

Doporučení pro zakládání repozitářů v prostředí NRP

M. Janíček, I. Brejchová, D. Mikšík a kolektiv EOSC-CZ a NRP projektů

verze 0.9.1 z 19. 5. 2025

Tento dokument je primárně určen zástupcům uživatelských komunit uvažujícím o založení nové instance repozitáře v prostředí **Národní repozitářové platformy (NRP)**. Napomůže zorientovat se v celém procesu zakládání a provozování repozitáře a neopomenout žádný zásadní aspekt. Cílem je přiblížit jednotlivé činnosti, které jsou nedílnou součástí procesu vytváření repozitáře, včetně těch, za které zodpovídá zástupce zadávající uživatelské skupiny, nikoli NRP. Dokument pomůže informovaně se rozhodnout, zda bude repozitář v prostředí NRP tím správným řešením. Zároveň může také sloužit jako rozcestník na podrobnější materiály k jednotlivým činnostem a jako pomůcka při rozvrhování práce na novém repozitáři v případě, že se uživatelská skupina rozhodne nový repozitář v prostředí NRP vybudovat.

Dokument vznikl jako výstup klíčové aktivity 7.1 *Metodické materiály pro software-slужby NRP* projektu NRP a bude průběžně aktualizován.

Historie verzí

Verze	Datum	Popis
0.9	23.04.2025	První veřejná verze
0.9.1	19.05.2025	Korektura srozumitelnosti textu



Spolufinancováno
Evropskou unií



Národní repozitářová platforma
pro výzkumná data (NRP)
CESNET, zájmové sdružení právnických osob
Korespondenční adresa: Generála Píky 430/26, 160 00 Praha
info@eosc.cz; www.eosc.cz

Obsah

Úvod.....	3
1. Před založením repozitáře	3
1.1 Ujasněte si cíl navrhovaného repozitáře	3
1.2 Výběr vhodného repozitářového systému	5
2. Před implementací instance repozitáře.....	6
2.1 Data a metadata	6
2.2 Právní a etické aspekty.....	8
2.3 Administrace – vybrané aspekty	9
2.4 Uživatelské rozhraní	11
2.5 Zajištění FAIR	12
3. V průběhu implementace prototypu	14
3.1 Dokumentace a podpora uživatelů	14
3.2 Zaškolení.....	14
3.3 Testování funkčnosti repozitáře	14
4. Po uvedení do běžného provozu	15
4.1 Monitorování a hodnocení	15
4.2 Komunikace s veřejností	15

Úvod

Dokument [Podmínky vytváření nových a úprava stávajících oborových repozitářů v projektu Open Science II](#)¹ (dále též označováno jen jako *Podmínky*) popisuje tři základní případy užití NRP:

I. Vybudování repozitáře s použitím základních repozitářových systémů, tedy jako instanci CESNET Invenio, CLARIN DSpace nebo ASEP/ARL (bod B. výzvy OS II, tedy budování nových repozitářů).

II. Vybudování repozitáře s využitím alternativních repozitářových systémů v odůvodněných případech (bod B. výzvy).

III. Napojování stávajícího repozitáře provozovaného dosud mimo systémy NRP (bod A. výzvy, který se týká výhradně existujících repozitářů).

Pro každý z nich stanovuje, co bude poskytnuto infrastrukturou jako služba a co je zodpovědností správce repozitáře. V dokumentu, který právě čtete, se podrobněji věnujeme první možnosti. Dílčí části tohoto dokumentu mohou být přínosné i pro ty budující repozitáře v jiném repozitářovém systému, v takovém případě ale nemůžeme poskytnout žádnou metodickou podporu týkající se toho, jak můžete jednotlivé body realizovat. Správce existujícího repozitáře se může rozhodnout buď integrovat stávající řešení (viz kapitola *Integrace stávajícího samostatně provozovaného repozitáře do prostředí NRP/NDI*) nebo převést data do nové instance základního repozitářového systému v prostředí NRP (v takovém případě můžete přeskočit kapitolu 1. Před založením repozitáře v tomto dokumentu a pokračovat na kapitolu 2. Před implementací instance repozitáře).

Z *Podmínek* přebíráme také většinu definic základních pojmů, doporučujeme se s nimi proto seznámit před přečtením tohoto dokumentu.

1. Před založením repozitáře

Před samotným rozhodnutím založit repozitář v prostředí NRP by si zástupci vědecké komunity, které má repozitář sloužit, měli zkusit stručně zodpovědět otázky z kapitoly 1.1 a své odpovědi na ně zdokumentovat. Pokud se i po jejich zodpovězení rozhodnete repozitář založit, metodici vám pomohou s výběrem vhodného repozitářového systému (ze systémů nabízených v rámci NRP) a provedou vás jeho následnou implementací.

Předpokládáme, že tvorbu repozitáře budete chtít financovat z grantu, aktuálně přes [Open Science II](#)² výzvy. Doporučení z této kapitoly se s požadavky na projektovou žádost překrývají a doplňují. Záměr s námi můžete konzultovat i před podáním projektové žádosti.

1.1 Ujasněte si cíl navrhovaného repozitáře

Před tím, než s námi budete konzultovat, vaši volbu, poprosíme Vás, abyste písemně zpracovali následující otázky, které nám poslouží k ujasnění vašich požadavků a napomohou

¹ Matyska, L. et al. (2024). *Podmínky vytváření nových a úprava stávajících oborových repozitářů v projektu Open Science II*. <https://www.eosc.cz/media/3812767/podminky-vytvareni-novych-a-uprava-stavajicich-repozitaru-v30.pdf>

² Open Science II. *Projektová podpora Univerzita Karlova*. <https://opp.cuni.cz/OPP-135.html>

nám tak zvolit vhodné řešení. Využít můžete také [tuto šablonu](#)³. Vůbec nevadí, pokud zatím nebudete umět úplně a přesně zodpovědět všechny otázky.

Proč má být repozitář založen?

- Vznikne popis problému, který má repozitář řešit a deklarace cíle nového repozitáře.

Jaká data mají být ukládána?

- Definujte rozsah (scope) repozitáře včetně toho, jaké typy dat do repozitáře patří a která už ne. Víte už teď o standardizovaném metadatovém formátu, který by byl aplikovatelný?
- Pokuste se řádově odhadnout:
 - datový objem jednoho modelového záznamu;
 - počet záznamů v repozitáři při uvedení do standardního provozu;
 - počet nových záznamů za rok po uvedení do provozu.

Jaké vědecké komunitě má sloužit?

- Která instituce (případně konsorcium) zastřešuje vaši komunitu? Je ochotna se podílet na dlouhodobé udržitelnosti repozitáře?
- Stručně popište vědeckou komunitu, která bude tvořit jádro uživatelské skupiny a její institucionální ukotvení. Jaký bude stav při:
 - spuštění instance repozitáře (již jste předběžně domluveni na spolupráci);
 - po cca. třech letech (budete je aktivně zvát, aby do repozitáře přispívali a máte důvod předpokládat, že možnosti využijí, např. kolegové ze spřáteleného ústavu);
 - ideální výhledový stav (např. všichni evropští ornitologové).
- Popište skupiny uživatelů (např. studenti, lidé z byznysu, laická veřejnost) repozitáře a modelového uživatele dat v každé z nich.
 - Mají mít některé skupiny různou úroveň přístupu k datům?
- Popište, kdo bude mít oprávnění vkládat a editovat data.

Jak zapadá nová instance repozitáře do existujícího ekosystému?

- Existující už repozitáře, které můžete pro váš typ dat použít? Pokud ano, jaké jsou jejich výhody a nevýhody? V čem očekáváte výhodu u nového repozitáře v NDI?
- Popište roli nového repozitáře v kontextu relevantních existujících systémů a nástrojů (související repozitáře, výzkumná prostředí, agregační platformy atd.) a jak by s nimi mohl být provázaný.

Co musí repozitář umět, aby byl pro komunitu užitečný?

- Vypište seznam svých požadavků na systém a jak moc jsou pro vás prioritní, např. pomocí metody MoSCoW – nutné (must have), důležité (should have), volitelné (could have), nebude (won't have).

³ Účel a základní vlastnosti navrhovaného repozitáře. <https://www.eosc.cz/media/3841816/ucel-repozitare-sablona.docx>

1.2 Výběr vhodného repozitářového systému

Důsledně se v komunitě zamyslete nad svými odpověďmi. Je založení nového oborového repozitáře adekvátním řešením vašeho problému? Bude mít vaše komunita dostatečné institucionální a personální zajištění (více viz Podmínky), aby v rámci Národní repozitářové platformy vytvořila, nastavila a provozovala svou novou instanci repozitáře?

Pokud ano, je dalším krokem kontaktování metodiků [NRP](#)⁴, kteří vám pomohou s doladěním vašich požadavků a výběrem a následně implementací vhodného řešení (Invenio, DSpace, ASEP/ARL). Jednotlivé repozitářové systémy mají v zásadě stejné jádrové funkce, mohou se ale lišit v detailech způsobu implementace konkrétních funkcí a v míře, kterou si můžete dané funkce přizpůsobit. Proto vám některý ze systémů doporučíme až po tom, co prodiskutujeme vaše požadavky. Více se o nabízených systémech můžete dozvědět z jejich webu a dokumentace, vyzkoušet si můžete také uživatelské rozhraní stávajících instancí.

Invenio

- [Web Invenia](#)⁵
- [Dokumentace Invenio RDM](#)⁶
- [Národní datový repozitář](#)⁷
- [Národní metadatový adresář](#)⁸

DSpace

- [Web DSpace](#)⁹
- [Přehled funkcí DSpace](#)¹⁰
- [Repozitář LINDAT/CLARIAH-CZ](#)¹¹

ASEP/ARL

- [Dokumentace](#)¹²
- [Repozitář AV ČR](#)¹³

Pokud by vám nevyhovoval žádný ze základních repozitářových systémů NRP, můžete se také rozhodnout vybudovat instanci vlastního repozitářového systému na zdrojích NRP, nicméně v takovém případě budete nést i další náklady na tento proces (viz kapitola 5 *Vybudování repozitáře provozovaného na zdrojích NRP bez využití základních repozitářových systémů v Podmínkách*).

⁴ Kontaktujte nás. <http://www.eosc.cz/sluzby/ukladani/repozitare-v-nrp/kontaktujte-nas>

⁵ Invenio. <https://inveniosoftware.org/>

⁶ Invenio RDM: The turn-key research data management repository. <https://inveniordm.docs.cern.ch/>

⁷ Národní datový repozitář. <http://datarepo.eosc.cz>

⁸ Národní metadatový adresář. <https://nma.eosc.cz/>

⁹ DSpace: Build an Open Digital Repository. <https://dspace.org/>

¹⁰ DSpace 7.x Documentation: Functional Overview.
<https://wiki.lyrasis.org/display/DSDOC7x/Functional+Overview>

¹¹ Repozitář LINDAT/CLARIAH-CZ. <https://lindat.mff.cuni.cz/repository/>

¹² ASEP: Datový repozitář. <https://asep-portal.lib.cas.cz/uvod-do-asep/datovy-repozitar/>

¹³ ASEP: Repozitář AV ČR. <https://asep.lib.cas.cz/arl-cav/cs/index/>

2. Před implementací instance repozitáře

Jakmile se rozhodnete vybudovat instanci repozitáře v prostředí NRP a vyberete si konkrétní repozitářový systém, budete muset vypracovat detailnější specifikaci repozitáře, který se bude implementovat. Tato kapitola vám pomůže nezapomenout na nic zásadního.

2.1 Data a metadata

Metadata nám umožňují data strukturovaně vyhledat a popsat pro další uživatele, kteří potřebují data správně kontextualizovat a interpretovat. Ne vždy je ale jednoznačná hranice mezi daty a metadaty ani mezi různými typy metadat. V tomto dokumentu budeme za metadata považovat strukturovaný popis záznamu a daty budeme označovat soubory, které je možné k záznamu nahrát. (Meta)datový model bude tvořit základy vašeho repozitáře, je proto důležité jeho výběr pečlivě zvážit na základě hluboké znalosti dat, která se do něho mají ukládat.

V zásadě jsou dva typy repozitářů z hlediska dat a metadatového modelu:

a) Obecný repozitář s heterogenními daty

Metadatový model slouží jako obal nad libovolným souborem dat, který spadá do záběru repozitáře, jak si jej zadefinujete. Takovým modelem je CCMM (Czech Core Metadata Model) pro popis datasetů, jenž vychází z DataCite a DCAT-AP, a který používá Národní metadatový adresář (NMA) a Catch-all repozitář. Jeho [specifikaci naleznete na GitHubu](#).¹⁴

Pokud vaše odborná komunita potřebuje primárně ukládat datasety nebo potřebuje uložit velmi heterogenní data, může se rozhodnout znovupoužít implementaci Invenia, která už nad CCMM v NRP vzniká, případně jej jen rozšířit o oborově specifická pole (např. oborové řízené slovníky), s čímž vám jako infrastruktura pomůžeme, pokud nepůjde o rozsáhlé změny.

b) Specializovaný repozitář s homogenními daty

V datech se identifikují datové objekty, každý z nich má vlastní metadatový model. Pokud je datových objektů víc, určí se, jak mají být provázané.

Tato volba klade podstatně vyšší nároky na správce repozitáře, který musí být schopen navržený metadatový model naimplementovat z vlastních zdrojů. Infrastruktura poskytne technickou specifikaci jak na to pro každý ze základních repozitářových systémů. Silně doporučujeme nevymýšlet metadatový model sami, ale přebrat existující metadatové schéma, pokud už vhodné existuje. NRP poskytuje nástroj [DataSpecer](#)¹⁵ na specifikaci metadatových modelů. Umožní Vám generovat strojově čitelné schéma v JSON a XSD na základě vaší specifikace. Repozitářové systémy také umožňují definovat více metadatových modelů v jedné instanci repozitáře. Zároveň je možné data reprezentované v jednom schématu transformovat na libovolně mnoho jiných, pokud specifikujete jak – nutná bude minimálně konverze do CCMM, umožňující

¹⁴ Med, M. et al. (2024). Czech Core Metadata Model specification. GitHub. <https://github.com/EOSC-CZ/CCMM>

¹⁵ Dataspecer: A tool for management of semantic data specifications. <https://dataspecer.com/>

harvestování metadat do NMA. Vámi navržený metadatový model by měl usilovat o naplnění FAIR principů, které jsou podrobněji rozvedeny v kapitole 2.5 Zajištění FAIR.

Úplná specifikace (meta)datového modelu bude obsahovat minimálně strojové názvy polí a názvy polí pro uživatelské rozhraní, datový typ polí a informaci o povinnosti a opakovatelnosti polí. Konkrétní forma, ve které budete model specifikovat, je předmětem domluvy mezi vámi a metodikem vybraného repozitářového systému.

Kromě samotného metadatového modelu se doporučujeme věnovat i následujícím aspektům dat a metadat:

Kolekce

- Data a metadata v repozitáři je možné sdružovat do kolekcí. Zvažte, zda jich chcete využívat a pokud ano, popište, jak budou vznikat a k čemu budou sloužit.

Identifikátory

- Metadatové záznamy v repozitáři musí být opatřeny perzistentními identifikátory (PID)¹⁶. Na vás je vybrat vhodný typ PID (např. DOI, handle, ...) a určit jak a kdy se budou přidělovat. Všechny změny v datech nebo metadatach by měly být zaznamenávány, typicky v podobě vzniku nové verze nebo revize. Jednotlivé repozitářové systémy mohou k verzování a přidělování PID přistupovat mírně odlišně, metodici jednotlivých systémů vám detailně vysvětlí, jak daný systém k verzování defaultně přistupuje a co je možné přizpůsobit. Pravidelné zálohování dat (a metadat) na bezpečná místa zajišťuje infrastruktura.

Formáty

- Dál doporučujeme udělat analýzu stávajících souborových formátů, ve kterých jsou data uložena a zjistit, zda jsou vhodné pro dlouhodobou archivaci. Pokud ne, doporučujeme je před uložením konvertovat na vhodnější formát. Z toho pak bude vycházet seznam formátů, které jsou do repozitáře přijímané a které se zavazujete podporovat. Existence a vynucování takové souborové politiky vám výrazně ulehčí dlouhodobou správu repozitáře. Opřít se můžete o seznamy formátů vhodných pro dlouhodobé uchovávání dat, například ten od Library of Congress¹⁷. Doporučujeme využívat v co největší možné míře otevřených datových formátů.

Vkládání dat a metadat

- Data a metadata je možné do repozitáře vkládat přes formulář v uživatelském rozhraní nebo přes API. Řízení procesu ukládání dat podrobněji rozebíráme v kapitole 2.3 Administrace – vybrané aspekty. Doporučujeme API využít minimálně na začátku pro inicializaci prototypu repozitáře částí dat spolu s metadaty, což umožní testovat uživatelské rozhraní nad reálnými daty. Data a metadata je však nutné před importem strukturovat a pročistit. Stejným způsobem je možné data importovat i do hotového repozitáře při spuštění, aby data uživatelská komunita nemusela zadávat ručně.

¹⁶ Identifikátory.cz: Stránky o perzistentních identifikátorech. <https://identifikatory.cz/cs/>

¹⁷ Library of Congress. *Recommended Formats Statement*.
<https://www.loc.gov/preservation/resources/rfs/index.html>

Propojení s dalšími systémy

- Repozitáře mohou fungovat jako jeden z prvků v širším ekosystému oborových nástrojů a postupů pro práci s daty, proto je také důležité vědět s jakými externími systémy bude repozitář propojen a za jakým účelem. Mezi takovéto externí systémy můžeme řadit například elektronické laboratorní deníky (ELN), využití repozitářů pro automatické ukládání dat z přístrojů, nebo napojení repozitáře jako zdroje dat pro další nástroje (např. Jupyter notebooky). Repozitář typicky umožňuje přes API data nejen vkládat, ale také k nim přistupovat, vždy je však nutné každý přístup z jiných systémů individuálně posoudit a ověřit si technické i administrativní (např. právní) souvislosti.

2.2 Právní a etické aspekty

V rámci přípravy repozitáře je třeba uvážit i některé právní aspekty – provozovatel repozitáře musí být oprávněn ukládat a zpřístupňovat data a vkladatelé musí mít možnost určit, jakým způsobem mohou uživatelé s jejich daty dále nakládat (specifické licence, omezení přístupu a embarga apod.). Specifické požadavky na možnosti omezení přístupu k datům se mohou promítnout do celkového nastavení repozitáře a jeho workflows. V souvislosti s touto problematikou hovoříme o následujících termínech:

Depoziční licence

- Při ukládání obsahu (datové sady) do repozitáře je nutné, aby vkladatel poskytl provozovateli repozitáře oprávnění k užití datové sady, konkrétně k jejímu uložení a případnému zpřístupnění veřejnosti. Toto oprávnění vkladatel poskytuje prostřednictvím tzv. depoziční licence. Vzorová znění licencí poskytne infrastruktura.

Uživatelská licence

- Provozovatel repozitáře by měl definovat okruh tzv. uživatelských licencí, ze kterých budou moci vkladatelé volit při ukládání datové sady. Konkrétní podmínky užití datové sady uživateli z řad veřejnosti (jak ne/mohou uživatelé s datovou sadou nakládat) se bude řídit právě touto vkladatelem zvolenou a připojenou uživatelskou licencí. Zpravidla půjde o varianty veřejných licencí [Creative Commons](https://creativecommons.org/licenses/)¹⁸, okruh nabízených licencí však může být širší – závisí na potřebách konkrétní instituce/provozovatele repozitáře. V úvahu připadá i možnost vložení datové sady do repozitáře bez uživatelské licence.

Datová sada bez připojené uživatelské licence

- V případě, že k datové sadě nebude připojena žádná uživatelská licence (umožňuje-li vůbec repozitář takovou možnost), bude datová sada v režimu „all rights reserved“. Prakticky to znamená, že je datová sada uživatelům k dispozici pouze „ke čtení“ a nad rámec čtení ji lze užívat pouze na základě výjimek z autorského práva stanovených zákonem (např. citovat výňatky, užívat pro osobní potřebu). Takovou datovou sadu nebude možné dále sdílet, ledaže by uživatel získal výslovný souhlas od osoby, která vykonává autorská majetková práva k datové sadě (typicky autor nebo zaměstnavatel).

¹⁸ Creative Commons. *About CC Licenses*. <https://creativecommons.org/share-your-work/cclicenses/>

Další možná omezení přístupu uživatelů k datové sadě

- Kromě podmínek zvolené uživatelské licence může být zpřístupnění datové sady veřejnosti omezeno také:
 - odkladem zpřístupnění (embargem);
 - volbou režimu dostupnosti datové sady v repozitáři, např. omezením okruhu uživatelů pouze na osoby z určité instituce;
 - podmíněním přístupu k datové sadě uzavřením licenční smlouvy/NDA;
 - podmíněním přístupu k datové sadě na základě hesla/schválení vkladatelem.

Data ukládaná v repozitáři mohou v závislosti na míře jejich citlivosti vyžadovat různou úroveň ochrany z hlediska zabezpečení a přístupu. Jde o širokou škálu: od dat, která mohou být zpřístupňována neomezeně komukoliv, přes data, která nejsou určena ke zveřejnění, až po data, jejichž zpřístupnění neoprávněným osobám může způsobit značné škody. V procesu přípravy repozitáře je třeba důkladně zvážit jaká je citlivost dat, která do něj budou ukládána a jakým způsobem bude nastavena bezpečnost práce s nimi.

Další významnou otázkou je uchování a případně zpřístupnění dat, které jsou osobními údaji. Zpracování tohoto typu dat se řídí právními předpisy na ochranu osobních údajů, zejm. obecným nařízením o ochraně osobních údajů (GDPR)¹⁹, z něž vyplývají konkrétní zásady a požadavky, které musí být bezpodmínečně naplněny.

Důkladně zvažte, jestli data, která do repozitáře plánujete ukládat, zahrnují citlivá data a/nebo osobní údaje a jaká technická a organizační opatření budete muset při jejich ukládání a zpřístupnění provést.

2.3 Administrace – vybrané aspekty

V této části se budeme věnovat jen vybraným aspektům správy repozitářů, jmenovitě rolím, workflows, dlouhodobému financování provozu repozitáře a správě uživatelů a jejich skupin. Ostatní základní koncepty repozitářů – záznam, komunita ad. – jsou popsány v kapitole *Obecné principy NRP* v dokumentu *Podmínky*.

Role

- Uživatelé přistupující do repozitáře mají přiděleny **role**, které spoluurčují, které vlastnosti a funkcionality repozitáře jim budou dostupné.
- Pro příklad uvádíme výpis obvyklých rolí, v případě potřeby je lze rozšířit
 - anonymní čtenář – nepřihlášený uživatel;
 - člen komunity – přihlášený uživatel, může např. číst záznamy, které nejsou přístupné nepřihlášeným uživatelům;
 - vkladatel (submitter) – přihlášený uživatel s oprávněním tvořit (v definovaném kontextu) v repozitáři záznamy;

¹⁹ Nařízení Evropského parlamentu a Rady (EU) 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (obecné nařízení o ochraně osobních údajů). <https://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:32016R0679>

- schvalovatel (curator) – uživatel kontrolující správnost a úplnost záznamů před publikováním (přijetím do repozitáře);
- administrátor – uživatel s neomezenými oprávněními zodpovídající za repozitář po technické stránce.
- Role řídicí oprávnění v instanci repozitáři nejsou nutně totožné s rolí osob, kteří zajišťují chod repozitáře. Odpovědnosti rolí správce repozitáře a datový kurátor jsou detailně popsány v dokumentu *Podmínky*. Roli správce repozitáře typicky vykonávají fyzické osoby, naproti tomu instituce vystupují typicky v roli partnera či smluvní strany provozovatele infrastruktury (CESNET/e-INFRA.cz).

Workflows

- Workflow je algoritmizovaný postup vytvoření nového záznamu nebo pro přístup k existujícímu záznamu v repozitáři.
- Typické jednoduché workflow pro vytvoření nového záznamu (příp. „depozici dat“) zahrnuje:
 - vytvoření návrhu záznamu (draft);
 - odeslání návrhu ke kontrole a schválení (review) datovým kurátorem;
 - schválení návrhu – záznam je zveřejněn;
 - neschválení návrhu – záznam je vrácen vkladateli k opravě.
- Workflows lze doplňovat a upravovat podle potřeby – rozsáhlé úpravy mohou vyžadovat vývojové kapacity na straně uživatelské skupiny.
- Má-li vkládání iniciovat nejen uživatel-fyzická osoba, ale také uživatel-stroj, modelově celý scénář popište a konzultujte s metodikem příslušného systému.

Financování, udržitelnost

- Jak je/bude provoz repozitáře financován? Jak vyplývá z textu *Podmínky*, správce repozitáře musí být k dispozici během celého životního cyklu repozitáře, role datový kurátor a uživatelská podpora „musí existovat a být personálně obsazeny po celou dobu produkčního provozu repozitáře“. S tím jsou spojeny náklady, s nimiž je nutné dlouhodobě kalkulovat. Je přirozené, že zdroje financování se budou v čase měnit a že v okamžiku zakládání repozitáře nemá uživatelská skupina k dispozici přesné informace o tom, jaké zdroje na pokrytí těchto nákladů použije v dlouhodobém horizontu. Doporučujeme držet tento aspekt provozování repozitáře jako trvalé téma strategických i provozních rozhodnutí uživatelské skupiny stojící za daným repozitářem – stabilní personální zajištění těchto rolí není pouze otázka financování, ale také funkčního zasazení osob pečujících o repozitář do struktur institucí, které je zaměstnávají.
- K dlouhodobé stabilitě repozitářů přispívá zapojení institucí do jejich provozu, a tím i financování. Lze tedy doporučit, aby uživatelská skupina začala co nejdříve hledat možnosti, jak do péče o zakládaný repozitář zapojit instituce, které mají na jeho provozu aktuální či potenciální zájem. Projekt NRP k tomu poskytne základní modely dlouhodobě udržitelného provozování repozitářů (předpokládaný termín Q2/2025).

Uživatelé, účty, skupiny

- Projekt NRP předpokládá, že výchozím (nikoliv nutně jediným) zdrojem informací o uživateli a skupinách bude autentizační a autorizační infrastruktura EOSC CZ

(EOSC CZ AAI). Očekává se proto, že se správce repozitáře (nebo jím pověřená osoba) seznámí s architekturou EOSC CZ AAI a propojením s repozitáři, který bude popsán v připravovaném dokumentu Iniciální architektura AAI pro NRP, a bude skupiny a uživatele řídit v prostředí Perun AAI.

2.4 Uživatelské rozhraní

Repozitáře podpořené z prostředků NRP mají povinnou publicitu, loga a pravidla týkající se jejich umístění dostanete k dispozici. Jak je správně použít bude uvedeno v připravovaném design manuálu. Kromě toho tam najdete i podmínky pro provoz repozitáře na poddoméně eossc.cz (pokud byste se pro tuto variantu rozhodli) a základní doporučení týkající se přístupnosti.

Pro Catch-all datový [repozitář](#), který se implementuje v Inveniu, pracujeme na vizuálu uživatelského rozhraní, ten pak může sloužit jako šablona, kterou si můžete následně přizpůsobit. To se bude hodit zejména pokud bude váš metadatový model rozšířením CCMM. Na vás bude určit barevné schéma repozitáře, logo repozitáře a doprovodné texty.

Drobné změny pro vás uděláme my, budete však muset přesně komunikovat co a jak se má upravit. V případě větších zásahů si budete muset zaplatit vývojáře z vlastních prostředků, který vámi specifikované požadavky naimplementuje. Co už je velká změna a co ještě ne s vámi předem detailně projednáme.

Čím víc se vámi vybraný metadatový model bude lišit od existujících šablon a čím rozsáhlejší a komplikovanější váš metadatový model bude, tím náročnější bude proces tvorby uživatelského rozhraní a tím víc vlastních kapacit budete potřebovat na tuto aktivitu dedikovat.

Uživatelské rozhraní jakéhokoli repozitáře se bude typicky skládat minimálně z následujících stránek na jejichž podobě bude potřeba se domluvit:

- Domovská stránka
- Stránka s metadatovým záznamem a možností stáhnout data
- Procházení (browse) a vyhledávání (search):
 - Podle jakých atributů metadat se mají záznamy filtrovat?
 - Podle jakých atributů metadat se mají záznamy řadit?
 - Jaké atributy metadat se mají zobrazit na kartě záznamu?
- Editační rozhraní a depoziční formulář:
 - Jaké widgety se budou používat při zadávání metadat? (drop down/autocomplete atd.);
 - Uzpůsobení dle speciálních požadavků vycházející z depozičního workflow.

Pokud chcete uživatelům poskytnout jiné náhledy nad daty, musíte se informovat u metodika repozitářového systému, zda takový náhled repozitářový systém defaultně podporuje. Pokud ne, budete si muset takové rozšíření naprogramovat z vlastních zdrojů jako další aplikační vrstvu nad daty. Výpočty nad daty už jdou typicky nad rámec aplikace repozitáře, k těmto účelům budou k dispozici nástroje pro přenos dat do prostředí e-INFRA CZ, kde jsou pro tento účel k dispozici specializované zdroje pro výpočetní operace nad takto připravenými daty.

Silně doporučujeme, aby byl repozitář primárně v angličtině. Musíte si ale rozmyslet, jaké další jazykové mutace repozitář bude muset mít a do jaké míry bude obsah přeložen:

- a) vícejazyčně bude domovská stránka a informace o platformě;
- b) vícejazyčně budou také návěští polí a taxonomie;
- c) vícejazyčně budou i textová pole v metadatových záznamech.

Metodici vás seznámí s podrobnostmi toho, jak je v jejich repozitářovém systému překlad obsahu konkrétně řešen. Různé varianty překladu jsou různě komplikované na realizaci, výběr tedy musíte konzultovat s metodikem příslušného systému. Pokud se rozhodnete pro více jazykových mutací, musíte dodat všechny překlady a zodpovídáte za to, že budou při změnách aktualizovány. V případě varianty c) musíte navrhnout, jak budou překlady dodávat uživatelé a za jakých okolností se přeložený obsah bude zobrazovat.

2.5 Zajištění FAIR

Podpora FAIR principů v repozitáři se protíná s jeho základními funkcemi. Nicméně vzhledem k významu FAIR principů v současném ekosystému výzkumných dat doporučujeme se podívat na podporu jejich naplnění i samostatně.

NRP poskytuje i další nástroje pro podporu tvorby výzkumných dat naplňujících FAIR principy, podrobnosti naleznete v dokumentu [Analýza a návrh postupů k naplňování FAIR principů v rámci NRP](#)²⁰, kde jsou také v rámci tabulky podrobně vysvětleny jednotlivé principy FAIR. Před začátkem plánování implementace doporučujeme ověřit, zda neexistují vhodné, či alespoň blízké oborové FIP (FAIR integration profiles). K dispozici budete také mít FIP wizard, nástroj na deklaraci konkrétních řešení, pomocí kterých jste se rozhodli jednotlivé aspekty FAIR principů naplnit. Každý repozitář by měl za svoji uživatelskou komunitu vytvořit a aktualizovat FIP záznam, na základě kterého bude hodnocena míra naplnění FAIR principů.

(Nalezitelnost) Findability

- Identifikátory
 - Jaký bude váš repozitář používat perzistentní identifikátor?
- Metadatové schéma
 - Je třeba využít oborové metadatové schéma a pokud ano, existuje už, nebo je teprve budete muset vytvořit? Jakým způsobem se bude mapovat na CCMM (pro napojení na NMA), případně na schéma DataCite (nutné při využití identifikátoru DOI)
- Budou pro popis datových objektů využívány řízené slovníky, či ontologie? Jaké? Bude případně možný popis (volně zvolenými) klíčovými slovy?
- Do jakých systémů (kromě NMA) by se měla metadata dále poskytovat?

(Přístupnost) Accessibility

- Protokoly
 - Jaké protokoly budou využívány pro přístup k datům? Bude potřeba využít jiné protokoly nad rámec obvyklého https?

²⁰ Trtíková, I., Bártková, P., Biedermannová, L., Jindřich Fejfar, Robert Pergl, David Novák, & Tomáš Čížek. (2025). *Analýza a návrh postupů k naplňování FAIR principů v rámci NDI/NRP*. <https://doi.org/10.71495/5dv6-v116>

- Autentizace a autorizace
 - Jak bude řešena autentizace a autorizace? Bude využit pouze e-INFRA AAI, nebo ještě další autentikační a autorizací systém?
 - Jak bude popsán přístup k datům s omezeným přístupem, respektive mechanismus přidělování přístupu?

(Interoperabilita) Interoperability

- Formáty dat
 - Jaké budou využívané/preferované formáty dat? Jsou všechny otevřené?
 - Struktura dat
 - Existuje oborově definovaná struktura dat/metadat, kterou by bylo možno využít?
 - Jsou datové objekty, které budou ukládány do repozitáře už v této struktuře vytvářeny?
- Vazby na jiná data a metadata
 - Jaké typy vazeb mezi daty je potřeba zvážit („odvozené od“, „cituje“, „je součástí“ apod.)
- Dokumentace
 - Jaký je vhodný formát dokumentace, který lze ukládat společně s datovým objektem? Je struktura dat dostatečně popsána?

(Znovuvyužitelnost) Reusability

- Licence
 - Jaké licence jsou vhodné pro vaše datové objekty? Je třeba pracovat se speciálními (dle potřeby definovanými) licencemi?
- Provenance
 - Jakým způsobem je vhodné zachytit provenanční informace? (historie vzniku a práce s datovým objektem, použité přístroje a nástroje apod.) Jak podrobné provenanční informace jsou pro vaše data relevantní?

3. V průběhu implementace prototypu

3.1 Dokumentace a podpora uživatelů

Správce repozitáře vytváří uživatelskou a technickou dokumentaci, instruktážní materiály pro vkladatele (např. jak vkládat data, jak vyplnit metadata) – NRP poskytuje šablony dokumentace a technický obsah, stejně jako nástroj pro tvorbu a udržování dokumentace.

Uživatelům podílejícím se na ověřování prototypů repozitáře je vhodné sdělit, jak mají správci repozitáře předávat zpětnou vazbu. Správce repozitáře by tedy měl jednak určit komunikační kanály, jednak odpovědnost za zpracování uživatelských požadavků – tuto činnost lze využít jako první krok k nastavení rutinní uživatelské podpory, kterou bude správce repozitáře zajišťovat po celou dobu existence repozitáře (viz Podmínky).

3.2 Zaškolení

Provozovatel repozitářového systému předá správci repozitáře technickou dokumentaci repozitáře v rozsahu dostatečném pro běžnou práci správce repozitáře, datového kurátora, vývojáře uživatelské skupiny. V případě otázek se můžete obrátit na metodika příslušného systému.

3.3 Testování funkčnosti repozitáře

Před nasazením na produkci je potřeba otestovat, že výsledný repozitář uspokojivě odpovídá specifikaci a potřebám zadávající uživatelské komunity. Správce repozitáře (nebo někdo jím pověřený) by si měl proklikat náhledy metadat, formuláře pro zadávání metadat, vyhledávání a případné další funkce dělané na míru repozitáři. Případné připomínky předá metodikovi repozitářového systému na vyhodnocení a případné zapracování. Počítejte s tím, že testování pravděpodobně bude probíhat ve více iteracích. Můžete vyhodnotit, že by bylo přínosné, aby některé části systému připomínkovali i koncoví uživatelé a odhalily se tak pro ně neintuitivní a nesrozumitelné aspekty instance repozitáře. Je možné se obrátit na ux@eossc.cz s žádostí o pomoc s přípravou uživatelského testování nebo ohledně nasazení nástroje pro monitoring chování uživatelů na stránce.

4. Po uvedení do běžného provozu

4.1 Monitorování a hodnocení

Význam a užitečnost repozitáře lze mj. měřit strojově zpracovanými statistikami. Každý repozitářový systém umožňuje sledovat některé údaje na úrovni celého repozitáře, komunit či kolekcí, až po jednotlivé záznamy, další data o využívanosti repozitáře lze získat z externích analytických nástrojů. Základní parametry sledované přímo repozitářovými systémy mj. zahrnují:

- kolekce/komunity – počet registrovaných uživatelů, rozložení registrovaných uživatelů dle jejich afiliace, počet záznamů v kolekci, nové záznamy od definovaného data, množství uživatelských dat, počet zobrazení metadat z kolekce, počet stažení datových sad z kolekce, rozložení typů licencí;
- jednotlivé záznamy – počet zobrazení metadat, počet stažení datové sady ze záznamu, objem dat.

Pokud jsou pro váš repozitář relevantní jiné údaje, je vhodné je podrobně definovat a popsat při specifikaci repozitáře, tak aby byly nástroje pro jejich sledování od začátku zahrnuty do vývoje repozitáře.

Pokud bude možné do repozitáře vkládat data bez individuální detailní kontroly každého záznamu (např. s využitím automatické kontroly vstupních polí), doporučujeme provádět pravidelnou kontrolu kvality dat i metadat.

U repozitářů, které jsou propojeny s externími systémy (např. metadatové modely, řízené slovníky), je naprosto nezbytné průběžně sledovat a vyhodnocovat změny zdrojových dat.

Repozitáře poskytují různé služby různým skupinám uživatelů, a protože společenský, technický i oborový kontext se neustále mění, je také vhodné průběžně sbírat přímou zpětnou vazbu od různých skupin uživatelů, vyhodnocovat ji, a následně využívat pro iterativní úpravy repozitáře.

4.2 Komunikace s veřejností

Po předání repozitáře do užívání uživatelské skupině, dokončení dokumentace a nastavení komunikačních kanálů pro podporu uživatelů je vhodné o nově dostupném repozitáři informovat odbornou i laickou veřejnost. Doporučujeme využít služeb oddělení vnějších vztahů na institucích s vazbou na repozitář – obvykle mají tato pracoviště vytvořenu síť formálních i neformálních vztahů a postupů pro efektivní šíření informací o vědeckých výstupech. Obrátit se můžete také na Oddělení vnějších vztahů EOSC CZ (pr@eosc.cz). Společně připravená sdělení pak mohou členové uživatelské komunity samozřejmě šířit i prostřednictvím svých osobních kanálů.