „ŠKOLENÍ A OSVĚTA V OBLASTI FUNKCIONALIT A SLUŽEB NRP“

V rámci KA7 bude poskytnuta podpora pro získávání a prohlubování znalostí a dovedností potřebných pro optimální využívání infrastruktury Národní repozitářové platformy, jádra Národní datové infrastruktury.

Cílovou skupinou KA7 NRP jsou zaměstnanci výzkumných institucí a vysokých škol působících na datově orientovaných pozicích, např. data steward, datový kurátor, správce repozitáře, administrátor analytik nástrojů pro provoz repozitářů. V textu je skupina profesí souhrnně nazývána „Datová podpora“.

Náplně činností i výstupy KA7 jsou celkově vždy určeny pro zaměstnance institucí přímo podpořených v projektu NRP, stejně tak zaměstnancům institucí mimo projektové partnerství.

V květnu 2023 proběhlo pod záštitou Univerzity Karlovy a pracovní skupiny EOSC pro vzdělávání a lidské zdroje formou dotazníkové šetření vzdělávacích potřeb mezi institucemi, o kterých se v té době uvažovalo, jako o potenciálních partnerech projektu (Příloha č. 2 ke Studii proveditelnosti). Průzkumu se v první vlně účastnilo 35 institucí různé velikosti. Motivací průzkumu bylo zjištění, jaký je případně zájem o jednotlivá témata školení a zejména předpokládaná kapacity objednávky na školení v tématech souvisejících s výzkumným daty. Respondenti byli dotazováni na svou představu o potřebě školení pro rok 2024. Toto časové omezení bylo dáno tím, že dotazovat se respondentů na jejich odhady pro delší období a zároveň zachovat konkrétnost dotazů by pravděpodobně přineslo obecné odpovědi – vzdělávat se v oblasti nakládání s výzkumnými daty, což by byla relevantní odpověď, ale nepomohla by identifikovat tematické vymezení. Průzkum ukázal, že v dané oblasti jsou velmi roztržité představy o tematických vymezení školení, ovšem jasně vyplynula potřeba konkrétních témat na školení nástrojů DMP, nástrojů na správu a sdílení dat, správa datových verzí, ale i technická řešení pro datové repozitáře nebo Call4Help pro národní nástroje.

Zjištěno bylo také, že plošně lze velmi těžko určit zájem o školení určené pro výzkumné pracovníky a studenty. Pro projekt NRP bylo tedy rozhodnuto, že je třeba posílit kompetence institucí, a to prostřednictvím orientace na pracovníky datové podpory institucí a v oblasti vzdělávání se zaměřit na posílení komplexního a celkového znalostního potenciálu pro připravenost aplikovat koncept FAIR data do českého prostředí. Průzkum přispěl k tomu, že v rámci projektu NRP je preferována forma vzdělávání formou “trénink trenérů” a forma cíleného vzdělávání (lépe řečeno posunování znalostí a kompetencí) k užívání nástrojů a systému NRP. Tím se projekt NRP v KA7 zaměřuje na cílovou skupinu, se kterou je již v rámci EOSC iniciativ dlouhodobě spolupracováno a je s ní navázána přímá komunikace a lze již v přípravě projektu odhadnout oblasti zájmu a potřebné kapacity pro vzdělávací a osvětové aktivity projektu.

Datovou podporu v této KA vyškolíme a zároveň jí poskytneme konzultační podporu, pomoc, nástroje i pomůcky pro efektivní užití infrastruktury NRP. Na základě školení a osvěty v KA7 bude schopna Datová podpora na jednotlivých institucích poskytovat přímé služby, včetně instruktáží a školení celoživotního vzdělávání výzkumných pracovníků své instituce. Pro tuto činnost vybavíme Datovou podporu metodikami, návody a budeme jí poskytovat služby pokročilého helpdesku.

Pro naplnění cílů KA7 budou využity různé formy komunikace a vzdělávání, které byly detekovány jako optimální, což platí zejména pro elektronické/online formy vzdělávání, a to ve formě synchronní i asynchronní (např. instruktážní audio/video záznam nebo e-learning). V roce 2023 proběhlo šetření v rámci pracovní skupiny pro lidské zdroje a vzdělávání EOSC, kde byl zjišťován názor na optimální formu vzdělávání a jednoznačně dostala prioritu forma distanční. V rámci přípravy metodických a vzdělávacích materiálů bude zohledňována možnost užití hybridní formy vzdělávání v off-line i on-line formě.

Aktivita je úzce navázána na všechny klíčové aktivity projektu, a to jak na “technická řešení” (KA2, KA3, KA4, KA5), tak i na KA6. Zde se jedná zejména o úzkou kooperaci s UX designovým týmem (KA6.3)

Odborný realizační tým KA7 je rozdělen funkčně do dvou základních skupin, které mají stanovené hlavní úkoly k pokrytí cílů NRP. Zároveň platí, že úzká spolupráce v rámci KA7 je podmínkou splnění cílů

NRP. Plnění celé KA7 koordinuje Garant klíčové aktivity. První skupina je určena tím, že bude poskytovat podporu v oblasti technické, typicky pro systémové analytiku, kurátory digitálních dat a repozitářů, IT specialisty jednotlivých českých výzkumných institucí, i mimo projektové partnerství. Tito odborníci jsou označováni jako “metodici”, jejich práce bude navazovat na technickou dokumentaci. Jsou to tito členové odborného realizačního týmu: Metodik CLARIN DSpace, Metodici DMP, Metodik AI, Metodik ASEP ARL, Metodik Invenio, Metodik kybernetické bezpečnosti. Je předpokládáno že se budou účastnit mezinárodních konferencí zaměřených na danou problematiku, aby bylo možné získat nejnovější poznatky a metody. Druhou skupinou jsou odborníci, kteří se orientují více na komunikaci s osobami, které na jednotlivých institucích (i mimo projektové partnerství) poskytují přímo služby výzkumníkům. Typicky je pro tuto pozici užíván termín data steward, nicméně je nutné mít pokrytu i problematiku legislativních a etických aspektů nakládání s výzkumnými daty. Právníci budou zodpovědní za konkrétní úkoly, které jsou nezbytné pro tvůrce a správce repozitářů, zejména: orientaci v autorském právu, právní expertizu v oblasti dat a umělé inteligence, právní analýzy týkající se veřejně dostupných dat, ochrany osobních údajů, odpovědnosti za etické principy, provozu služeb pro uživatele, komerčního použití dat a také právní analýzy pro zadávání veřejných zakázek. Dále budou poskytovat základní orientaci pro tvůrce, data stewardy, další členy realizačního týmu i uživatele NRP v otázkách autorského práva, s důrazem na EU Copyright Directive. Právníci dále budou úzce spolupracovat s týmem zajišťujícím problematiku compliance (KA6.2). Komunikaci na úrovni univerzální podpory a organizace práce Komunity Data Stewardů bude zajišťovat Garant klíčové aktivity, ve spolupráci s data stewardy v realizačním týmu.

7.7.1 KA7.1 METODICKÉ MATERIÁLY PRO SOFTWARE/SLUŽBY NRP

- Instruktažní výukové/metodické materiály, zejména na principu “how to” pro jednotlivé software a služby NRP i jejich kombinované využití.
- Spolupráce se zaváděním a ověřováním software a služeb NRP.
 - návaznost na práci UX design týmu KA6
 - zpětná vazba z pilotního ověření funkcionalit software/služeb NRP
 - metodické postupy pro užívání software/služeb NRP
- Metodické materiály/návody pro užívání software/služeb NRP pro Datovou podporu.
- Šablony metodických materiálů pro užití software/služeb NRP na institucích (editovatelné verze, které umožňují doplnit metodický návod o institucionální kontext, např. kontakty na místní Datovou podporu, odkazy na konkrétní repozitáře). Tímto aspektem bude ošetřeno, že jednotlivé instituce, které budou participovat na výstupech projektu, nebudou muset vytvářet své vlastní školicí materiály, ale materiály pouze doplní, případně rozšíří.
- Zpřístupňování metodických materiálů prostřednictvím Informačního portálu EOSC-CZ.
- Dílčí moduly pro jednotlivé software/služby NRP, vytvářené po ověření v prvním pilotním provozu.
- Využití AI při koncipování dílčích úkolů. Metodické materiály pro software a služby NRP v oblasti aplikace nástrojů umělé inteligence
- Metodické návody pro užívání testovacích verzí software/služeb NRP (sandboxů).
- Organizace setkání k jednotlivým software/službám NRP v průběhu jejich vývoje a pilotního testování.
- Harmonogram: 2024–2028, návazně na vytváření infrastruktury NRP.

7.7.2 KA7.2 KOMUNITA DATA STEWARDŮ

- Pracovní skupina pro národní spolupráci Datové podpory českých výzkumných institucí a vysokých škol.
- Cílem je aktivovat a podpořit osoby, které budou mít zájem o intenzivní participaci na výstupech projektu NRP.
- Komunita data stewardů není určena všem zájemcům o danou problematiku a budování ekosystému, ale pouze zaměstnancům institucí s určeným pracovním zařazením do Datové podpory.
- Organizace a podpora komunikace Datové podpory institucí a vysokých škol podpořených projektem i mimo projektové partnerství NRP.
- Průběžná zpětná vazba z užití metodických postupů NRP, komunikace ohledně zpětné vazby s dalšími KA.
- Pilotní ověřování kvality a relevance vzdělávacích materiálů bude probíhat prostřednictvím komunikace v rámci komunity data stewardů, v první řadě z institucí přímo podpořených NRP a následně dalších institucí, které zpětnou vazbu poskytnou. Tato komunita bude také sledovat potřeby cílové skupiny – výzkumných pracovníků při participaci na výstupech projektu NRP. Potřeby cílové skupiny budou zjišťovány také v rámci spolupráce s KA6.3, z výsledků mapování vědecké komunity. Instituce, které nebudou zřizovat plnohodnotné datově orientované podpůrné týmy pokryjeme sdružením takových institucí do oborových skupin a hledáním možnosti participativního systému podpory v rámci oborově spřízněných výzkumných institucí.
- Reflexe požadavků výzkumných týmů na naplnění požadavků poskytovatelů dotací.
- Reflexe výstupů a zapojení do spolupráce na výstupu KA5 Roadmapa dalších pokročilých služeb NRP/NDI a její průběžná aktualizace.
- Návrh modelových projektů pro organizační/personální zajištění principů EOSC v českých výzkumných institucích.
- Sdílení zkušeností ze zahraničí.
- Harmonogram: 2024–2028.

7.7.3 KA7.3 KONZULTAČNÍ SLUŽBY A ŠKOLICÍ SLUŽBY POKROČILÉ PODPORY NRP

- Helpdesk pokročilé podpory (level 2) pro software/služby NRP (spolupráce s KA6.4).
- Individuální konzultační služby pro instituce v oblasti využívání software/služeb NRP.
- Skupinové konzultace/akce pro instituce v oblasti využívání software/služeb NRP.
- Technická náročnost a komplexnost systémů NRP vyžaduje specializovanou podporu. S rozšiřováním uživatelské základny v návaznosti na vývoj systémů očekáváme nárůst dotazů a nejasností týkajících se využívání služeb NRP. Zatímco skupinové konzultační služby umožní efektivně sdělit informace o NRP většímu počtu osob v rámci instituce, což podpoří růst uživatelské základny; individuální konzultační služby umožní metodikům a data stewardům zaměřit se na specifické potřeby jednotlivých účastníků konzultací. Očekáváme celkem 250 účastí (indikátor 212 031) za dobu trvání projektu. Předpokládáme nárůst účastníků konzultací zejména v druhé polovině projektu po zprovoznění repozitářových systémů. Množství účastníků skupinových konzultací bude pravděpodobně vyšší než u individuálních konzultací.
- Návrh modelových projektů pro technické zajištění splnění principů EOSC v ČR.
- Konzultační podpora pro aplikace/e-návody k užívání software NRP.
- Zveřejnění materiálů prostřednictvím webu EOSC-CZ.