

Obsah

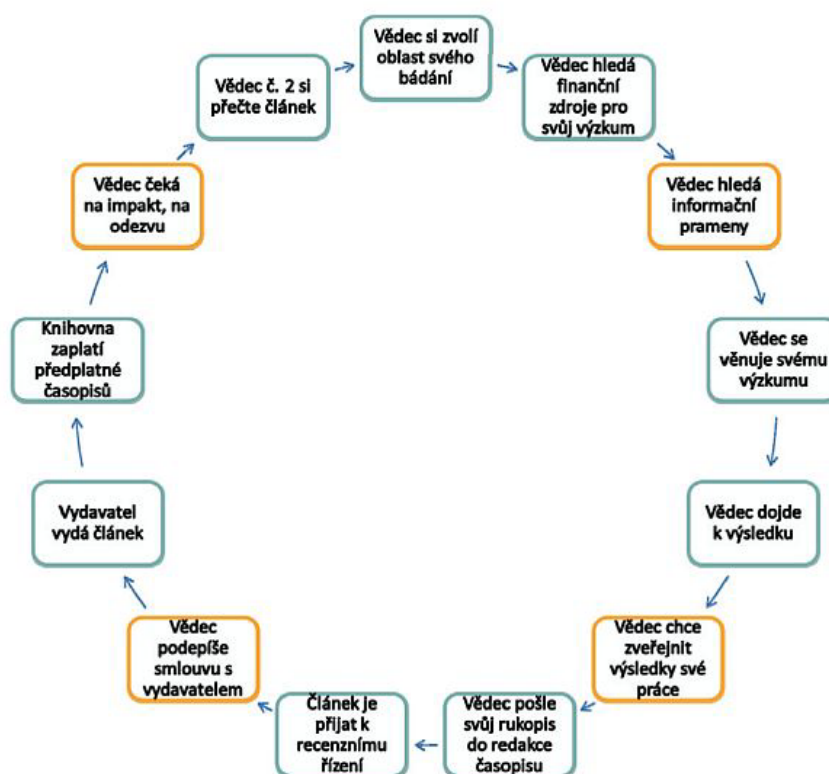
Obsah.....	2
Úvod.....	3
Vytvoření dat	5
Zpracování dat	7
Analýza dat.....	11
Uchovávání dat	13
Zpřístupnění dat	16
Opětovné využití dat	19
Závěr	22
Seznam použité literatury.....	23

Úvod

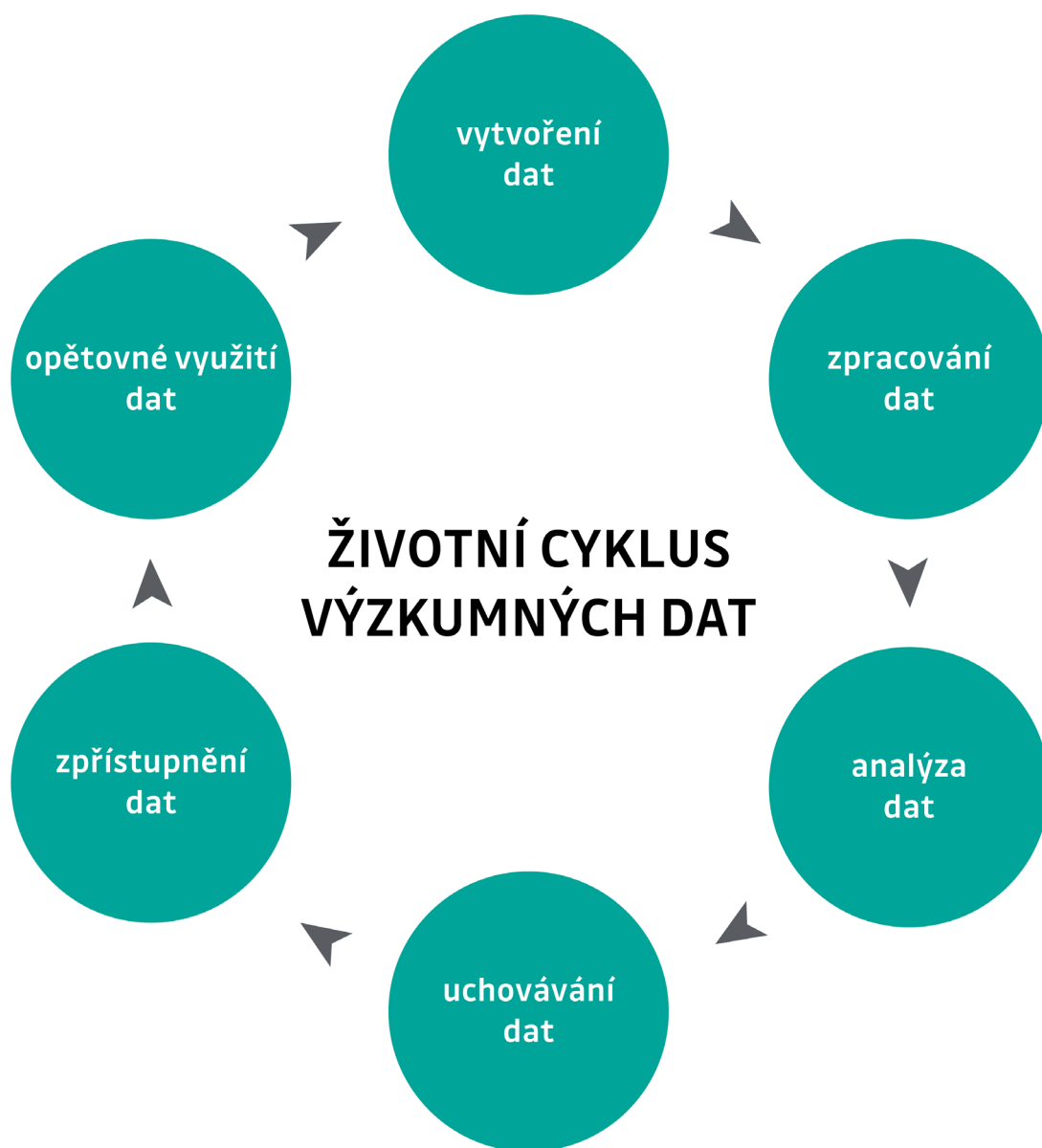
Předložená metodika se snaží čtenáře seznámit se základními a „obecnými principy“ při práci s výzkumnými daty v rámci politiky Open Science, což je úkol značně náročný, protože každý vědecký projekt je jedinečný a vyžaduje individuální přístup. Následující text předkládá pouze některá pravidla a postupy, které by měly výzkumníkům pomoci při aplikaci teoretických požadavků politiky otevřené vědy do praxe. Rozhodně nelze tvrdit, že po přečtení následujících stránek budou vědci schopni aplikovat všechny principy pro otevření dat. Vždy narazí na některé konkrétní problémy, při jejichž řešení budou muset požádat o pomoc odborného pracovníka, tzv. Data Stewarda nebo Open Science metodika (či Metodika/Metodičku open science), aby mohli splnit podmínky uložené zadavatelem projektu.

Dalším zásadním problémem je rychlý rozvoj v oblasti politiky Open Science, který se promítá také do její praktické aplikace. Následující text se sice pokusí zachytit všechny nejnovější a nejaktuálnější postupy, přesto se nejedná o jeho definitivní znění, ale pouze o druhou aktualizovanou verzi, jež bude v budoucnosti postupně doplňována a rozšiřována o další informace, aby poskytovala co možná nejucelenější a nejpraktičtější obraz o aktuálním stavu politiky Open Science na VŠB-TUO.

Předložený text má především usnadnit práci Data Stewardů v uvedené instituci, jelikož popsané principy se budou opakovat u každého výzkumu uplatňujícího politiku Open Science. Po přečtení předložené metodiky budou moci výzkumníci konzultovat pouze menší individuální problémy během zveřejňování svých dat. Pro snazší a přehlednější pochopení principů Open Data bude následující text uspořádán podle tzv. **šestifázového životního cyklu dat**, který je založen na koloběhu publikačního procesu.



Koloběh publikační činnosti [1]



Šestifázový životní cyklus dat [2]

Vytvoření dat

Již při vytváření dat se každý výzkum stává unikátním, a tudíž nelze na něj aplikovat určité obecné principy, které by zajistily jeho správné provedení. Přesto existuje několik vhodných doporučení, které by odborníkům mohly pomoci při otevírání získaných dat.

Před zahájením každého vědeckého projektu v rámci Open Science je vždy povinné vytvořit Data Management Plan (někdy se překládá do češtiny jako Plán managementu dat, dále jen DMP), což je dokument na několik stránek, jehož struktura je většinou pevně stanovena již v samotném zadání projektu (např. Program Kappa v rámci TA ČR vyžaduje předložení první verze DMP do 6 měsíců po zahájení projektu [3]). Je samozřejmé, že během výzkumu se objeví značné množství problémů a neočekávaných skutečností, o nichž před zahájením práce neměli výzkumníci ani tušení, a proto je DMP tzv. „živý dokument“, který by se měl pravidelně aktualizovat během samotného výzkumu (nejlépe s každou průběžnou zprávou o průběhu výzkumu by měla být předložena i nová verze DMP). Poslední revizi je nutné provést po skončení celého projektu. Nicméně i přesto je nutné vytvořit hned na začátku výzkumu první verzi, která by v oblasti otevřených dat měla obsahovat [4]:

1. **Data Summary/Shrnutí dat**
2. **FAIR data/FAIR data**
3. **Other research outputs/Další výstupy výzkumu**
4. **Allocation of resources/Alokace zdrojů**
5. **Data security/Bezpečnost dat**
6. **Ethics/Etika**
7. **Other issue/Ostatní**

Konkrétní podobu DMP určuje sám zadavatel projektu, který by měl výzkumníkům poskytnout formulář se všemi klíčovými body a otázkami a vědci by do něj měli pouze doplnit konkrétní údaje a odpovědi. I když každý zadavatel má svou vlastní verzi DMP, výše uvedené rubriky by se měly nacházet v každém formuláři, které jsou si díky tomu značně podobné.

Podrobný rozbor jednotlivých částí obsahu DMP je nad možností předložené metodiky, a proto se následující text zaměří pouze na bod 2., konkrétně **jak budou zajištěny FAIR principy?** I když je nezbytná spolupráce s Data Stewardem, který musí vždy zkontrolovat a případně opravit data před jejich zveřejněním, mohou již samotní výzkumníci na základě několika jednoduchých pravidel zajistit, že úpravy dat před zveřejněním budou jen minimální:

- a) Findable (dohledatelnost) = Repozitáře většinou automaticky přidělují perzistentní identifikátory (dále jen PID) k vloženým a zveřejněným datasetům. Nejběžněji se používají **DOI** (Digital Object Identifier) a **Handle**. Existuje i speciální typ perzistentního identifikátoru pro rozpracovaný dataset **ePIC** (Persistent Identifiers for eResearch) [5]. Dalšími důležitými identifikátory jsou také **ORCID** (Open Researcher and Contributor ID) pro jednoznačné ztotožnění konkrétního výzkumníka a **ROR** (Research Organization Registry), který identifikuje vědeckou instituci. Výzkumníci také mohou, ale nemusí, vytvořit skupinu klíčových slov nebo napsat krátký popis.

- b) Accessible (zpřístupnění dat) = výzkumníci mohou, ale nemusí, vytvořit skupinu klíčových slov nebo napsat krátký popis v libovolném jazyce. Další metadata jsou uložena v samotném datasetu v souboru **readme** (viz 2. Zpracování dat).
- c) Interoperable (interoperabilní) = **Formáty souborů**, ve kterých budou hodnoty zveřejněny. Vhodné jsou běžně používané formáty (např. pdf, doc, xls, txt, jpg atd.). Rovněž je možné ke speciálním typům souborů zveřejňovat kód (nástroje), po jejichž stažení a instalaci bude možné otevřít soubory na kterémkoliv elektronickém zařízení. Popsaný postup by měl být aplikován v rámci tzv. oborových repozitářů (viz 5. Zpřístupnění dat).
- d) Resuable (opětovně použitelná) = Před zveřejněním dat je nutné si vybrat **typ licence** a její působnost (např. mezinárodní, platící jen pro ČR atd.; viz 6. Opětovné využití dat). Každý výzkumník se musí rozhodnout, jaká data otevře a zda u některých nastaví omezení (viz 3. Analýza dat).

Pokud se tvůrci rozhodnou, že některá data **nebudou zveřejňovat**, musí své důvody vysvětlit již na začátku projektu v DMP (porušení GDPR, citlivá data atd.).

Pro usnadnění přípravy Data Management Planu již na internetu existují nástroje. Autorovi jsou kladeny otázky, na něž se snaží co nejpřesněji odpovědět, čímž postupně vytváří DMP. Také má možnost podívat se na starší verze jiných projektů a použít je jako inspiraci. U některých nicméně mohou nastat potíže při **registraci**. Pokud se chce výzkumník přihlašovat v rámci své organizace, musí ji nejdříve vybrat v přednastaveném seznamu. Jelikož všechny nejpoužívanější nástroje jsou provozovány mimo území ČR, jejich databáze vždy neobsahují české instituce, a proto je nutné se nejdříve obrátit na jejich technické podpory a požadovat zavedení nové položky, což celý proces značně prodlužuje. Mezi nejznámější příklady uvedených online nástrojů patří:

1. Argos ([Argos \(openaire.eu\)](https://argos.openaire.eu))
2. Data Stewardship Wizard ([Data Stewardship Wizard \(ds-wizard.org\)](https://ds-wizard.org))
3. DMP Online ([DMPonline \(dcc.ac.uk\)](https://dmponline.dcc.ac.uk))
4. DMPTool ([DMPTool](https://dmptool.org))

Posledním zásadním problémem je, kde se má první a všechny ostatní verze DMP zveřejnit. I když dosud neexistuje jednoznačná odpověď, přesto není vhodné jej zveřejňovat na internetu ve zmíněných nástrojích. Data Stewardi zatím doporučují vytvořit „**administrativní**“ **dataset**, ve kterém budou ve zvoleném repozitáři zveřejněny všechny dokumenty řešící otázky správy dat a budou zde i průběžně aktualizovány.

Tvorbu popsaného datasetu by měli zajišťovat administrativní pracovníci vědeckého projektu nebo jeho hlavní řešitel (uvedený termín používá Univerzita Karlova a definuje jej jako výzkumníka, který má odpovědnost za celý výzkumný projekt) [6] nebo administrátor projektu, přičemž vždy musí být obsah zveřejněných souborů zkontrolován s Data Stewardem.

Zpracování dat

Při zpracování informací je důležité rozlišení pojmů **popis** a **organizace dat**, aby bylo možné zajistit požadované FAIR principy a také usnadnit spolupráci mezi jednotlivými výzkumnými skupinami v rámci projektu.

Organizací dat by se výzkumníci měli zabývat již při přípravě DMP (viz 1. Vytvoření dat), přičemž se všichni autoři musí domluvit na společných postupech, jelikož výsledky jejich výzkumů musí být jednotně uspořádané a přehledné. **Popisování souborů a jednotlivých složek by vždy mělo být pro autora co nejvíce přirozené, aby se mu se soubory co nejsnáze pracovalo.** Aby byl dataset následně srozumitelný i pro jiné výzkumníky, musí být právě opatřen vyplněným souborem **readme**.

Pokud si tvůrci dat nejsou jistí, že jejich systém popisování je všem srozumitelný, mohou se inspirovat některými obecně užívanými pravidly. Vždy by do systému měly být zahrnuty specifické požadavky konkrétních vědních oborů a cílů výzkumu a hledat nejvhodnější pracovní postupy. Pro popisování výzkumných dat se „doporučuje“:

- Pro chronologické řazení je nejlepší používat datum ve formátu **YYYYMMDD**.
- Pokud autoři nepoužívají nástroje, které umožňují automatické verzování (např. Open Science Framework, dále jen OSF), je nutné uvést ručně verzi souboru.
- Není vhodné **používat některé znaky** (např.: @ % * aj.), protože některé systémy popsane symboly považují za pokyny k provedení určitých činností.
- Jestli chtějí někteří výzkumníci odlišit své soubory, mohou do jejich názvu připsat své **iniciály nebo jiný specifický symbol**. Podobný postup lze použít také pro odlišení respondenta, který provedl poslední změny v datasetu.
- Totéž platí, když vědec pracuje **na více projektech současně** a chce si své informace roztrždit. (V uvedeném případě je mnohem praktičtější data roztrždit do samostatných složek, aby se zbytečně neprodlužovaly názvy souborů – poznámka autora metodiky).
- Důležité je také vybrat si **jazyk** pro popisovaná data (angličtina pro snazší komunikaci se zahraničím nebo mateřský jazyk instituce)
- Když výzkumníci používají různé typy souborů, mohou je rozlišovat **předem domluvenými identifikátory** (např. měření, dotazník, test atd.).
- Všichni autoři by se také měli domluvit na **pořadí jednotlivých částí popisu a rozdělovačích** (např. podtržítka, pomlčka aj.). (Při popisování souborů je nutné mít vždy na paměti, že by žádné dva soubory neměly mít stejné názvy, aby nevznikly potíže během ukládání do repozitáře (viz 5. Zpřístupnění dat) – poznámky autora metodiky).

Na základě uvedených informací lze sestavit konkrétní metodiku pro popisování dat v rámci jednoho projektu, včetně tabulky s přehledem všech využívaných zkratk a symbolů pro lepší přehlednost, která by mohla vypadat:

Pojmenování souborů: **VY1_001_234_XXXXXX_Y_002_v3**

- **VY1** – výzkumný záměr (3 znaky)
- **001** – pořadí publikace v rámci výzkumného záměru (3 znaky)
- **234** – poslední tři čísla nadřazeného indikátoru (3 znaky)
- **XXXXXX** – zkratka instituce (VSB-TUO, VSCHT aj.) (max. 7 znaků)

- **Y** – typ souboru (1 znak, např. M = Manuscript; D = Data; S = Supplementary information; PU = Publisher version atd.)
- **002** – pořadí souboru v rámci datasetu (3 znaky, protože např. v repozitáři Zenodo může jeden dataset obsahovat maximálně 100 souborů)
- **v3** – verze datasetu (2 znaky) [7]

Uvedený systém představuje propracovanější postup popisu dat v datasetu většího rozsahu a propracované struktury (zhruba 50–100 souborů). Pokud chce výzkumník zveřejnit výzkumná data do 50 položek, stačí mnohem jednodušší popis s menší počtem zkratek (např.: **HC_M1.txt**, kde **HC** znamená hořecnatý cement a **M1** měření první).

Při vytváření názvů souborů je nutné také uvažovat o způsobu, jakým budou budoucí data uspořádána na interních discích a po zveřejnění ve veřejném uložišti. Výzkumník sice může mít své soubory rozdělené přirozeně do jednotlivých složek, nicméně některé repozitáře ukládají pouze soubory beze složek. V uložišti tak vznikne pouze dlouhý seznam jednotlivých názvů souborů a jejich autor by proto měl dávat pozor, aby neměly dva soubory stejný název a zachovaly se alespoň určitou strukturu organizace. Pro větší přehlednost byla použita jednoduchá forma popisu souborů. Absenci obecné metodiky lze opět pouze nahradit některými praktickými „doporučeními“ [8]:

a) řazení podle data:

2023-4-28_mereni_Novak.txt
 2023-4-28_poznamky_Novak.pdf
 2023-4-28_mereni_Lebeda.txt
 2023-4-28_poznamky_Lebeda.pdf

b) řazení podle jména výzkumníka:

Lebeda_mereni_2023-4-28.txt
 Lebeda_poznamky_2023-4-28.pdf
 Novak_mereni_2023-4-28.txt
 Novak_poznamky_2023-4-28.pdf

c) řazení podle druhu dat:

Mereni_Lebeda_2023-4-28.txt
 Mereni_Novak-2023-4-28.txt
 Poznamky_Lebeda_2023-4-28.pdf
 Poznamky_Novak_2023-4-28.pdf

d) řazení podle pořadových čísel:

001_Mereni_Lebeda_2023-4-28.txt
 002_Poznamky_Novak_2023-4-28.pdf
 003_Mereni_Novak_2023-4-28.txt
 004_Poznamky_Lebeda_2023-4-28.pdf

Popis výzkumných dat je tvořen daty o „datech“, čili **metadaty**. Každý repozitář používá určité **metadatové schéma**, která vytváří „formulář“ tvořený povinnými (např. autoři, datum uložení, název, typ datasetu atd.) a nepovinnými položkami (klíčová slova, krátký popis aj.). Pokud výzkumník vyplní povinné položky, vytvoří se automaticky záznam, který by měl zůstat v repozitáři i poté, co budou uložená data smazána, aby se uvolnilo místo pro nová.

I když žádný repozitář zatím neuvádí konkrétní časový interval, po kterém budou otevřená data smazána, v české legislativě je již zakotvena povinnost, aby informace otevřené v rámci veřejně financovaných projektů byly **bezplatně přístupné třetím stranám ještě 12 měsíců po ukončení poskytování podpory** [9].

Aby si každý repozitář nemusel vytvářet vlastní schéma, existuje několik základních „šablon“, které používá více veřejných uložišť. Nejběžnější je **DataCite Metadata Schema**

([DataCite Schema](#)). Často se také využívají např. **Data Documentation Initiative** ([Home](#)), **Dublin Core™ Metadata Element Set** ([DCMI: Dublin Core™ Metadata Element Set, Version 1.1: Reference Description](#)), **OLAC Metadata** ([OLAC Metadata](#)) aj.

Výzkumníci by proto měli věnovat maximální pozornost tvorbě metadat k výsledkům své práce. Tudiž každý dataset zveřejněný v repozitáři by měl obsahovat soubor **readme** [10], který sděluje každému zasvěcenému i nezasvěcenému čtenáři tyto důležité informace:

1. Jaké údaje zveřejněná data obsahují.
2. U rozsáhlejších záznamů popisuje uložení souborů v jednotlivých složkách.
3. Informace o poslední aktualizaci datasetu.
4. Poskytuje kontakt na odpovědného autora dat.

Samotný soubor readme je podle požadavků konkrétních projektů ve formátu txt nebo xls, pro snazší aktualizaci v něm obsažených údajů, především data poslední aktualizace. Jazyk zápisu si opět vybírá sám výzkumník a měl by mít zhruba tuto strukturu [10]:

<p>This README file was generated on 2024-03-28 by Last updated: 2024-03-28 ----- GENERAL INFORMATION ----- 1. Title of Dataset: 2. DOI: 3. Contact information Name: Institution: Email: ORCID: 4. Date of data collection: 5. Geographic location: 6. Description of dataset:</p>	<p>Tento soubor ReadMe byl vytvořen 2024-03-28 panem/paní Poslední aktualizace: 2023-11-28 ----- Základní informace ----- 1. Název datasetu: 2. DOI: 3. Kontaktní informace Jméno: Instituce: E-mail: ORCID: 4. Data archivace (zveřejnění) datasetu: 5. Místo archivace (zveřejnění): 6. Popis datasetu:</p>
--	--

Uvedený postup je jen určitým doporučením, protože některé instituce vytvářejí svá vlastní pravidla pro soubor readme, která se většinou mezi sebou odlišují jen nepatrně. Jako další alternativu lze uvést **Obecné doporučení pro metadatový popis výsledků výzkumu (zejména publikací a dat)** [11], kterou v roce 2022 zveřejnila **Národní technická knihovna v Praze** (dále je NTK). Uvedený text navrhuje uvádět v rámci metadat ke každému otevřenému datasetu tyto údaje:

- název datové sady,
- celá jména původců (tj. autorů a přispěvatelů),

- datum (plánovaného) zveřejnění,
- vydavatel,
- popis datové sady,
- informace o dostupnosti dat (např. licence a jiné údaje o dostupnosti).

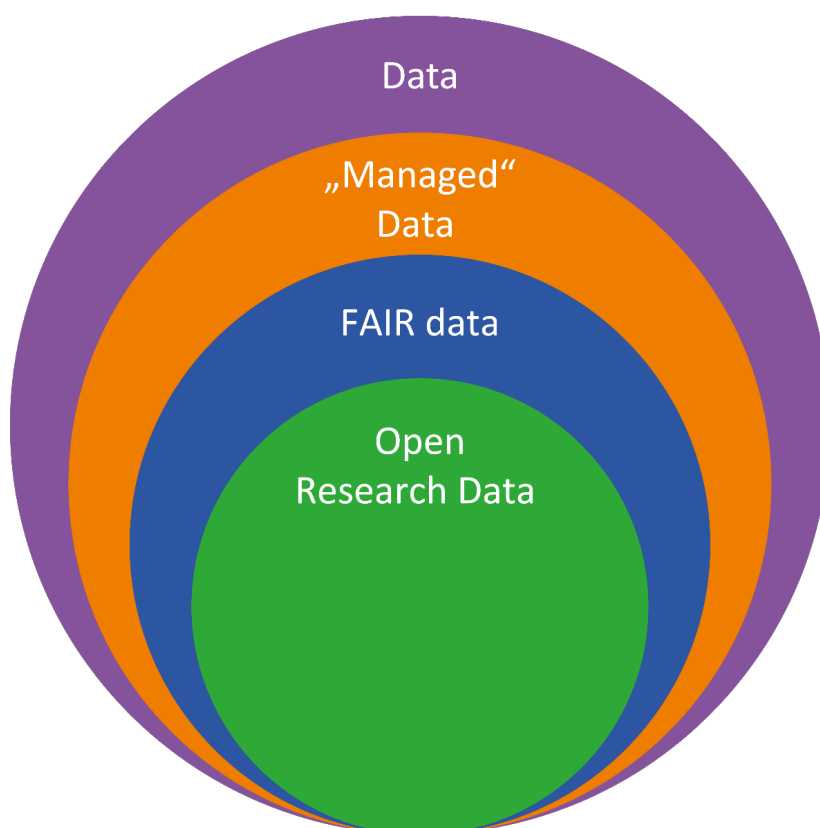
Ve většině případů se jedná pouze o společné disky, na kterých pracuje několik výzkumníků současně, a když jsou se svou prací hotovi, mohou ji zpřístupnit třetím stranám. Proto uvedené online programy nelze pokládat za repozitáře, protože neslouží primárně ke zveřejňování FAIR dat. Také neřeší problematiku názvů otevřených souborů ani jejich metadata – např.:

1. GitHub ([GitHub: Let's build from here · GitHub](#) – přímo napojen na repozitář Zenodo)
2. GitLab ([The DevSecOps Platform | GitLab](#))
3. R Markdown ([R Markdown \(rstudio.com\)](#))

Vždy před zveřejněním souborů v repozitáři se výzkumník musí obrátit na svého Data Stewarda, aby zkontroloval, zda byly dodrženy všechny FAIR principy, aby byly dodrženy základní principy Open Science. I když až po této kontrole je možné data otevřít, přesto by se jejich tvůrci měli snažit dodržovat výše uvedené postupy, aby zbytečně neprodlužovali jejich zveřejňování a publikování vlastních odborných textů.

Analýza dat

Po sestavení Data Management Planu a vyřešení problémů s uspořádáním získaných dat musí každý výzkumník nejdříve zanalyzovat získané informace. Během každého výzkumu je sesbíráno, vypořazováno nebo vygenerováno značné množství dat, které se následně rozdělují do **4 úrovní** [12], přičemž v repozitáři se většinou zveřejňují pouze data čtvrté nejnížší a také rozsahem nejmenší úrovně.



Nově kolorované úrovně dat [12]

První úroveň tvoří veškerá nashromážděná **Data**. Jelikož nejsou uspořádána a ani popsána, nazývají se „**binární odpad**“ nebo **raw data**, která se někdy rozlišují na kvalitativní (numerická) a kvantitativní (nenumernická). Každý výzkumník proto musí data roztřídit na nepotřebná a na použitelná pro další práci nebo odbornou publikaci.

Užitečné soubory se potom přesunují do druhé úrovně, tzv. „**Managed**“ **Data**, kde musí být popsána jednoduchým a pro každého srozumitelným způsobem (viz 2. Zpracování dat). Až takto zpracované informace mohou být umístěny v uložisti (viz 4. Uchovávání dat).

Před zveřejněním je nutné vytvoření tzv. **FAIR dat** podle obecně stanovených principů, což je značně náročný proces, a proto již na internetu existují různé nástroje [13], které obsahují určitá „doporučení“ pro vhodné zpracování dat podle daných principů. Hlavní výhodou popsaných internetových odkazů jsou jejich průběžné aktualizace, které zaručují, že během neustálého rozvoje politiky Open Science přinášejí nejnovější informace.

Přesto hlavní zásady FAIR principů zůstávají zachovány od jejich zformulování v roce 2016 [14]. Protože byly podrobně popsány v předchozím textu (viz 1. Vytvoření dat), lze jen doplnit jejich stručný přehled ve 4 základních bodech:

1. **Dohledatelnost** (Findable) zajišťují **perzistentní identifikátory** pro autory, odbornou publikaci, otevřená data, vědecké instituce atd.
2. Pro **zpřístupnění dat** (Accessible) musí každý výzkumník opatřit svá data potřebnými **metadaty** podle požadavků zvoleného repozitáře a souborem **readme**.
3. **Interoperabilní** (Interoperable) díky využívání běžných **typů souborů**, které by si mohl každý zájemce otevřít ve svém osobním počítači.
4. **Opětovně použitelná** (Reusable) informace musí být opatřeny vhodnými **licencemi**.

Když jsou data zpracována podle FAIR principů, musí jejich autor ještě provést poslední fázi analýzy, aby získal **Open Research Data**. Nejdříve se vědec musí rozhodnout, které informace použije při psaní svého odborného textu jako primární zdroje a které nejsou tak významné pro jeho práci. Samozřejmě se doporučuje zveřejnit co největší množství výzkumných dat, přesto vzhledem k omezenému množství souborů i jejich velikosti v repozitářích (viz 2. Zpracování dat) musí autor vždy uvážit, jaké informace je nezbytné otevřít. Pokud jsou data příliš velká nebo početná pro jeden záznam a je nutné je všechny zveřejnit, nezbyvá nic jiného, než vytvořit více datasetů se stejným názvem a rozlišit je pouze číslovkami (římskými či arabskými) či krátkými dodatky v názvu (např. část 1, část 2 atd.). V případě, že překročil maximální počet souborů, které repozitář povoluje v jednom datasetu, může celou složku zkomprimovat, což většina veřejných uložišť pokládá pouze za jeden soubor. Hlavní nevýhodou je skutečnost, že každý zájemce si poté musí stáhnout celou složku a nejen soubor, o který má zájem. **Výzkumník vždy musí zveřejnit všechna data, která jsou potřeba pro replikaci jim publikovaného odborného textu.**

Když jsou vyřešeny všechny výše popsané problémy, je nutné se zamyslet ještě nad jedním významným hlediskem. Politika Open Science chce sice zpřístupnit výsledky vědeckého výzkumu prakticky každému zájemci, existují některé výjimky, které nemohou být otevřeny, jelikož vždy musí být zajištěny:

1. **Ochrana národní bezpečnosti** (např. důležité vládní kontrakty)
2. **Ochrana osobních údajů (GDPR)** (např. osobní údaje účastníků veřejných výzkumů)
3. **Ochrana duševního vlastnictví** (dosud nepublikované práce jiných autorů)

Uvedená omezení se týkají především veřejných výzkumů v oblasti statistiky, sociologie, lékařských výzkumů atd. Aby mohla být zveřejněna ostatní výzkumná data, existují speciální nástroje (např. **Amnesia** ([Amnesia Anonymization Tool - Data anonymization made easy \(openaire.eu\)](https://openaire.eu))), které zajišťují ve třech jednoduchých krocích tzv. **anonymizaci dat**. Výzkumník nahraje svá data na server. Poté označí informace, které nechce nebo nemůže zveřejnit, a nechá nástroj, aby je zašifroval pro ostatní čtenáře. Zpracované soubory si vědec potom stáhne zpátky do počítače a může z nich vytvořit dataset určený pro repozitář.

Předtím než jsou data otevřena se výzkumník vždy musí obrátit na Data Stewarda, aby provedl kontrolu FAIR principů. I když se autor již během tvorby a popisování dat snažil dodržovat politiku Open Science, přesto je musí před zveřejněním vždy nechat zkontrolovat.

Uchovávání dat

Jakmile je vytvořena metodika pro popis souborů a začnou se vytvářet data, jednotliví výzkumníci si je začnou ukládat na lokální disky, síťová či cloudová úložiště, přenosná zařízení (např. USB disky), mobilní zařízení (mobily, notebooky, tablety aj.) a různé elektronické pracovní nástroje (např. elektronické laboratorní deníky, měřící přístroje atd.). Pro přenos jednotlivých souborů mezi sebou poté používají e-maily, sdílené disky a již mohou příležitostně a dočasně využívat i interní úložiště některých institucí.

Pro upřesnění je nutné uvést, že elektronické laboratorní deníky (někdy se též nazývají terénní a používá se pro ně zkratka **ELN**, která je odvozena z jejich anglického názvu *Electronic Laboratory (Field) Notebook*) prodělaly svůj historický a technologický vývoj a podle těchto hledisek se rozdělují do pěti základních skupin:

1. **papírové** laboratorní deníky
2. **notebooky** (záznamy v souborech Microsoft WORD apod.)
3. **laboratorní e-notebooky** (soubory na cloudu jako OneNote nebo Evernote)
4. **elektronické laboratorní deníky Standard** (první nedokonalé verze)
5. **inventory/LIMS** (moderní a zatím nejvyspělejší verze) [15]

Když je práce na projektu ukončena a poskytovatel již nepřiděluje žádné veřejné finanční prostředky, měli by všichni vědečtí pracovníci shromáždit všechna výzkumná data a po provedení jejich analýzy vybrat informace, které by následně chtěli uchovat pro případné budoucí zájemce o svou práci. Většinou se jedná o veškerá „**Managed**“ data. Po shromáždění vybraných souborů by všechny měly být uloženy v jednom interním úložišti, aby k nim další vědci měli snadný a rychlý přístup, samozřejmě se svolením správců úložiště autorů dat. Není chybou, pokud si jednotlivé instituce zapojené do projektu ponechají svá dat ve svých vlastních úložištích, pouze tím poněkud zkomplikují jejich budoucí znovuvyužití.

Problematikou uchovávání dat se zatím nejpodrobněji zabýval Angus Whyte ve svém odborném textu *Five steps to decide what data to keep* publikovaném v roce 2014 [16]. Kromě podrobné analýzy dané problematiky sestavil **5 kroků**, které by měly každému výzkumníkovi pomoci určit, jaká data je vhodné zachovat v interním úložišti:

1. K čemu by data mohla v budoucnosti sloužit? Jak je lze opětovně využít?

Angus Whyte navrhuje, aby každý výzkumník u získaných informací uvážil:

- **Ověření:** Data, která je třeba uchovat pro zopakování výsledků výzkumu.
- **Další analýza:** Některé informace mohou být později podrobeny dalším analýzám podle jiných nebo nových metod.
- **Budování akademické reputace:** Je praktické otevřít soubory, jež by mohly posloužit jako podklady pro další odborné texty a zvýšit citovanost autorů.
- **Vývoj komunitních zdrojů:** Informace sloužící určité skupině uživatelů (např. výzkumníků příbuzných oborů).
- **Další publikování:** Zveřejněná data může vědecká komunita použít k různým odborným debatám.
- **Výuka:** Některé soubory lze použít jako podklady pro výuku.
- **Soukromé využití:** Informace uložené v repozitářích by měly být využitelné i k jiným než vědeckým účelům.

Výzkumníci by primárně měli plnit požadavky zadavatele projektu, což jsou většinou ověření a další analýzy získaných dat. Ostatních pět hledisek lze považovat pouze za sekundární možnosti, které mohou nastat až po delším časovém období.

2. Jaká data se musí uchovávat na základě smluvních či zákonných podmínek?

Uchovávání dat také zásadně ovlivňují instituce poskytující finance na projekt nebo zvyklosti zapojených vědeckých organizací. Kromě typů souborů mohou také rozhodovat, zda budou informace volně dostupné nebo s omezeným přístupem.

3. Jaká data by mohla mít dlouhodobou hodnotu?

Angus Whyte ve svém článku opět uvádí celý seznam kritérií, které mají výzkumníkům pomoci při určení dlouhodobé hodnoty získaných dat:

a) Jsou data dostatečně kvalitní?

- **Popis:** Důležitý je nejvýstižnější popis dat v názvu a metadatech.
- **Kvalita:** Je nutné posoudit úplnost a informační obsah souborů.

b) Je pravděpodobné, že po datech bude poptávka?

- **Známí uživatelé:** Autoři by měli uvážit, kdo by mohl mít zájem o shromážděná data.
- **Doporučení:** Poskytovatelé finančních prostředků nebo vědecká komunita doporučí výzkumníkovi uložit některé informace.
- **Integrační potenciál:** Pro integraci dat do konkrétní vědecké komunity je nutné dodržovat odbornou terminologii.
- **Reputace:** Uložené informace by měly posílit renomé vědců, kteří je vytvořili nebo sbírali.
- **Dosah:** Soubory mohou mít vliv na širší veřejnost, pokud se týkají významného objevu.

c) Lze data snadno replikovat?

- Každý výzkumník musí zvážit možnost, zda bude replikování dat pro další vědce obtížné nebo drahé. Také může nastat možnost, že výzkum již nebude možné zopakovat.

d) Existují překážky pro další využití?

- **Schválení/Pročištění:** Každý výzkumník musí u svých dat posoudit, zda neporušují etické či licenční podmínky, aby je mohl zpřístupnit veřejnosti.
- **Otevřený formát:** Všechny soubory musí být ve formátech, které lze otevřít v běžně dostupných programech, aniž by si možný budoucí uživatel musel pořizovat placenou licenci nebo programy.
- **Nezávislost:** Samozřejmě je možné uchovávat soubory, které vyžadují k otevření běžnému uživateli nedostupný program, nicméně pokud jej mají běžně k dispozici vědci v daném vědním oboru, lze i taková data zveřejnit.

e) Je to jediná kopie?

- **Jedinečnost:** Výzkumník vždy musí dlouhodobě uchovávat soubory, které jsou unikátní a kompletní.
- **Ohrožená data:** Před uložením dat do interního úložiště si vědec musí vždy ověřit dobu, po kterou v něm budou uložena, a současně i důvěryhodnost jeho správce, aby nemohli zneužít informace určené k uchování.

Na základě předchozího popisu analýzy získaných informací lze jen doplnit, že do interního úložiště by měla být uložena především všechna **Open Research data** a pokud je místo i všechna **FAIR data**. Důkladné analýze a třídění by měli vědci podrobit až „Managed“ data (viz 3. Analýza dat), přičemž by neměli zapomenout, že nejdůležitější jsou data, která jsou

nezbytná pro replikaci jimi publikovaných textů (viz 3. Analýza dat). Angus Whyte a jeho kolega Andrew Wilson se problematikou archivace zabývali v jiném článku *How to Appraise and Select Research Data for Curation* publikovaném v roce 2010 a odhadli, že by se ze všech dat („Managed“ data, viz 3. Analýza dat) mělo archivovat **4-8 %**. [17]

4. Jaké budou náklady na uchování dat?

Autoři musí uvážit již vynaložené finanční prostředky především na vybavení a různé poplatky za licence aj. Rovněž je nutné promyslet i budoucí vydání na přípravu výzkumných dat k archivaci a za uložení a správu souborů v interním uložišti.

5. Předchozí čtyři kroky by měly výzkumníkům poskytnout pravidla pro dokončení analýzy dat a určení, které informace je třeba dále uchovávat.

Výzkumníci by se nakonec měli zaměřit na dlouhodobou hodnotu získaných dat a veškeré vynaložené náklady, finanční i časové.

Uvedené odborné texty se nezabývají jednou zásadní otázkou, **jak dlouho by data měla zůstat v interním uložišti**, když není možné je mít uložené natrvalo. Samozřejmě by se všichni autoři měli snažit uchovávat výsledky svého výzkumu po co možná nejdelší dobu. Angus Whyte doporučuje výzkumná data archivovat alespoň 3 roky. [16] V ČR musí být podle zákona zveřejněná v repozitáři ještě 12 měsíců po ukončení financování projektu [9]. Některé instituce v ČR ve svých politikách správy výzkumných dat již stanovily termíny pro uchování souborů, které byly vytvořeny v rámci veřejných výzkumných projektů (např. **Univerzita Karlova** ustanovila **10 let** [6]). U souborů uložených v interním uložišti si pravděpodobně dobu uchování dat určí jeho provozovatel po dohodě s jejich autory.

Zpřístupnění dat

Shromážděná a zanalyzovaná výzkumná data se zpřístupňují prostřednictvím datových časopisů nebo repozitářů. **Datové časopisy** publikují články, které popisují určitou datovou sadu uloženou v repozitáři, a jsou oborově zaměřené.

I když repozitáře je možné definovat několika různými způsoby, asi nejdůležitější poskytují ve své prezentaci Michal Růžička [18], podle které slouží nejen k ukládání a zveřejňování výzkumných dat, ale také pro aktivní práci s tzv. „živými“ daty. Pokud repozitář neposkytuje automaticky všechny uvedené služby, nelze jej pokládat za důvěryhodný.

Univerzita Karlova rozlišuje ve svém institucionálním repozitáři údaje **povinné** a **povinně volitelné**. [19] Ostatní veřejná uložiska označují ve svých záznamech většinou hvězdičkou údaje, které musí autoři vyplnit, jinak nebude jejich dataset otevřen. I když existuje několik „šablon“ metadatových schémat (viz 2. Zpracování dat), přesto by každý popis dat v záznamu **měl vždy obsahovat**:

1. **DOI** – pokud DOI zatím není přiděleno, většina repozitářů umožňuje políčko dočasně nevyplňovat.
2. **Typ zdroje (Resource type)** – většinou se jedná o seznam možností, ze kterého se nejčastěji vybírá termín dataset, i když lze zvolit i jinou možnost podle aktuálních požadavků.
3. **Název (Title)** – co možná nejkratší ale současně nejdůležitější, nestačí napsat několik odborných zkratk, které mohou být nepochopitelné pro některé výzkumníky.
4. **Datum zveřejnění (Publication Date)** – repozitář většinou automaticky vygeneruje aktuální datum, které je možné ponechat nebo ručně přepsat na jiné.
5. **Autoři (Creators)** – vždy se uvádí všichni autoři textu: křestní jména, příjmení, názvy jejich institucí a lze doplnit i jejich ORCID ID, pokud jej zařídili.

Anglické názvy jsou používány v repozitáři Zenodo, nicméně u jiných cizojazyčných uložisk se mohou lišit, i když princip tvorby záznamu zůstává stejný. Pokud repozitář povinně nevyžaduje vyplnění výše uvedených údajů před zveřejněním souborů, měl by jej výzkumník opustit, protože ho nelze považovat za důvěryhodný. [20]

Při výběru vhodného uložiska pro otevření výzkumných dat by se měl každý autor řídit oborovými zvyklostmi nebo postupy běžně používanými ve své instituci. Nicméně když nejsou vhodné pro jeho vědecké záměry, může využít různých online nástrojů, které mu pomohou vyhledat vhodné veřejné uložisko pro jeho soubory, např.:

1. CESSDA ([CESSDA - Consortium of European Social Science Data Archives](#))
2. FAIRsharing ([FAIRsharing | Home](#))
3. Re3data ([Home | re3data.org](#)) aj.

Všechny existující repozitáře se dnes rozdělují do tří základních skupin:

1. **Obecné** – nazývané též „sirotčí“, je možné v nich ukládat data jakéhokoliv oborového zaměření, přičemž existují na mezinárodní (např. Zenodo) nebo na národní bázi (např. Národní repozitář).
2. **Oborové** – zaměřené pouze na jeden vědní obor, kterému jsou technicky přizpůsobeny, čímž usnadňují práci výzkumníků; většinou nadnárodní.

3. Institucionální – vytvářejí je a spravují jednotlivé vědecké organizace, proto se většinou vytváří na národní bázi, přičemž do nich primárně ukládají odborné texty svých studentů nebo členů, přesto je možné v nich zveřejňovat i FAIR data). [21]

Výzkumníci, kteří společně pracují na jednom projektu, se sami musí rozhodnout v jakém repozitáři budou zveřejňovat získaná data. Pokud všichni pracují v jedné instituci, mohou využít služeb institucionálního veřejného uložště. Když je do projektu zapojeno více vědeckých organizací, je možné využívat oborové repozitáře. Pokud se autoři nemohou dohodnout nebo nemají ještě dostatek zkušeností se zveřejňováním výzkumných dat, je vhodné využít služeb obecných veřejných uložště, které díky své univerzálnosti nejsou náročné na pochopení a tvorbu záznamů. Zatím neexistuje právní předpis, který by vědcům nařizoval používat pouze jeden repozitář pro celý veřejný projekt. Přesto je značně nepraktické používat pro otevírání výzkumných dat více než jedno uložště, protože informace se potom stávají méně přehlednými nejen pro budoucí zájemce, ale také pro samotné autory. Přesto nelze považovat využívání více veřejných uložště v rámci jednoho výzkumu za zásadní chybu.

Pokud výzkumníci budou zveřejňovat více odborných textů v rámci jednoho projektu, mohou si vytvořit v repozitáři vlastní **komunitu**, což usnadní jejich komunikaci i zveřejňování FAIR dat. Komunitu vždy založí jeden administrativní pracovník, který do ní následně jako její správce zapisuje ostatní vědce. Když se autoři rozhodnou zveřejnit data ke svým publikacím, je důležité při tvorbě záznamu zvolit i příslušnou komunitu. Jestli výzkumník popsanou položku během tvorby záznamu přehledne nebo na ni zapomene, nemusí zbytečně vytvářet nový záznam. Většina veřejných uložště nabízí možnost přiřadit dataset k jedné nebo i více vědeckým komunitám i po jeho otevření.

V každém repozitáři je možné zveřejnit jen jeden soubor (např. metodiku, preprint nebo postprint článku, samostatnou tabulku atd.), přičemž by k němu měl být doplněn i **readme**. Pokud se vědci rozhodnou otevřít dva a více souborů, musí z nich vytvořit tzv. dataset, což je složka všech souborů určených pro otevření (viz 3. Analýza dat). Může obsahovat pouze jednotlivé soubory, což je většinou případ odborných publikací menšího rozsahu (např. články v časopisech, ve sbornících atd.), nebo přirozeně propracovaný systém složek. **Mezi zveřejněnými daty musí být vždy i soubor readme** (viz 1. Vytvoření dat). Když dataset obsahuje systém složek, zveřejní se v repozitáři pouze jednotlivé soubory. Pokud chce autor zachovat své původní přirozené rozdělení informací, musí složku pro zveřejnění zkomprimovat na soubor rar, .zip aj. Struktura dat sice bude zachována, nicméně případný budoucí výzkumník si musí stáhnout celou složku, která může mít velikost i několik GB. Když jsou otevřeny jednotlivé soubory, mohou si zájemci stáhnout jen konkrétní informace, o které mají zájem.

Jak uvedli ve své metodice zaměstnanci Vysoké školy chemicko-technologické v Praze [22], po zveřejnění datasetu již není na Zenodo možné provádět jeho úpravy. Ve skutečnosti se otevřené soubory již nemohou mazat nebo doplňovat, nicméně aktualizace dat uložených v repozitáři patří mezi základní pravidla FAIR principů. Proto je na Zenodo možné vytvořit úplně novou verzi záznamu s nezměněnými původními údaji. Jedinou nevýhodou je, že se musí nahrát kompletně nový dataset, tudíž všechny staré soubory, jež nechtějí výzkumníci aktualizovat, musí být opět znovu přetaženy do veřejného uložště. Právě i pro tyto případy je nezbytné uchovávat všechny důležité výzkumné soubory v jiném, nejlépe interním, uložšti (viz 4. Uchovávaní dat). Každé nové verzi je automaticky přidělen i nový perzistentní identifikátor, a proto si každý autor musí dávat pozor, zda použil v citaci vhodný řetězec.

Jelikož o popisech jednotlivých souborů rozhoduje především samotný výzkumník, lze obecně stanovit jen několik základních „doporučení“ pro usnadnění orientace v datech, až budou zveřejněna (uvedené postupy vytvořil autor předložené metodiky na základě svých vlastních zkušeností při otevírání dat v Národním repozitáři a Zenodo):

- Pokud se autor rozhodne používat zkratky, měl by uvažovat, jestli budou srozumitelné všem odborníkům na danou problematiku. Pokud si nebude jistý, je vhodné názvy rozepsat na celá slova.
- Žádné dva soubory by neměly mít stejný název, protože do repozitáře se ukládají do jediné složky a pokud mají identické označení, systém uloží pouze jeden a druhý ignoruje, tudíž nejsou zveřejněna všechna důležitá data.
- Při popisování souborů se autoři musí zvolit jazyk popisu. Data a údaje v záznamu mohou popsát v souladu s cizojazyčným textem nebo je mohou zveřejnit v jiném vhodném jazyce.
- Také je důležité zkontrolovat počet souborů a jejich velikost. Např. Zenodo dovoluje uložit k jednomu záznamu maximálně 100 souborů najednou o celkové velikosti 50 GB.
- Uložená data v repozitáři by se neměla zbytečně rozdělovat. Pokud počet souborů není větší než jejich maximální povolené množství, autor by je měl zveřejnit v jediném záznamu.
- Když vědec potřebuje zveřejnit více než povoluje repozitář, měl by se snažit data přirozeně rozdělit na co možná největší skupiny, aby nebylo vygenerováno zbytečně velké množství perzistentních identifikátorů, které bude následně uvádět v citacích ve svém textu.

I když podle politiky Open Science by se měla výzkumná data sdílet v **každé fázi projektu**, každý autor by měl počkat, dokud neshromáždí dostatečné množství informací pro vytvoření odborného textu alespoň menšího rozsahu (např. článek, vědecká zpráva aj.). Někdy může i samotný zadavatel projektu požadovat, aby byla data zveřejňována v určitých časových intervalech, většinou společně s aktualizacemi DMP (viz 1. Vytvoření dat).

V případě technických problémů s repozitářem by se měl výzkumník obrátit na správce své komunity nebo přímo na technickou podporu zvoleného veřejného uložiště. Správci repozitáře poté mohou žadateli popsát přesný postup pro odstranění problému nebo ho sami vyřešit v případě, že výzkumník nemá přístup k potřebným službám.

Opětovné využití dat

Výzkumná data se zveřejňují, aby se vědecká práce stala transparentnější a důvěryhodnější ve vědecké komunitě i před širokou veřejností. I když vše bylo detailně probráno v předchozím textu, je praktické nyní provést stručné shrnutí, proč by měly být informace shromážděné během výzkumu sdíleny:

1. Proč data sdílet?

Pokud otevření výzkumných souborů nepožaduje přímo zadavatel projektu, sdílení údajů zvyšuje **prestíž** všech zapojených vědců a jejich institucí. V případě rozsáhlejších vědeckých projektů tak lze **urychlit** výzkumný proces. Nejdůležitější je **zvýšení citovanosti** odborných publikací. Jak dokazují některé zahraniční studie, u textů s daty otevřenými podle FAIR principů byl o 25 % [23] až 70 % [24] větší počet citací než u článků bez zveřejněných souborů.

2. Jaká data sdílet?

Uvedená problematika je podrobně zanalyzovaná (viz 3. Analýza dat), lze jen zopakovat, že je nutné zveřejnit všechny soubory, které **jsou potřeba k replikaci publikované studie**.

3. Kdy data sdílet?

Jak opět už bylo uvedeno v předchozí části (viz 5. Zpřístupnění dat), data by měla být sdílena **kdykoliv během výzkumného procesu**.

Výzkumník může mít pro zveřejnění výsledků tři základní důvody:

- a) Nemá prostředky na další výzkum a poskytuje své výsledky jiným kolegům, aby na ně mohli navázat a práci dokončit.
- b) Spolupracuje s řadou i zahraničních institucí a své soubory poskytuje podle dříve uzavřené dohody svým spolupracovníkům, aby mohli ihned pokračovat ve výzkumu a nemuseli čekat, až budou data zveřejněna v odborném tisku.
- c) Ve většině případů vytvořil vědec na základě získaných údajů vlastní odborný text a podle podmínek Open Science musí svá data zpřístupnit všem jeho čtenářům.

Ve všech případech je nutné vybavit data **otevřenou licencí**, aby je mohli dále používat další výzkumníci nebo široká veřejnost. I když nejběžněji používanými jsou licence **Creative Commons**, existují i další vhodné varianty, např.:

- Apache License
- Berkeley Software Distribution (BSD)
- Community Data License Agreement (CDLA)
- Eclipse Public License (EPL)
- GNU General Public License (GNU GPL)
- MIT License
- Mozilla Public License
- Open Data Commons (ODC)
- Public Domain aj.

Licence Creative Commons [25] jsou podrobněji definovány čtyřmi základními prvky (*BY*, *NC*, *ND* a *SA*), které tvoří pak různé další typy licencí. Podrobný rozbor celého systému je nad možností této metodiky [26] – [27], a proto lze jen výzkumníkům doporučit, aby používali dvě nejliberálnější verze *CC BY* a *CC BY-SA*, které umožňují nejširší možné využití odborných

informací zveřejněných v repozitářích. Nejvhodnější je proto používat verzi 4.0, která je mezinárodní a zveřejněné informace i publikace zpřístupňuje prakticky ve všech státech. [25]

Po opatření odborné publikace licencí se musí její autoři rozhodnout, jaký způsobem ji otevřou svým kolegům i veřejnosti, na což opět existuje několik postupně vyvinutých způsobů publikování, které se v češtině označují jako **cesta** nebo **forma**:

1. **Bronzová cesta/forma**

Odborné publikace jsou zveřejněny na stránkách vydavatele, nicméně jsou většinou nepoužitelné, protože jim chybí licence.

2. **Černá cesta/forma**

Označení pro texty zpřístupněné nelegálně.

3. **Hybridní cesta/forma**

Kombinace uzavřených i otevřených textů, které jsou dostupné prostřednictvím předplatného, přičemž jejichž autoři platí APC (Article Processing Charge = poplatek za zpracování článku či publikace).

4. **Platinová či diamantová cesta/forma**

Publikování v otevřených časopisech bez APC. Finanční zdroje si redakce zajišťuje prostřednictvím externího zdroje (např. inzercemi nebo reklamou), tudíž poplatky nehradí ani vědci ani čtenáři.

5. **Zelená cesta/forma**

Výsledky jsou archivovány v repozitářích v tzv. **verzích plných textů** [19]:

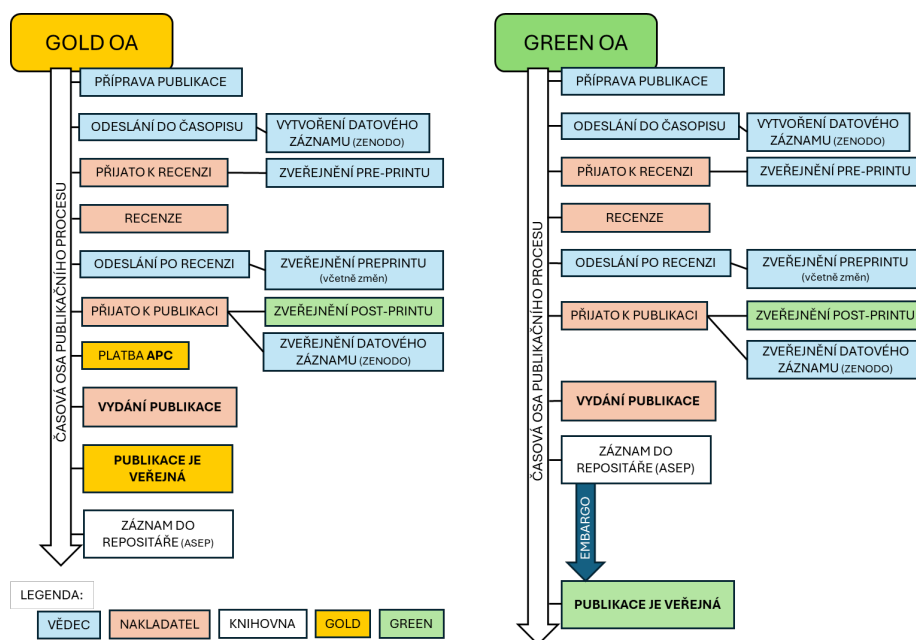
- **Rukopis** (draft): pracovní verze výsledku, která se většinou neotevívá.
- **Preprint** (submitted version): autorem dokončená verze textu, která byla zaslána k recenznímu řízení.
- **Postprint** (accepted version): verze výsledku po recenzním řízení, do které již autor zapracoval požadované změny, ale ještě není připravena pro tisk.
- **Vydavatelská verze** (published version): verze odborného textu publikovaná vydavatelem s konečným formátováním a grafickými úpravami pro časopis.
- **Aktualizované verze** (updated version): verze výsledku aktualizované po jeho vydání, přičemž jejich počet je zatím neomezený.

Vše si zajišťuje sám autor, a proto ani on ani čtenáři nehradí žádné poplatky. Otevření článků může někdy předcházet časové embargo 12 či 24 měsíců.

6. **Zlatá cesta/forma**

Postup pro publikování v otevřených časopisech, přičemž všechny poplatky APC musí uhradit výzkumník, čtenář má vše zdarma. Text je přístupný ihned na stránkách redakce, jelikož je okamžitě opatřen licencí, většinou *Creative Commons*. Existují i Open Access časopisy, které nevyžadují žádné poplatky. Jejich přehled poskytuje online databáze **DOAJ** ([Directory of Open Access Journals – DOAJ](#)).

V roce 2025 se v politice Open Science nejčastěji uplatňují **zelená** a **zlatá cesta/forma**. V případě **zelené** verze používá většina výzkumníků v ČR pro zveřejňování obecné (např. Zenodo) nebo institucionální repozitáře (např. DSpace), i když v zahraničí se již pomalu začínají prosazovat oborová veřejná uložiska, která jsou pro daný účel teoreticky nejvhodnější.



Vývojové diagramy zlaté a zelené cesty/formy [28]

Podle normy ČSN ISO 690:2022 (01 0197) existují pravidla odkazování pro datasey [29]:
 Tvůrce. *Název. Vydání, verze.* Formát a typ zdroje. Systémové požadavky. Místo: vydavatel, datum publikace. *Licence.*

Tvůrce. *Název. Vydání, verze.* Formát a typ zdroje. Systémové požadavky. In: *Název* hostitelského archivu. Místo: vydavatel, datum publikace. *Licence.* Dostupnost. [*Velikost souboru*]. [Datum citování].

- Příklad: KLIMES, Jiri a HOFIERKA, Jaroslav. *Dataset for "Binding energies of molecular solids from fragment and periodic approaches"*. Version 1. Online, dataset. In: Zenodo. 24. srpen 2021. Licence: CC BY 4.0. Dostupné z: <https://doi.org/10.5281/zenodo.5248078>. [632,1 MB]. [cit. 2023-03-10].
- Příklad: PLH, Jiří. *Nominální a reálné příjmy, 2018, NUTS3*. Online, dataset. In: Otevřená data TUL. Technická univerzita v Liberci, 11. června 2021, aktualizace 22. července 2022. Licence: CC BY-NC-SA 4.0. Dostupné z: <https://arcp.is/10Ofbb1>. [cit. 2023-03-10].

Fialově zvýrazněné části textu jsou nepovinné. V odborné literatuře se citace píše celé černě, kromě perzistentních identifikátorů. Velikost datasetu je povinné uvádět pouze u složek velkého rozsahu. Jelikož norma sama neuvádí žádná konkrétní čísla, lze stanovit jen přibližnou hodnotu na zhruba 500 MB. I když existují online citační manažery, nejsou přímo nastaveny funkce pro tvorbu citačních záznamů výzkumných dat. Autor si nicméně může např. v nástroji Citace PRO Plus ([Citace PRO](#)) zvolit možnost **Vytvořit** → **Příspěvek, obrázek na webu**. Zde lze vytvořit potřebnou citaci, ale bez nepovinných částí (konkrétně *verze*, *licence* a *velikost souboru*), které výzkumník musí dopsat ručně.

Jelikož Data Stewardi se zabývají **Open Daty**, zatímco problematiku otevřených publikací řeší **Open Access**, musí se výzkumníci s případnými dotazy obrátit na jiné administrativní pracovníky v rámci svého výzkumného projektu nebo přímo na zadavatele. Experti na otevřená data v této fázi jejich životního cyklu pomáhají jen minimálně, především s citováním datasetů v odborných textech a výběrem licence.

Závěr

Předložená metodika poskytuje pouze stručný souhrn základních postupů, které by měli vědci dodržovat během životního cyklu vlastních výzkumných dat. Jak už bylo několikrát uvedeno v předchozím textu, není možné stanovit přesná obecná pravidla, která by platila ve všech oborech a odpovídala konkrétním požadavkům jednotlivých vědeckých projektů. Prakticky v každé fázi je nezbytný individuální přístup každého výzkumníka, který se v případě nejistoty nebo nejasností může obrátit na Data Stewarda nebo Open Science metodika, aby měl jistotu, že splnil všechny podmínky otevřené vědy. Metodika proto neposkytuje univerzální návod pro aplikaci politiky Open Science na vědecký výzkum a ani úplně nenahrazuje práci Data Stewarda, který se vždy bude muset zapojit do vědeckého projektu. Pouze seznamuje čtenáře s několika obecnými principy, jež se opakují při každé práci s otevřenými daty.

I po přečtení předchozích stran by se žádný výzkumník neměl obávat pokládat otázky k dané problematice. **Žádný dotaz není špatný, hloupý nebo zbytečný!** Jelikož celá politika je pro českou vědeckou obec poměrně nová, každý výzkumník během své práce narazí na mnoho otázek a problémů, k jejichž řešení slouží právě pozice Data Stewardů a Open Science metodiků a různé odborné publikace, včetně předloženého textu.

Jelikož politika Open Science se v České republice stále pomalu vyvíjí, mohou dva a více poradců, pokud se na ně obrátí stejný výzkumník, narazit na neshody ve svých doporučených postupech. Vědec se poté sám musí rozhodnout, kterého odborníka si zvolí pro svou práci s otevřenými daty. Lze jen doufat, že v nedaleké budoucnosti budou stanoveny obecnější pravidla na základě praktických zkušeností, díky kterým už nebude docházet k neshodám mezi Data Stewardsy nebo Open Science metodiky.

Seznam použité literatury

- [1] MASARYKOVA UNIVERZITA. *Open Access – Současná situace v oblasti publikování*. Online. 2013. Dostupné z: https://is.muni.cz/el/med/podzim2015/VLLL021/um/08/open_access/ch01.html#prettyPhoto. [cit. 2024-05-14].
- [2] VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA. 5.3. *Otevřená výzkumná data – studijní opora: 5.3.4. Plán managementu dat*. Online. Studijní opora: 5.3.4. Plán managementu dat. Dostupné z: <https://lms.vsb.cz/mod/book/view.php?id=1151625&chapterid=10488>. [cit. 2024-04-23].
- [3] UNIVERZITA KARLOVA – CENTRUM PRO PODPORU OPEN SCIENCE. *Technologická agentura ČR (TA ČR)*. Online. 2023. Dostupné z: <https://openscience.cuni.cz/OSCI-208.html>. [cit. 2024-04-11].
- [4] NÁRODNÍ TECHNICKÁ KNIHOVNA. *Horizon Europe Data Management Plan template (vzor)*. Verze 2.0. 2022. Dostupné také z: <https://repozitar.techlib.cz/server/api/core/bitstreams/2742fd81-9253-4350-9055-164ce710ec89/content>.
- [5] EPIC – *Persistent Identifiers for eResearch*. Online. 2024. Dostupné z: http://www.pidconsortium.net/?page_id=1060. [cit. 2024-04-11].
- [6] *Politika správy výzkumných dat na UK*. Online. In: Univerzita Karlova. 2024. Dostupné z: <https://cuni.cz/UK-13492.html>. [cit. 2024-04-18].
- [7] VYSOKÁ ŠKOLA CHEMICKO-TECHNOLOGICKÁ. *Konverze a skladování energie (ECO&Stor)*, reg. č. CZ.02.01.01/00/22_008/0004617. 2024.
- [8] LONGVA, Leif. *How to structure and document research data - Organising*. Online, video. 2023. Dostupné z: <https://www.youtube.com/watch?v=w9qPR0gDsBA>. [cit. 2025-03-04].
- [9] AION CS - INFO@AION.CZ. *Zákon č. 130/2002 Sb. Zákon o podpoře výzkumu a vývoje z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje)*. Online. *Zákony pro lidi*. Dostupné z: <https://www.zakonyprolidi.cz/cs/2002-130?citace=1>. [cit. 2024-04-23].
- [10] LONGVA, Leif. *How to structure and document research data - the ReadMe file*. Online, video. 2023. Dostupné z: <https://www.youtube.com/watch?v=jGdJIF3rKws>. [cit. 2025-03-04].
- [11] NÁRODNÍ TECHNICKÁ KNIHOVNA. *Obecné doporučení pro metadatový popis výsledků výzkumu (zejména publikací a dat)*. Online. 2022. Dostupné z: <https://repozitar.techlib.cz/server/api/core/bitstreams/8cad20c6-14de-4429-b8ed-74f514c052da/content>. [cit. 2024-04-12].
- [12] ŠKOLICÍ CENTRUM EOSC-CZ. *Úvod do FAIR dat*. Online, video. 2024. Dostupné z: Youtube, https://www.youtube.com/watch?v=_DC0vQS_lj0. [cit. 2025-03-04].

- [13] GO FAIR. *FAIR Principals*. Online. GO FAIR. 2017. Dostupné z: <https://www.go-fair.org/fair-principles/>. [cit. 2024-04-18].
- [14] WILKINSON, Mark D.; DUMONTIER, Michel; AALBERSBERG, IJsbrand Jan; APPLETON, Gabrielle; AXTON, Myles et al. The FAIR Guiding Principles for scientific data management and stewardship. Online. *Scientific Data*. 2016, roč. 3, č. 1. ISSN 2052-4463. Dostupné z: <https://doi.org/10.1038/sdata.2016.18>. [cit. 2024-04-23].
- [15] CEBECAUER, Marek. *Elektronické laboratorní (terénní) deníky - (ELN)*. Online. 2024. Dostupné z: https://www.youtube.com/watch?v=cvtxb_eZqOw. [cit. 2025-03-04].
- [16] WHYTE, Angus. *Five steps to decide what data to keep*. Online. In: DCC Because good research needs good data. 2014. Dostupné z: <https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep#7>. [cit. 2024-04-18].
- [17] WILSON, Andrew a WHYTE, Angus. *How to Appraise and Select Research Data for Curation*. Online. In: DCC Because good research needs good data. 2010. Dostupné z: <https://www.dcc.ac.uk/guidance/how-guides/appraise-select-data>. [cit. 2024-04-18].
- [18] RŮŽIČKA, Michal. *Očekávané služby Národní datové infrastruktury (NDI)*. Online, přednáška. 2022. Dostupné z: [https://www.e-infra.cz/file/10f3a8516a342db81f61ef32b245fafa/670/2_Oc%CC%8Ceka%CC%81vane%CC%81%20sluz%CC%8Cby%20Na%CC%81rodni%CC%81%20datove%CC%81%20infrastruktury%20\(NDI\)_Ru%CC%8Az%CC%8Cic%CC%8Cka_Praha_1.6.pdf](https://www.e-infra.cz/file/10f3a8516a342db81f61ef32b245fafa/670/2_Oc%CC%8Ceka%CC%81vane%CC%81%20sluz%CC%8Cby%20Na%CC%81rodni%CC%81%20datove%CC%81%20infrastruktury%20(NDI)_Ru%CC%8Az%CC%8Cic%CC%8Cka_Praha_1.6.pdf). [cit. 2025-03-04].
- [19] UNIVERZITA KARLOVA – ÚSTŘEDNÍ KNIHOVNA UK. *Metodika open access a autoarchivace na Univerzitě Karlově*. 2022.
- [20] *Digitální repozitář*. Online. Dostupné z: https://wikisofia.cz/wiki/Digit%C3%A1ln%C3%AD_repozit%C3%A1%C5%99. [cit. 2025-03-04].
- [21] *Institucionální repozitář*. Online. Dostupné z: https://wikisofia.cz/wiki/Institucion%C3%A1ln%C3%AD_repozit%C3%A1%C5%99#cite_note-kral-3. [cit. 2025-03-04].
- [22] SCHÄTZ, Martin; VALIŠ, Jan; KOUNTENTAKI, Georgia a SKLENÁŘ, Adam. Datasets archiving in the Zenodo repository and in the UCT Prague Zenodo Community. Online. *Zenodo*. 2024, roč. 2024, s. 4. Dostupné z: <https://doi.org/10.5281/zenodo.10890096>. [cit. 2024-04-18].
- [23] COLAVIZZA, Giovanni; HRYNASZKIEWICZ, Iain; STADEN, Isla; WHITAKER, Kirstie; MCGILLIVRAY, Barbara et al. The citation advantage of linking publications to research data. Online. *PLOS ONE*. 2020, roč. 15, č. 4. ISSN 1932-6203. Dostupné z: <https://doi.org/10.1371/journal.pone.0230416>. [cit. 2024-04-18].

- [24] PIWOWAR, Heather A.; DAY, Roger S.; FRIDSMA, Douglas B. a IOANNIDIS, John. Sharing Detailed Research Data Is Associated with Increased Citation Rate. Online. *PLoS ONE*. 2007, roč. 2, č. 3. ISSN 1932-6203. Dostupné z: <https://doi.org/10.1371/journal.pone.0000308>. [cit. 2024-04-18].
- [25] UNIVERZITA KARLOVA. *Licence Creative Commons - základní informace*. Online. 2016. Dostupné z: <https://publications.cuni.cz/page/licenses?locale-attribute=cs>. [cit. 2024-04-20].
- [26] UNIVERZITA KARLOVA – CENTRUM PRO PODPORU OPEN SCIENCE. *Licence*. Online. 2023. Dostupné z: <https://openscience.cuni.cz/OSCI-68.html>. [cit. 2024-04-23].
- [27] CREATIVE COMMONS ČESKÁ REPUBLIKA. *Jak používat CC*. Online. 2024. Dostupné z: <https://www.creativecommons.cz/pro-autory/publikace-del/>. [cit. 2024-04-23].
- [28] VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA. *OS pro ZoR I*. 2024.
- [29] FARKAŠOVÁ, Blanka; GARAMSZEGI, Tereza; JANSOVÁ, Linda; KONEČNÝ, Lukáš; KRČÁL, Martin et al. *Výklad normy ČSN ISO 690:2022 (01 0197) účinné od 1.12.2022*. V Brně: Citace.com, 2023. Dostupné také z: <https://www.citace.com/Vyklad-CSN-ISO-690-2022.pdf>.