

# FAIR příručka pro data steward komunitu v ČR



Hlavní editor:

- Karolina Podloucká (NTK) <https://orcid.org/0000-0002-5623-2949>

Editori:

- Tereza Kolmačková (KNAV) <https://orcid.org/0000-0003-0829-0770>
- Věra Franková (UK) <https://orcid.org/0000-0002-1927-9596>
- Anna Tůmová (UK) <https://orcid.org/0000-0002-6716-0987>
- Matěj Uchýtil (UK)

Kolektiv autorů:

- Karolina Podloucká (NTK) <https://orcid.org/0000-0002-5623-2949>
- Adéla Jílková (NTK) <https://orcid.org/0000-0002-3158-9626>
- Jan Skůpa (VUT) <https://orcid.org/0000-0001-8033-9634>
- Jindřich Fejfar (KNAV AV ČR) <https://orcid.org/0000-0002-2048-3682>
- Eliška Pospíchalová (KNAV AV ČR) <https://orcid.org/0009-0008-6092-7146>
- Eliška Blažková (KNAV AV ČR) <https://orcid.org/0009-0004-5454-9514>
- Matyáš Hiřman (UK) <https://orcid.org/0000-0003-0153-9414>
- Martin Schätz (NTK/VŠCHT) <https://orcid.org/0000-0003-0931-4017>

Tato příručka vznikla na základě podkladů rozpracované FAIR data příručky. Tímto bychom rádi poděkovali autorům za jejich dosavadní práci.

Editori:

- Eva Hnátková (NTK/VŠCHT) <https://orcid.org/0000-0002-3237-9305>
- Zdenka Dudová <https://orcid.org/0000-0002-7615-1396>
- Dagmar Hanzlíková (UK) <https://orcid.org/0000-0001-9287-399X>

Kolektiv autorů:

- Eva Hnátková (NTK/VŠCHT) <https://orcid.org/0000-0002-3237-9305>
- Zdenka Dudová <https://orcid.org/0000-0002-7615-1396>
- Tereza Šimová (FLÚ AV ČR/ČZU) <https://orcid.org/0000-0002-1774-4335>
- David Šlosar (KNAV AV ČR/UISK FF UK) <https://orcid.org/0000-0001-5168-5327>
- Veronika Zemanová <https://orcid.org/0000-0002-2544-6872>
- Jiří Marek <https://orcid.org/0000-0003-2132-762X>
- Tereza Šírová (ÚI AV ČR) <https://orcid.org/0000-0001-5927-1065>
- Daniela Tršová <https://orcid.org/0000-0002-7019-8310>
- Kristýna Zychová (ČZU) <https://orcid.org/0000-0002-2021-7894>

DOI pro českou verzi: <https://doi.org/10.71495/hxhc-6f57>

DOI pro anglickou verzi: <https://doi.org/10.71495/ejrs-8t41>

*Tato práce vznikla za podpory projektu Národní repozitářová platforma pro výzkumná data, reg. č.: CZ.02.01.01/00/23\_014/0008787 financovaného z EFRR.*

*The work was supported from ERDF "National Repository Platform for Research Data" No. CZ.02.01.01/00/23\_014/0008787.*

Toto dílo podléhá licenci [Creative Commons Uveďte původ 4.0 Mezinárodní \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).



## Obsah

|                                                                                                                    |    |
|--------------------------------------------------------------------------------------------------------------------|----|
| Úvod .....                                                                                                         | 5  |
| FAIR principy v kontextu otevřené vědy a managementu výzkumných dat .....                                          | 5  |
| Výzkumná data a jejich management .....                                                                            | 5  |
| Otevřená výzkumná data .....                                                                                       | 5  |
| FAIR principy .....                                                                                                | 6  |
| FAIR vs. otevřená data vs. data management .....                                                                   | 7  |
| Dohledatelnost (findability) .....                                                                                 | 9  |
| F1. (Meta)datům je přiřazen jedinečný a perzistentní identifikátor .....                                           | 9  |
| F2. Data jsou popsána dostatečnými metadaty .....                                                                  | 12 |
| F3. Metadata obsahují identifikátor datové sady .....                                                              | 12 |
| F4. (Meta)data jsou registrována nebo indexována v prohledávatelných zdrojích .....                                | 12 |
| Dostupnost (accessibility) .....                                                                                   | 13 |
| A1. (Meta)data jsou vyhledatelná podle svého identifikátoru pomocí standardizovaného komunikačního protokolu ..... | 13 |
| A1.1 Protokol je otevřený, bezplatný a univerzálně implementovatelný .....                                         | 13 |
| A1.2 Protokol umožňuje v případě potřeby autentizační a autorizační postup .....                                   | 14 |
| A2. Metadata jsou přístupná, i když data již nejsou k dispozici .....                                              | 14 |
| Interoperabilita (Interoperability) .....                                                                          | 17 |
| I1. (Meta)data používají formální, dostupný, sdílený a široce aplikovatelný jazyk pro reprezentaci znalostí .....  | 18 |
| I2. (Meta)data používají slovníky, které se řídí FAIR principy .....                                               | 19 |
| I3. (Meta)data zahrnují kvalifikované reference na další (meta)data .....                                          | 19 |
| Opětovná využitelnost (reusability) .....                                                                          | 22 |
| R1. (Meta)data mají množství přesných a relevantních atributů .....                                                | 22 |
| R1.1 (Meta)data jsou zveřejňována pod jasnou a dostupnou uživatelskou licencí .....                                | 23 |
| Další právní překážky otevírání (meta)dat .....                                                                    | 24 |
| Problematika osobních údajů .....                                                                                  | 24 |
| Anonymizovaná data .....                                                                                           | 25 |
| Pseudonymizovaná data .....                                                                                        | 25 |
| R1.2 (Meta)data jsou spojena se svým původem (data provenance) .....                                               | 27 |
| R1.3 (Meta)data splňují standardy vědecké komunity pro daný obor .....                                             | 28 |
| Sémantická interoperabilita .....                                                                                  | 29 |
| Strojová funkčnost .....                                                                                           | 31 |
| Užitečné odkazy .....                                                                                              | 34 |

## Úvod

Dokument je určen pro pracovníky na pozici data steward se základní orientací a znalostmi v oboru. Přináší přehled a popis jednotlivých elementů FAIR principů (F = findability, A = accessibility, I = interoperability, R = reusability), které byly poprvé definovány a publikovány v roce 2016 a v současné době již jsou plnou součástí cyklu vědeckých dat. Je na ně kladen čím dál tím větší důraz.

Příručka vychází z podkladu rozpracované FAIR data příručky, rozvíjí ji a obohacuje o praktické části.

Cílem dokumentu je seznámit data stewardy s definicí FAIR principů. K lepšímu pochopení dané problematiky navíc přináší i praktické příklady v jednotlivých sekcích. Nad rámec základních principů nabízí příručka rozšířené kapitoly, které autoři pokládají za podstatné k dokreslení celé problematiky. Věnuje se tak i více technickým tématům jako anonymizace dat, sémantická interoperabilita nebo strojová funkčnost. Tato témata jsou podstatná v mezioborové spolupráci nebo i pro data stewardy působící v instituci s širokým vědeckým záběrem.

## FAIR principy v kontextu otevřené vědy a managementu výzkumných dat

### Výzkumná data a jejich management

Výzkumná data mají většinou velkou hodnotu, protože jejich získání vyžaduje mnoho času, úsilí, financí i jiných prostředků a jejich výstupy mohou mít velký dopad. V průběhu celého výzkumného procesu je důležitá kvalitní správa výzkumných dat (research data management, RDM), která zahrnuje různé postupy a strategie. Mezi jednotlivé kroky tzv. životního cyklu správy výzkumných dat patří plánování, organizace, dokumentace, ukládání a archivace dat. Tyto kroky zajišťují, že data budou používána efektivně a budou dostatečně zabezpečena proti ztrátě či zneužití. RDM je také klíčem k opětovnému využití a sdílení dat jak v rámci jednotlivých výzkumných týmů, tak na národní i mezinárodní úrovni. V oblasti RDM existuje řada návodů a instruktážních materiálů specifických pro jednotlivé obory či konkrétní instituce. V mnoha projektech i na mnoha institucích byla vytvořena nová pracovní pozice data steward (správce dat) zajišťující odbornou pomoc, koordinaci a dohled nad správou dat. Jako obecné zásady pro správu dat byly formulovány FAIR principy<sup>[1]</sup>, které zajišťují, aby data byla dohledatelná, dostupná, interoperabilní a opětovně využitelná.

### Otevřená výzkumná data

Otevřená věda (open science, OS) je koncept v rámci výzkumného procesu, který usiluje o otevřený přístup k vědeckým výsledkům s využitím nových digitálních technologií a nástrojů.<sup>[2]</sup>

<sup>3]</sup> Hlavním cílem tohoto konceptu je zlepšení dostupnosti a opětovné využití výsledků výzkumu, šetření finančních zdrojů vynakládaných na vědu a výzkum, možnost navázání nové spolupráce a zajištění transparentnosti výzkumu. Otevřená věda tak umožňuje úspěšné replikování výsledků, přináší větší důvěru ve výzkum, usnadňuje hodnocení výstupů vědecké diskuse a zvyšuje impakt vědců i výzkumných institucí. To vše přispívá ke zrychlení a zefektivnění výzkumu a zlepšení jeho kvality.<sup>[2, 4, 5]</sup>

Otevřená věda zahrnuje několik oblastí, mezi které patří otevřený přístup k vědeckým publikacím („open access“) a výzkumným datům („open data“).<sup>[6–9]</sup> Otevřená data představují obecně data, která jsou volně přístupná a kdokoli je může používat, upravovat a sdílet pro jakékoliv účely.<sup>[10]</sup> Někdy však mohou nastat legitimní situace, kdy přístup k datům musí být omezen a data ponechána z oprávněných důvodů uzavřená, např. v případě ochrany duševního vlastnictví, osobních údajů a soukromí, lidských práv, bezpečnosti, oprávněného komerčního zájmu apod.

Evropská komise doporučuje pro data vzniklá v rámci výzkumu financovaného z veřejných prostředků jejich zpřístupnění v souladu s FAIR principy a podle zásady „otevřená, jak jen možno, uzavřená, jen jak nutno“ („as open as possible, as closed as necessary“).<sup>[11]</sup> Otevřený přístup k datům podle FAIR principů je v různé míře stále častěji vyžadován poskytovateli financí (např. [Technologická agentura ČR](#), [Grantová agentura ČR](#) atd.) a stává se součástí politik univerzit a výzkumných institucí (např. [Politika správy výzkumných dat na Univerzitě Karlově](#)). Při publikování vědeckých článků vydavatelé často požadují zveřejnit data, na nichž jsou založeny výsledky výzkumu, a doložit informace o dostupnosti dat (např. [PLoS](#), [Springer Nature](#), [Wiley](#)).

Česká republika v rámci [Národní politiky výzkumu, vývoje a inovací 2021+](#) usiluje o zajištění otevřeného přístupu k výsledkům výzkumu a vývoje v souladu s evropskou legislativou. Dále se v ČR správa výzkumných dat a přístup k nim řídí zákonem [130/2002 Sb.](#), o podpoře výzkumu a vývoje.

### FAIR principy

FAIR principy byly publikovány v roce 2016 v článku [Wilkinson et al. 2016](#), kde jsou formulovány čtyři základní principy, které definují, že data mají být: **F**indable = dohledatelná, **A**ccessible = dostupná, **I**nteroperable = interoperabilní, **R**eusable = opětovně využitelná. Dále jsou principy rozpracovány do celkem patnácti prvků, které detailněji popisují, jaké charakteristiky mají vykazovat současná data, metadata a použité nástroje a infrastruktury, aby data a/nebo metadata byla co nejlépe dohledatelná, přístupná a opětovně využitelná.<sup>[1]</sup>  
<sup>[2]</sup> Samotné FAIR principy nejsou standardem ani normou a nepřikazují používání konkrétních nástrojů ani technologií, protože ty se mohou měnit v čase i oborově.<sup>[1]</sup> FAIR principy se nicméně stávají čím dál častěji součástí požadavků poskytovatelů financí nebo institucionálních politik pro zajištění dobré výzkumné praxe.<sup>[13]</sup>

FAIR principy jsou velmi obecné a nejsou oborově specifické, aby je bylo možné aplikovat na širokou škálu vědeckých výsledků. Vztahují se na data i metadata. Jednotlivé prvky FAIR spolu navzájem souvisejí, ale zároveň jsou nezávislé a oddělitelné. Díky této modularitě je možné je

využívat podle různých okolností, např. při práci s citlivými daty uvádět pouze metadata.<sup>[1]</sup> FAIR je popisováno jako spektrum, na kterém lze dosáhnout různého stupně „férovosti“ (odvozeno od slova FAIRness v rámci FAIR principů) a postupně tuto „férovost“ zlepšovat.<sup>[14]</sup>

Žádný z principů FAIR nepožaduje, aby data byla otevřená nebo zdarma. Nicméně je nutné stanovení jasných a transparentních podmínek přístupu a opětovného použití dat. FAIR data tedy nemusí být otevřená, ale měla by mít jasně přidělenou licenci.<sup>[1]</sup>

Data by měla splňovat FAIR principy z pohledu činností řízených jak lidmi, tak stroji. Zvláště důležitá je tzv. strojová funkčnost (machine actionability), kdy výpočetní systémy jsou schopné rozpoznat typ dat a jeho užitečnost, vyhodnotit podmínky použití podle licence a následně data zpracovat bez pomoci člověka nebo s jeho minimálním zásahem. Lidé nejsou schopni pracovat v takovém rozsahu a rychlosti, jaké vyžaduje objem a složitost současných vědeckých projektů a dat, a tato oblast je odkázána na činnost výpočetních systémů.<sup>[1]</sup>

FAIR principy mohou být užitečné pro různorodé cílové skupiny, které se jimi mohou řídit. Tyto skupiny zahrnují například výzkumníky, kteří chtějí sdílet svoje data či opětovně využívat data jiných, profesionální tvůrce dat, vývojáře softwarů, poskytovatele financí nebo vědeckou komunitu včetně data stewardů.

Procesu tvorby FAIR dat se říká „FAIRifikace“ a existují různé online návody na přípravu workflow, například:

- [FAIRToolkit \(Pistoia Alliance\)](#)
- [FAIRification framework \(FAIR Cookbook, ELIXIR, FAIRplus\)](#)
- [FAIRification Process \(GO FAIR\)](#)

K dispozici jsou i nástroje k hodnocení, do jaké míry jsou výzkumná data FAIR, například:

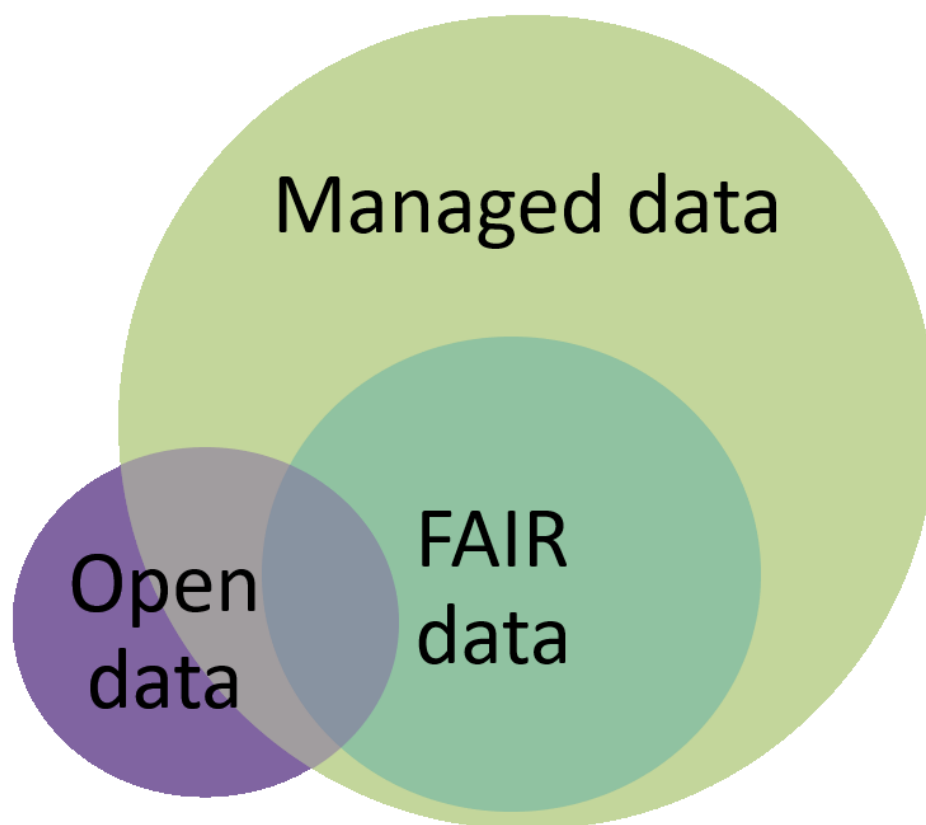
- [F-UJI](#)
- [FAIR DataSet Maturity](#)
- [FAIR Data Self Assessment Tool](#)
- [FAIR checker](#)

### FAIR vs. otevřená data vs. data management

FAIR, otevřená data a management výzkumných dat jsou tři různé koncepty, které se však navzájem překrývají. Každý klade důraz na něco jiného a nejlepších výsledků lze dosáhnout, pokud jsou používány společně. Důležité je v časných stádiích výzkumu začít s kvalitní správou výzkumných dat (ještě lépe v období přípravy výzkumných prací), jinak v pozdějších fázích už může být obtížnější docílit otevřených a FAIR dat. Principy otevřených a FAIR dat mohou zase pomoci zapojit výzkumníky do data managementu s motivací, že lépe dohledatelná data jsou více vidět a mohou zlepšit impakt výsledků výzkumu.

FAIR a otevřená data neznamenaí totéž. Data mohou být FAIR i otevřená najednou, pouze FAIR, pouze otevřená, nebo ani jedno z obou. Principy FAIR i otevřených dat jsou zaměřeny na sdílení dat. FAIR data však nemusejí být nutně otevřená. Ideální je vytvořit data co nejvíce FAIR

a otevřená, jak je to možné. Pokud v určitých situacích z oprávněných důvodů není možné poskytnout otevřená data, může být možné splnit alespoň požadavky FAIR. Ani FAIR, ani otevřená data neříkají nic o kvalitě dat.<sup>[15]</sup>



**Obrázek č. 1** Grafické znázornění prolínání tří různých konceptů dat.<sup>[72]</sup>

## Dohledatelnost (findability)

Hlavním předpokladem pro zpřístupnění a opětovné využití dat je jejich nalezení, kterému se bude věnovat tato kapitola. Význam dohledatelnosti (findability) je možné předvést na následujícím příkladu. Byl napsán vědecký článek. To stojí poměrně dost času, úsilí a práce. Přestože článek může mít excelentní kvalitu, pokud o něm nebude dostatečná informace, lidé se o něm nedozvědí, nedostane se do vědecké komunity, což může mít za následek malou citovanost, nízké povědomí o autorech a o prestiži instituce, ale také může zpomalit další vědecký výzkum, neboť na výsledky publikované v článku nebude možné navázat. Pokud ovšem článek bude vložen na vhodné místo (repozitář), bude dostatečně popsán (metadata) a bude mu přiřazen perzistentní identifikátor (díky němuž bude snadno dohledatelný), autoři tím přispějí k povědomí o své práci a ostatní budou moci díky identifikátoru článek jednoduše najít, šířit či citovat. Podobný postup by měl následovat i s vědeckými daty.

### F1. (Meta)datům je přiřazen jedinečný a perzistentní identifikátor

**Perzistentní (trvalý) identifikátor (zkráceně PID)** je nástroj, který slouží k jednoznačné identifikaci osob, organizací a dalších objektů (např. knih, článků, datových sad) v systému vědecké komunikace. Jedinečné odkazy spojené s entitou a/nebo jejími metadaty se v průběhu času nemění a umožňují trvalé vyhledávání, přístup, citování, propojování a opětovné využití výsledků výzkumu. Jedná se o celosvětově unikátní, trvalé a strojově použitelné digitální identifikátory využívající metadatové schéma.<sup>[16]</sup>

V tomto kontextu se identifikátory většinou skládají z internetového odkazu a jejich proklik v prohlížeči umožní přístup na webovou stránku (tzv. landing page) s dalšími údaji či přímo s daty samotnými. Identifikátory jsou nezbytné pro spolupráci mezi lidmi a stroji, která je klíčová pro vizi otevřené vědy. Kromě toho identifikátory pomohou ostatním správně citovat naši práci při opětovném použití našich dat. Některé digitální objekty už mohou mít PID přidělený, když se k nám dostanou. Ten je tedy vhodné používat konzistentně a další identifikátor už nepřidělovat.<sup>[17]</sup>

### Funkční požadavky na perzistentní identifikátory

#### PIDs musí být:

- globálně unikátní;
- globálně čitelné jako URI (uniform resource identifier) v souladu s http (např. <https://doi.org/10.48813/k2xs-y923>);
- trvalé – vytvářené tak, aby fungovaly bez ohledu na trvání systémů či organizací;
- spravovány dedikovanou organizací s definovanými mechanismy správy a rozhodování;
- vzájemně propojitelné/interoperabilní s jinými identifikátory prostřednictvím prvků metadat, které popisují jejich vztahy;
- obsahovat metadata, která popisují jejich nejdůležitější vlastnosti;
- indexovatelné a vyhledatelné podle svých metadatových prvků spolu se všemi ostatními důvěryhodnými identifikátory.<sup>[16]</sup>

V akademickém prostředí je využívána celá řada různých perzistentních identifikátorů k identifikaci objektů (např. autorů, digitálních objektů, publikací a časopisů, datových sad, organizací aj.). V tabulce níže jsou příklady nejčastěji používaných druhů unikátních identifikátorů. Některé identifikátory jsou ve fázi rozvoje, např. [RAiD](#) (výzkumné aktivity), [ePIC](#) (výzkumná data před zveřejněním), [ConfIDent](#) (odborné konference).

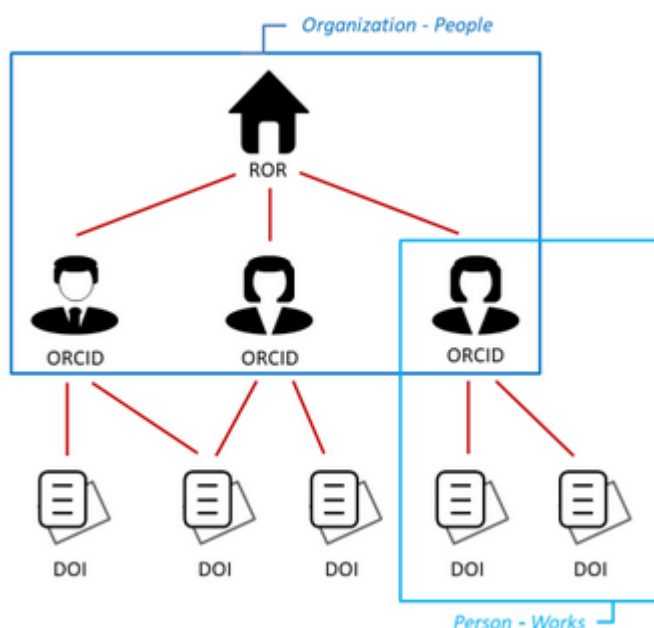
| Druh identifikátoru | Typ užití                                                                                                                                                              | Příklad                                                                                         |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| ORCID iD            | Osoba (autor, výzkumník/ce)                                                                                                                                            | <a href="https://orcid.org/0000-0001-8888-635X">https://orcid.org/0000-0001-8888-635X</a>       |
| ResearcherID        | Autorům publikace, která je indexována v databázi Web of Science (WoS), je automaticky vytvořen záznam autora, kterému je přiřazeno ResearcherID.                      | B-6035-2012                                                                                     |
| Scopus Autor ID     | Identifikátor a profil autora patřící k bibliografické a citační databázi Scopus, kterou provozuje vydavatelství Elsevier.                                             | 6603082428                                                                                      |
| DOI                 | Digitální objekty (články, datové sady, DMP (Data Management Plan, plán správy dat), audiovizuální záznamy, konferenční příspěvky, preprint, software, standardy atd.) | <a href="https://doi.org/10.1038/sdata.2016.18">https://doi.org/10.1038/sdata.2016.18</a>       |
| Handle              | Digitální objekty                                                                                                                                                      | <a href="https://hdl.handle.net/20.500.14391/1549">https://hdl.handle.net/20.500.14391/1549</a> |
| ROR                 | Výzkumné organizace                                                                                                                                                    | <a href="https://ror.org/028txef36">https://ror.org/028txef36</a>                               |
| IGSN                | Výzkumné vzorky                                                                                                                                                        | <a href="https://doi.org/10.60510/ICDP5054ESYI201">10.60510/ICDP5054ESYI201</a>                 |
| ISSN                | Časopisy                                                                                                                                                               | 1214-8790                                                                                       |
| ISBN                | Knihy                                                                                                                                                                  | 978-80-247-2279-5                                                                               |
| ISMN                | Hudebniny                                                                                                                                                              | 979-0-2600-0043-8                                                                               |

Tabulka č. 1 Příklady perzistentních (trvalých) identifikátorů.

**Příklady globálně jedinečných a trvalých identifikátorů:**

- Jedna konkrétní osoba na planetě Zemi má tento globálně jedinečný a trvalý identifikátor: <https://orcid.org/0000-0001-8888-635X>.
- Identifikátor, který jednoznačně odkazuje na článek o „férovosti“ FAIR principů: <https://doi.org/10.2218/ijdc.v12i2.567>.
- Lidský protein polycystin-1 má globálně jedinečný a trvalý identifikátor, který udává databáze UniProt: <http://purl.uniprot.org/uniprot/P98161>.
- Polycystická choroba ledvin typu 1 má celosvětově jedinečný a trvalý identifikátor uvedený v databázi OMIM: <http://omim.org/entry/173900>.<sup>[17]</sup>

**PID graf** – PIDs jsou důležité nejen pro jednoznačnou identifikaci publikace, datové sady (datasetu) nebo osoby, ale metadata těchto trvalých identifikátorů mohou zajistit jednoznačné propojení mezi trvalými identifikátory stejného typu, např. články v časopisech citující jiné články v časopisech, nebo různých typů, např. propojení výzkumného pracovníka a datových sad, které vytvořil.<sup>[18]</sup> Obrazově znázorněno v grafice níže.



**Obrázek č. 2** Grafické znázornění propojení různých typů perzistentních identifikátorů, které vede k usnadnění akademické komunikace.<sup>[19]</sup>

**Podpora perzistentních identifikátorů v České republice:**

- [Česká národní agentura ISBN a ISMN](#) provozovaná Národní knihovnou České republiky
- Národní centrum pro perzistentní identifikátory provozované Národní technickou knihovnou, které zastřešuje následující centra:
  - a) [Národní centrum ISSN](#)

- b) [Národní centrum ORCID](#), které vzniklo v roce 2023. Členské instituce mohou pracovat s ověřenými ORCID IDs výzkumníků/ic ve svých systémech a na profil ORCID svých výzkumníků zapisovat informace o afiliaci, publikační činnosti atd.
- c) [Národní centrum DOI](#), které vzniklo v roce 2023. Členské instituce mohou registrovat DOI pro různé typy objektů, včetně IGSN pro vzorky.<sup>[16, 20]</sup>

Ke snazší orientaci v problematice perzistentních identifikátorů mohou také pomoci webové stránky <https://identifikatory.cz>. Na nich lze nalézt aktuální informace o PIDs a druzích jejich podpory na národní úrovni.

## F2. Data jsou popsána dostatečnými metadaty

Šance nalézt výzkumná data je značně podpořena bohatým metadatovým popisem, jako jsou například popisné informace o kontextu, kvalitě, stavu nebo vlastnostech dat (detailněji popsáno v části R1). Bohatá metadata umožňují počítači automaticky provádět rešerše, které jsou pro člověka velmi časově náročné. Předpokladem principu F2 je, že uživatel by měl být schopen nalézt data na základě informací poskytnutých v metadatech, a to i v případě, že data nemají přidělen perzistentní identifikátor. Zvláště při rešeršní činnosti zvyšuje kvalitní metadatový popis šanci na objevení a další využití dat.<sup>[17]</sup>

## F3. Metadata obsahují identifikátor datové sady

Datová sada a popisná metadata jsou obvykle samostatné soubory. Jak už samotný název tohoto principu napovídá, je důležité, aby tyto dva celky byly propojené. Tedy, že globálně jedinečný a trvalý identifikátor datové sady by měl být obsažen také v metadatech. V případě, že data samotná již nejsou (nebo nemohou) být k dispozici, je doporučeno, aby byla k dispozici alespoň zmínka o jejich existenci. Mnoho repozitářů již v dnešní době automaticky generuje PIDs při vložení dat a ty lze pro tento účel použít.<sup>[17]</sup>

## F4. (Meta)data jsou registrována nebo indexována v prohledávatelných zdrojích

Samotné identifikátory a bohatá metadata nezajistí „naležitelnost“ dat na internetu bez dalšího kroku. Kvalitní datové zdroje mohou zůstat nevyužity jen proto, že o jejich existenci nikdo neví. Pokud není datová sada dostupná, pak ji nikdo (ani žádný stroj) nemůže objevit. Proto je zapotřebí, aby data i metadata byla vložena do vhodného důvěryhodného (datového) repozitáře, který je přístupný, prohledávatelný a indexovaný. Podle potřeby lze použít buď institucionální, disciplinární (oborový), nebo multioborový (též generický, obecný nebo sirotčí) repozitář.<sup>[16, 17]</sup> K výběru vhodného repozitáře může posloužit registr datových repozitářů [Registry of Research Data Repositories](#). V tomto webovém nástroji je možné prohledávat prostřednictvím názvu, oboru/vědecké disciplíny, země, ve které je repozitář provozován, nebo podle typu obsahu (např. datové sady, textové soubory, strukturovaný text, obrázky, konfigurační data a další). Další nástroj, který může pomoci nalézt vhodný repozitář, je [FAIRsharing](#), kde je možné prohledávat v sekci Databases. S výběrem vhodného repozitáře autorům také pomohou doporučení u některých vydavatelů a časopisů, např. <https://www.nature.com/sdata/policies/repositories>.

## Dostupnost (accessibility)

Tento princip zajišťuje naplňování dostupnosti dat a metadat. Z reálného hlediska nemohou být všechna data otevřeně přístupná (z důvodů osobních či citlivých dat, národní bezpečnosti, komercializace apod.), proto se doporučuje (a v projektových podmínkách často určuje) postupovat podle pravidla „otevřená, jak jen možno, uzavřená, jen jak nutno“ („as open as possible, as closed as necessary“).

Dnes je již poměrně zavedenou praxí, že repozitáře umožňují rozlišovat mezi způsobem zveřejnění tak, aby data byla dostupná komukoliv, nebo tak, že jsou určena pouze pro omezený okruh uživatelů. Pokud není možné zpřístupnit data samotná, jejich metadata by měla být přístupná v repozitáři bez omezení, neboť i informace o existenci dat s restrikcemi je velmi hodnotná. To vše za využití standardních komunikačních protokolů.

### A1. (Meta)data jsou vyhledatelná podle svého identifikátoru pomocí standardizovaného komunikačního protokolu

Identifikátory používané pro metadata jsou především DOI, případně URN, PURL, HDL (Handle). Přidělení DOI zajišťují např. repozitáře, které zaregistrují DOI na základě založení záznamu či uložení datasetu. Identifikátor ideálně odkazuje na tzv. landing page. Standardizovaným komunikačním protokolem se zde rozumí především HyperText Transfer Protocol – http(s).<sup>[21, 22]</sup>

| Název   | Hodnota                 | Odkazování                                                                                                      |
|---------|-------------------------|-----------------------------------------------------------------------------------------------------------------|
| DOI     | 10.5281/zenodo.4420115  | <a href="https://doi.org/10.5281/zenodo.4420115">https://doi.org/10.5281/zenodo.4420115</a>                     |
| Handle  | 1813/7895               | <a href="https://hdl.handle.net/1813/7895">https://hdl.handle.net/1813/7895</a>                                 |
| URN:NBN | urn:nbn:cz:tst02-000008 | <a href="https://resolver.nkp.cz/urn:nbn:cz:tst02-000008h">https://resolver.nkp.cz/urn:nbn:cz:tst02-000008h</a> |

**Tabulka č. 2** Příklady perzistentních (trvalých) identifikátorů.

#### A1.1 Protokol je otevřený, bezplatný a univerzálně implementovatelný

Pro přístup k meta(datům) je nutné zajistit přístup pomocí otevřeného protokolu bez finančního nebo implementačního omezení. Ve velké většině případů tak jde o využití HTTP nebo HTTPS protokolu, je však možné využít i API, FTP, FTPS, SFTP nebo SCP protokolů a dalších. Nutno dodat, že v dnešní době není použití protokolů HTTP a FTP doporučováno, protože veškerá komunikace probíhá těmito protokoly v nešifrované podobě. V případě přístupu prostřednictvím např. API rozhraní je důležité zajištění potřebné specifikace možných operací, dokumentace a v případě citlivých údajů s omezeným přístupem i autentizace přistupujícího, což zajistí možnost zajištění automatizovaného přístupu k datům, popřípadě jejich integraci do dalších služeb pro autorizované entity.<sup>[23]</sup>

Příklad dokumentace API: <https://docs.ckan.org/en/2.10/api/index.html>

## A1.2 Protokol umožňuje v případě potřeby autentizační a autorizační postup

Ne všechna data mohou být přístupná pro všechny bez ověření totožnosti. Je tak nutné zajistit možnost bezpečné autentizace a autorizace při přístupu k výzkumným datům. Možnosti přístupu by měly být jasně uvedeny v metadatech nebo v ReadMe souboru. Zahrnuty by měly být informace o tom, kdo data může používat, případně jaká jsou omezení v užívání a také jak je možné získat přístup k datům a za jakých podmínek. Autentizaci a autorizaci uživatelů většinou řídí samy repozitáře. Při programování vlastních systémů je vhodné využívat standardizované protokoly, jako je např. OAuth, SAML.<sup>[24]</sup>

Kromě přímé autentizace uživatele pomocí hesla lze využít federativní identity (např. eduID.cz, mojeID, eduGAIN). Při práci s API mohou být uživatelé autorizováni pomocí API klíčů nebo OAuth tokenů. Organizace často využívají Role-Based Access Control (RBAC), kde uživatelé získávají oprávnění na základě role v systému.

## A2. Metadata jsou přístupná, i když data již nejsou k dispozici

Ne všechna data mohou být přístupná pro všechny, ne všechna data mohou být dostupná po neomezenou dobu. V takových případech by měla zůstat přístupná alespoň metadata, která jsou tak svědectvím např. o měřeném experimentu. Metadata je důležité (ba až nutné) ukládat a publikovat ve standardních formátech typu Dublin Core nebo RDF, které mají dostupnou online specifikaci. Metadata by měla být vždy přístupná prostřednictvím trvalých identifikátorů typu DOI nebo Handle.<sup>[25]</sup>

## Další aspekty dostupnosti dat

### Dlouhodobá dostupnost výzkumných dat

Výzkumná data mohou být relevantní mnoho let po svém vzniku, a proto je nezbytné zajistit jejich ochranu a správu tak, aby byla použitelná i v budoucnosti. Tato problematika není důležitá jen z pohledu vědeckého poznání, ale i z hlediska efektivity a opakovatelnosti výzkum. Správné zachování dat umožňuje jejich znovupoužití, minimalizuje potřebu drahých opakovaných experimentů a zvyšuje věrohodnost vědeckých výsledků.

Kromě správného výběru formátu, který je detailněji řešen v následující kapitole Interoperabilita, existují i další opatření, která mohou pomoci zajistit dlouhodobou dostupnost dat. Významnou roli zde hrají repozitáře, jež mohou poskytovat specializované služby pro ochranu a uchování dat v čase. Při výběru repozitáře je proto klíčové ověřit, zda nabízí garanci dlouhodobé dostupnosti, jaké má strategie ochrany dat a zda splňuje mezinárodní standardy pro digitální archivaci. Vhodné repozitáře často disponují certifikacemi, jako je [CoreTrustSeal](#), které potvrzují jejich schopnost uchovávat data spolehlivým a dlouhodobě udržitelným způsobem.

### Jednotlivé prvky, které zaručují dlouhodobou dostupnost nebo jí pomáhají:

#### Bitová integrita dat

Jedním z klíčových prvků zajišťujících dlouhodobou dostupnost dat je kontrola jejich neporušenosti, která se provádí pomocí bitové integrity. Tento proces zahrnuje výpočet

kontrolních součtů, jako jsou algoritmy MD5, SHA256 nebo modernější SHA3, které umožňují ověřit, zda nedošlo k neautorizované změně nebo poškození dat během jejich přenosu, ukládání nebo archivace.

Kontrolní součet se vytváří v okamžiku nahrání souboru do repozitáře a lze jej v budoucnu porovnávat s aktuální verzí souboru, čímž se zajistí, že data nebyla nepozorovaně pozměněna. Pokud by došlo k jakémukoli narušení integrity, systém repozitáře by měl umožnit obnovu dat z redundantních záloh. Řada repozitářů, jako například [Zenodo](#) nebo [Dataverse](#), aktivně implementuje mechanismy pro detekci a opravu chyb v souborech, čímž minimalizuje riziko ztráty nebo nevratného poškození výzkumných dat.

Kromě kontrolních součtů je vhodné využívat také techniky jako **audit integrity dat**, spočívající v pravidelném přepočítávání [hash hodnot](#) souborů a porovnávání s původními hodnotami. Některé pokročilé systémy archivace umožňují dokonce **automatickou opravu dat**, pokud je nalezena odchylka od původního souboru, což výrazně přispívá k ochraně informací před postupnou degradací.<sup>[26]</sup>

### Migrační strategie

Digitální technologie se neustále vyvíjejí a některé formáty souborů mohou v průběhu času zastarat, což může vést k problémům s jejich čitelností a využitelností. Proto je důležité mít zavedenou **migrační strategii**, která zajistí, že data budou pravidelně převáděna do aktuálních a dlouhodobě udržitelných formátů. Tento proces je klíčový zejména pro specifické vědní obory, které využívají proprietární nebo méně běžné formáty souborů, jejichž podpora by v budoucnosti mohla zaniknout.

Migrační strategie obvykle zahrnují několik hlavních opatření. Jedním z nich je **monitorování vývoje softwarových standardů** a pravidelná kontrola kompatibility formátů s novými verzemi programů. Pokud je zjištěno, že u některého formátu hrozí zastarávání, repozitáře mohou automaticky převádět soubory do novějších otevřených formátů, jako je například **CSV místo proprietárních tabulkových formátů**, nebo **XML a JSON místo specifických binárních datových formátů**.

Důležité je také zajistit, že migrace dat nebude mít negativní dopad na jejich obsahovou správnost. Každý převod dat by měl být doprovázen podrobnou dokumentací, jež popisuje provedené změny a zachovává informace o původním formátu. V tomto směru se vyplatí sledovat repozitáře, které mají migrační strategie jasně definované a transparentně komunikované, jako je například **DataverseNO Preservation Policy**, která stanovuje přesné postupy pro pravidelné konverze dat do aktuálních formátů.<sup>[27]</sup>

### Geografická redundance

Bez ohledu na to, jak spolehlivá se může zdát datová infrastruktura, vždy existuje riziko, že uložená data mohou být nenávratně ztracena v důsledku technických havárií, přírodních katastrof nebo kybernetických útoků. Proto je jedním z nejefektivnějších opatření k zajištění dlouhodobé dostupnosti dat **geografická redundance**, tedy uchovávání datových kopií v různých geografických lokalitách.

Tato strategie spočívá v tom, že data nejsou uložena pouze v jednom repozitáři nebo datovém centru, ale jsou distribuována mezi více nezávislých úložišť umístěných v odlišných regionech. V případě výpadku jednoho úložiště mohou být data obnovena z jiného místa, což zajišťuje jejich **vyšší odolnost vůči katastrofám**, jako jsou požáry, zemětřesení nebo záplavy.



## Interoperabilita (Interoperability)

**Interoperabilita (meta)dat** je zjednodušeně vlastnost fungovat a interagovat s jinými daty, s aplikacemi a operačními systémy pro potřeby jejich zpracování, analýzy, porovnání a ukládání. Z toho plynou tyto požadavky:

- data poskytovat v běžně používaných a ideálně otevřených formátech;
- poskytovaná metadata se řídí relevantními standardy;
- pokud je to možné, používat pro popis řízené slovníky, tezaury a ontologie;
- poskytovat reference a odkazy k dalším souvisejícím datům.<sup>[28–30]</sup>

Příklady dobré praxe v rámci správy dat zahrnují zohlednění aspektu interoperability jak u samotných dat (řízené seznamy, problémy se syntaxí či formátem), tak i metadat digitálního souboru (schémata, názvy proměnných, metadatové nástroje).

Společně dohodnutý přístup k poskytování evropských veřejných služeb interoperabilním způsobem definuje [The European Interoperability Framework \(EIF\)](#). Základní úrovně interoperability jsou:

### 1. Právní

Právní interoperabilitou je myšlen sdílený výklad a chápání zákonů upravujících výměnu informací a spolupráci (např. zda je dovoleno předávat informace o účastnících výzkumu, jak přistupovat k ochraně soukromí nebo anonymizaci).<sup>[31, 32]</sup>

### 2. Organizační

Podle EIF se organizační interoperabilitou rozumí způsob, jakým organizace sladují své obchodní procesy, odpovědnost a svá očekávání k dosažení vzájemně prospěšných cílů. Vzhledem k celkovému dohodnutému cíli otevřené vědy by se dle dokumentu [EOSC Interoperability Framework](#) organizační interoperabilita měla zaměřit na dokumentaci, integraci nebo sladění procesů různých organizací poskytujících služby, aby bylo zajištěno, že výzkumní pracovníci mohou dosáhnout svých cílů v oblasti otevřené vědy. Mělo by být také jasné, kdo je odpovědný za poskytování (a také za vývoj, údržbu a správu) společných služeb interoperability, jako jsou mimo jiné katalogy služeb, registry a společné služby PID.<sup>[32–35]</sup>

### 3. Sémantická

Smyslem sémantické interoperability dle Evropské komise je zajištění toho, aby byl zachován a pochopen přesný formát a význam vyměňovaných dat a informací při výměně mezi jednotlivými stranami. Podle EIF zahrnuje sémantická interoperabilita jak sémantické, tak syntaktické aspekty.

Sémantický aspekt představuje společné základní modely a kodifikaci dat, včetně používání datových prvků se standardizovanými definicemi z veřejně dostupných souborů hodnot a kódovacích slovníků (terminologický standard), což umožňuje sdílené porozumění a význam.

Tyto slovníky jsou ve většině případů dohodnuté na mezinárodní úrovni, a tak zajišťují jednoznačnou interpretaci informací.

Syntaktický aspekt se týká popisu přesného formátu informací, které mají být poskytnuty z hlediska gramatiky a formátu.<sup>[32]</sup>

#### 4. Technická

Technická interoperabilita by měla mít v praxi schopnost systémů či aplikací přijímat (meta)data od sebe navzájem a od jiných subjektů. Dále si pod tím lze představit provádění daných úkolů vhodným a uspokojivým způsobem bez nutnosti dodatečných požadavků na zpracování dat či zásahu „operátora“ v rámci úplné automatizace činností výměny dat.

Aspekty technické interoperability zahrnují specifikace rozhraní, služby propojení, služby integrace dat, prezentace dat a jejich zpracování a bezpečné komunikační protokoly. Hlavní překážkou interoperability jsou v první řadě starší systémy neschopné vzájemného propojení s novějšími systémy či obecná kompatibilita přístrojů mezi sebou.<sup>[32]</sup>

Praktické příklady: Standardy softwarů, fyzické hardwarové komponenty, systémy a různé platformy podporující interakci počítače s počítačem.<sup>[36]</sup>

### 11. (Meta)data používají formální, dostupný, sdílený a široce aplikovatelný jazyk pro reprezentaci znalostí

Počítače by měly být schopné vzájemné výměny a interpretace dat. Data by obecně měla být strojově čitelná bez nutnosti využití specializovaných (proprietárních) či ad hoc algoritmů, překladačů nebo mapování. V této souvislosti interoperabilita znamená, že každý počítačový systém má mít alespoň minimální znalost o formátech dat z druhého systému. Struktura dat, která jsou považována za FAIR, musí být zdokumentována jak způsobem čitelným pro člověka, tak i strojově čitelným schématem.

Optimální formou standardizace je definice datového formátu nezávisle na konkrétním softwaru s detailním popisem struktury formátu a přesným vysvětlením jeho částí. Vhodné je definici doplnit nástrojem pro validaci dat, který umožní určit, zda vytvořená data odpovídají danému formátu, např. XML validátor pro XML. Nejvyšší formou standardizace jsou formáty schválené uznávanou mezinárodní standardizační organizací, např. standard pro formát [CSV](#), [ISO](#), [OASIS – Open Document](#) apod.<sup>[37–39]</sup> Takto standardizované datové formáty jsou vhodné pro výměnu dat z hlediska jejich dlouhodobé čitelnosti a interpretovatelnosti.

Hojně využívané datové formáty s podporou implementovanou v široké škále softwarů různých výrobců mají vyšší pravděpodobnost **dobré interoperability a dlouhodobé čitelnosti**. Pozitivní vliv na rozšíření datových formátů a jejich implementaci v softwaru různých výrobců má jejich otevřenost a dobrá definice (standardizace).<sup>[40]</sup>

**Příklady:**

Vhodné otevřené formáty:

- tabulková data – CSV, TSV, JSON, RDF (JSON-LD, Turtle, ...);
- hierarchická data – XML, JSON, RDF (JSON-LD, Turtle, ...);
- grafová a propojená data – RDF (JSON-LD, Turtle, ...);
- geodata (prostorová data) – GeoJSON, ESRI Shapefile, OGC GML, OGC GeoPackage.

Doporučené jazyky pro definici schémat:

- CSV – [CSV schema](#), [CSV on the Web](#);
- XML – [XML schema](#);
- JSON – [JSON schema](#);
- RDF – [RDFS](#), [OWL](#), [SHACL](#).<sup>[41–48]</sup>

Některé organizace na svých stránkách přímo uvádějí doporučené formáty, např. [DANS](#) nebo [UK Data Service](#).

## 12. (Meta)data používají slovníky, které se řídí FAIR principy

Řízené slovníky používané k popisu datových souborů musí být ověřovány a definovány za využití globálně jedinečných a perzistentních identifikátorů pro jednotlivé výrazy a ontologie. Vše musí být jednoduše nalezitelné a přístupné pro kohokoliv, kdo využívá dané datové soubory.

**Příklady slovníků „Knowledge Organisation Systems Types“:** AGROVOC, GeoNames

**Příklady metadatových slovníků:** Dublin Core, W3C Data Catalog Vocabulary, Multi-Crop Passport Descriptors

### Jak najít nejužitečnější dostupné slovníky?

1. FAIRsharing – <https://fairsharing.org>  
Registr terminologických artefaktů, modelů/formátů, pokynů pro vykazování a schémat identifikátorů.
2. Linked Open Vocabularies (LOV) – <https://lov.okfn.org/dataset/lov>
3. The Basel Register of Thesauri, Ontologies and Classifications (BARTOC) – <http://bartoc.org> <sup>[49–51]</sup>

## 13. (Meta)data zahrnují kvalifikované reference na další (meta)data

Kvalifikovaná reference představuje křížový odkaz, který vysvětluje jeho záměr. Cílem je vytvoření co nejvíce smysluplných vazeb mezi (meta)datovými zdroji, aby byla obohacena

kontextová znalost o datech, která je vyvážena časem a energií vynaloženými na vytváření kvalitního datového modelu.

Konkrétně jde o to, že je třeba určit, zda jedna datová sada navazuje na jinou datovou sadu, zda jsou k doplnění dat zapotřebí další datové sady nebo zda jsou doplňkové informace uloženy v jiné datové sadě. Zejména je nutné popsat vazby mezi soubory údajů. Dále je třeba řádně uvést všechny soubory údajů (tj. včetně jejich globálně jedinečných a perzistentních identifikátorů). V neposlední řadě je možné na sebe navzájem odkazovat i samotné PIDs, během jejich vzniku v rámci jejich metadatových popisů. Tyto provazby mezi datovými sadami dávají vzniknout mj. tzv. propojeným datům, považovaným za rozšířenou interoperabilitu dat, viz vztah linked data vs. FAIR data popsány na stránkách [The Road to FAIR](#).<sup>[52]</sup>

Paradigma propojených dat předpokládá čtyři pravidla podle [Road to FAIR](#):

1. Používat URI jako názvy k identifikaci objektů.
2. Používat URI HTTP, aby bylo možné tyto názvy vyhledávat.
3. Při nalezení URI je žádoucí poskytnout užitečné informace pomocí standardů (RDF, [SPARQL](#)). Dotazovací jazyk SPARQL lze například prakticky využít i prostřednictvím wikidat.
4. Uvést odkazy na další URI, aby bylo možné identifikovat další objekty.<sup>[52]</sup>

Paradigma propojených dat podle **Jonathana Blaneyho**, „[Introduction to the Principles of Linked Open Data \(LOD\)](#)“:

1. Používat uznávaný standardní formát LOD. Pro správné fungování systému LOD musí být data strukturována pomocí uznávaných standardů, aby je počítače, které se na data dotazují, mohly konzistentně zpracovávat. Existuje řada formátů LOD, z nichž některé jsou popsány níže.
2. Odkazovat na entitu stejným způsobem jako ostatní lidé. Pokud existují údaje o stejné osobě/místě/věci na dvou nebo více místech, je třeba zajistit, že na osobu/místo/věc je odkazováno ve všech případech stejným způsobem.
3. Otevřeně sdílet svá data. Otevřeným přístupem je myšleno, aby je mohl kdokoli použít bez placení poplatků a ve formátu, který nevyžaduje proprietární software.<sup>[53]</sup>

V roce 2010 vydal **Tim Berners-Lee** systém hodnocení otevřenosti propojených dat, tzv. [5-Star Linked Open Data](#). Zastává názor, že je dobré publikovat data otevřeně, s využitím otevřených formátů a veřejných standardů. Data získávají na informační hodnotě při propojení s jinými daty.<sup>[54]</sup>



Obrázek č. 3 Systém hodnocení propojených dat podle Tima Berners-Leeho.<sup>[54]</sup>

**Systém hodnocení:** 1–5 hvězdiček (nejlepší je 5)

★ Zpřístupnit své věci na webu (v jakémkoli formátu) pod [otevřenou licenci](#).

- viz [Jak stanovit podmínky užití datových sad?](#)

★★ Zpřístupnit je jako strukturovaná data (např. v MS Excel namísto screenshotu tabulky).

- Datová sada je poskytována ve strojově čitelném formátu, který umožňuje automatizované strojové zpracování.

★★★ Zpřístupnit je v otevřeném formátu (např. CSV místo MS Excel).

★★★★ Používat URI, aby bylo možné na data jednoznačně odkazovat.

★★★★★ Propojit svá data s jinými daty k získání kontextu.

- Datová sada splňuje standard propojených dat.<sup>[55–57]</sup>

## Opětovná využitelnost (reusability)

Cílem jakéhokoliv otevírání obsahu, obzvláště digitálních objektů, není pouhé zpřístupnění daného objektu, ale zejména jeho opětovné užití, anglicky „re-use“. Je to právě opětovné využití již existujících zdrojů pro přípravu, rozvoj nebo budování něčeho nového, co přináší ekonomické a další benefity. Aby se tak mohlo stát, existují právní, technická a manažerská (kurátorská) doporučení.

V návaznosti na principy FAIR se jedná o:

1. Uvedení vhodné (co nejotevřenější) licence k dané datové sadě, pokud je to potřeba, případně indikace, že se k dané datové sadě žádná práva neváží (viz kapitola *R1.1 (Meta)data jsou zveřejňována pod jasnou a dostupnou uživatelskou licenci*).<sup>[58]</sup>
2. Návrhy řešení možných překážek pro zveřejňování (meta)dat (*R1.2 Další právní překážky otevírání (meta)dat*).<sup>[58]</sup>
3. Uvedení původu datové sady a jejích úprav (*R1.3 (Meta)data jsou spojena se svým původem (data provenance)*).<sup>[59]</sup>
4. Využití všeobecně známých oborových standardů po technické i metodické stránce pro metadatový popis datové sady (*R1.4 (Meta)data splňují standardy vědecké komunity pro daný obor*).<sup>[60]</sup>

### R1. (Meta)data mají množství přesných a relevantních atributů

Pro snazší nalezení a opětovné využití dat je klíčové, aby k datům byla připojena bohatá metadata popisující jejich jednotlivé atributy. Co nejpřesnější popis dané datové sady přispívá ke zvýšení schopnosti uživatele (stroje nebo člověka) rozhodnout, zda jsou data v konkrétním kontextu skutečně užitečná. Podrobný a bohatý popis (meta)dat vede k možnosti být automaticky (s minimálním lidským úsilím) propojen nebo integrován s relevantními datovými zdroji.<sup>[61]</sup>

Publikovaná (meta)data by měla odkazovat na své zdroje s dostatečně podrobnými metadaty a informacemi o původu (provenance), aby bylo možné je správně citovat.

Je vhodné, aby byla poskytnuta nejen metadata, která zvyšují nalezitelnost datové sady (princip F), ale také metadata, která bohatě popisují kontext, v němž byla data generována. Tím se rozumí doplňující dokumentace k vědeckému procesu. Tvůrce dat by se neměl pokoušet předvídat potřeby spotřebitele/uživatele dat.

Obecně platí, že autor (meta)dat by měl být při poskytování metadat co nejštědřejší, a to včetně informací, které se mohou zdát irelevantní.

V rámci tohoto principu je potřeba se vypořádat například s těmito tématy:

- v jakém období / v jaké lokalitě sběr dat probíhal;
- informace o nástroji nebo verzi softwaru, který byl použit ke sběru či analýze dat;
- demografická data o respondentech, způsob oslovování respondentů;

- seznam proměnných / názvů polí v tabulce a jejich popis (např. hodnoty, kterých mohou nabývat);
- vysvětlivky zkratk či kódových označení;
- laboratorní deník;
- formulář dotazníku;
- vzor informovaného souhlasu;
- licenční podmínky pro využití nasbíraných dat;<sup>[62]</sup>
- a mnoho dalších informací.

### R1.1 (Meta)data jsou zveřejňována pod jasnou a dostupnou uživatelskou licenci

Datové sady, které jsou autorskými díly podle zákona č. 121/2000 Sb., autorského zákona (dále jen „AutZ“), jsou automaticky ze zákona chráněny autorským právem. To znamená, že autorské právo chrání autorovo duševní vlastnictví zde představuje určitou **právní bariéru pro opětovné využití dat** dalšími osobami. **Jako hlavní nástroj pro odstranění těchto právních bariér slouží veřejné licence.** Nejznámější a nejpoužívanější sadou veřejných licencí jsou pak licence [Creative Commons](#). Osoba oprávněná k poskytnutí licence (tedy typicky autor, případně zaměstnavatel u zaměstnaneckého díla) může zvolenou veřejnou licenci připojit k datové sadě, čímž kužití datové sady poskytuje oprávnění předem neurčitěmu a neomezenému okruhu uživatelů. Smlouva je pak uzavřena tím, že uživatel začne data užívat v souladu s podmínkami připojené licence.

Každá licence Creative Commons v sobě zahrnuje některé ze **čtyř licenčních prvků** (prvek BY je přítomen vždy), indikujících podmínky, které je nutno při užití licencované datové sady dodržet:

- **BY** – musí být uvedeno jméno autora, název díla, licence a zdroj;
- **SA** – musí být zachována stejná licence;
- **NC** – nesmí se užívat komerčně;
- **ND** – není dovoleno vytvářet odvozená díla.<sup>[63]</sup>

Použití co nejméně restriktivních veřejných licencí má svoje opodstatnění v tom smyslu, že výstupní vědecká data jednoho experimentu jsou pro jiný experiment daty vstupními. S ohledem na to **není** vhodné používat licence obsahující prvky NC a ND, neboť tyto dva prvky brání efektivnímu opětovnému využití výzkumných dat.

Naopak je žádoucí:

- V co nejvyšší míře využívat nejméně restriktivní licenci [Creative Commons Uveďte původ 4.0](#) (CC BY 4.0).
- Pro databázi chráněnou zvláštním právem pořizovatele databáze je vhodné použít [Creative Commons Zero Universal Dedication \(CC0\)](#)<sup>1</sup>. Právo pořizovatele databáze je výjimkou, u které lze tuto licenci využít.

<sup>1</sup> Jde o zvláštní variantu licence Creative Commons, která představuje tzv. waiver – vzdání se práva. Dílo opatřené CC0 je tedy každému k dispozici pro volné užití.

- Pokud datová sada **není** autorským dílem ani databází ve smyslu AutZ (jde o prostá data, typicky výsledky přístrojového měření), je vhodné do metadat uvést informaci, že datová sada nepodléhá ochraně autorského práva, a tedy že je volně k užití.

Jako vodítko pro výběr vhodné licence Creative Commons pro vlastní výstup může sloužit tento nástroj: <https://chooser-beta.creativecommons.org/>

## Další právní překážky otevírání (meta)dat

Praktické aspekty otevírání nebo „FAIRifikace“ vědeckých dat jsou součástí životního cyklu dat a je nutné na ně myslet v průběhu tvorby, zpracování a následného publikování datových sad. Před zpřístupněním vědeckých dat k opětovnému užití je nutné mít na mysli, že kromě autorskoprávní ochrany (viz kapitola R1.1) mohou data spadat pod některý z dalších **ochranných režimů**, které stanoví právní předpisy.

Jde zejména o případy, kdy:

1. **Výzkumná data obsahují osobní údaje.** Zpřístupnění osobních údajů ve formě otevřených dat je vyloučeno. Pokud by však data byla v rámci jejich zpracování anonymizována, je možné řádně anonymizovaná data zpřístupnit, protože již nespádají do ochranného režimu GDPR, více v kapitole Problematika osobních údajů.
2. **Výzkumná data obsahují jiná zákonem chráněná data.** Zpřístupnění takových výzkumných dat (např. obchodního tajemství, utajovaných informací) je vyloučeno. Jejich nezveřejnění je v souladu se zásadou „otevřená, jak jen možno, uzavřená, jen jak nutno“ („as open as possible, as closed as necessary“) zcela legitimní.

Pokud datová sada není z jakéhokoliv důvodu určena pro zpřístupnění nebo má být zpřístupněna pouze částečně, důvod pro nezveřejnění datové sady, případně podmínky pro zvýšení „férovosti“ dat by měly být uvedeny v metadatech popisujících datovou sadu.

## Problematika osobních údajů

Přítomnost osobních údajů (tj. jakéhokoliv údaje, který může vést k identifikaci fyzické osoby) v datové sadě může představovat překážku při jejím zpřístupnění. S osobními údaji musí být vždy nakládáno v souladu s obsahem souhlasu se zpracováním osobních údajů, který uděluje jednotlivý účastník výzkumu. Náležitosti souhlasu jsou dány nařízením EU č. 2016/679 o ochraně osobních údajů (GDPR), zároveň je v souhlasu většinou upraveno, v jaké formě (tj. identifikovatelné, anonymní nebo pseudonymizované, viz níže), za jakým účelem a s kým lze data sdílet, případně po jakou dobu mohou být uchovávána.

Jestliže tedy nelze datovou sadu s identifikovatelnými údaji na základě souhlasu zpřístupnit, je možné se s touto překážkou vypořádat anonymizací dat.

Údaje, které umožňují identifikaci fyzické osoby, se nazývají identifikátory. Identifikátory se dají rozdělit do dvou skupin na přímé a nepřímé.<sup>[64]</sup>

### Přímé identifikátory

- identifikátory, které umožňují identifikaci samy o sobě
- např. rodné číslo, jméno a příjmení atd.

### Nepřímé identifikátory

- nelze vztáhnout ke konkrétní osobě bez použití dodatečných informací
- např. pohlaví, věk, adresa, IP adresa, typ onemocnění atd.

### Anonymizovaná data

- Neumožňují přímou či nepřímou identifikaci fyzické osoby, a to ani zpětně.
- Nelze je proto vztáhnout ke konkrétní osobě (tj. nelze je přiřadit k určitému účastníkovi výzkumu).<sup>[64]</sup>
- Nevztahuje se na ně tudíž nařízení EU č. 2016/679 o ochraně osobních údajů (GDPR).

Anonymizace se provádí především odstraněním **přímých identifikátorů**. Nevýhodou anonymizace dat je ztráta smysluplnosti a/nebo další využitelnosti dat. Proto v některých případech data anonymizovat nelze. Výhodou je, že anonymizovaná data je možné volně sdílet mezi výzkumníky i bez nutnosti získat souhlas účastníka výzkumu se zpracováním osobních údajů (GDPR), jelikož už se nejedná o data s osobními údaji.<sup>[66]</sup>

Anonymizaci je možné provádět i pro **nepřímé identifikátory**. V tomto případě využíváme proces generalizace, při kterém se údaje shrnují do obecných předem specifikovaných kategorií. Při tomto procesu dochází k nenávratné ztrátě dat, je proto důležité důkladně promyslet, zda je generalizace potřeba. Platí, že čím větší je riziko reidentifikace, tím vyšší by měla být úroveň generalizace aplikovaná na celou datovou sadu.<sup>[66]</sup>

Například u dat dvou jedinců se datum narození může změnit z A: 29. 1. 1948; B: 11. 9. 1948 na A: 01.1948; B: 09.1948, či dokonce A: 1948; B: 1948.

### Pseudonymizovaná data

- Taková data, která nelze vztáhnout ke konkrétní osobě bez použití dodatečných informací.
- Vztahuje se na ně nařízení EU č. 2016/679 o ochraně osobních údajů (GDPR).<sup>[65]</sup>

Metodu pseudonymizace je vhodné používat v případě, že je potřeba si na určité úrovni **udržet spojení mezi informací/daty a konkrétním jedincem**. K tomuto spojení ovšem budou mít přístup jen ti, kteří mají k dispozici tzv. mapovací tabulky s přímými identifikátory a pseudonymy. Pseudonymizaci lze použít například kvůli možnosti opětovného kontaktu účastníka studie i po ukončení jeho účasti v klinické studii, třeba za účelem doplnění informací.<sup>[64]</sup>

Při pseudonymizaci jsou osobní údaje různými způsoby zastírány, například použitím **náhodného kódu** místo jména. Identitu účastníků výzkumu proto zná pouze zpracovatel osobních údajů dle GDPR (většinou omezený počet zaměstnanců instituce, např. ošetřující lékař, který jedince do výzkumu zařadil) a výzkumný tým už pracuje pouze se vzorkem označeným náhodným kódem, a dotýčný je proto pro vědce vždy anonymní osobou. V případě potřeby může být na základě šifrovacího klíče, který je vždy na straně zpracovatele osobních údajů, identita účastníka výzkumu „rozklíčována“ a osobní údaje doplněny (např. o výsledky dalších laboratorních testů, diagnostických vyšetření atd.).

Na rozdíl od plně anonymizovaných dat podléhají pseudonymizovaná data stále **GDPR a zákonu č. 110/2019 Sb., o zpracování osobních údajů**, jelikož je lze přiřadit ke specifické fyzické osobě a je na základě toho nutné počítat s možným omezením sdílení dat.<sup>[65]</sup>

### Jak data anonymizovat?

Data je možné anonymizovat kdykoliv během jejich životního cyklu, od samotného sběru až po dobu těsně předtím, než mají být publikována. Způsob anonymizace a její rozsah je potřeba pevně stanovit a aplikovat na všechna příslušná data.

Data lze anonymizovat různými způsoby:

- lze jednotlivé údaje anonymizovat manuálně
- lze si napsat krátký skript například v jazyce Python, který anonymizaci datové sady provede
- lze využít anonymizační software [AMNESIA](#)
  - open source software spravovaný OpenAIRE
  - online demo verze zvládá soubory až o 5000 řádcích, samotný software zvládá i větší soubory
  - určené pro prosté csv a txt datové sady
- dalším vhodným nástrojem může být např. software [ARX](#)

### Co když data anonymizovat nejde?

Existují případy, kdy data není možné anonymizovat. V takovém případě je třeba:

- řešit a kontrolovat přístupy k datům již během výzkumu (právní oprávnění, záruky při sdílení v kolektivu apod.);
- nesdílet a popsat skutečnost například již v přípravě projektu – DMP;
- patřičně zdůvodnit, proč taková data sdílet nejde;
- sdílet případně metadata bez osobních údajů.

Pro správné nastavení publikování vědeckých dat v otevřeném režimu nebo jejich „FAIRifikaci“ je nutné zejména vyjasnit jednotlivé role různých aktérů v rámci institucionálního prostředí. Jedná se tedy především o manažerské rozhodnutí.

Na úrovni instituce je nutné, aby byla jasná politika, kdo je oprávněn uzavírat licenční a další smlouvy související s publikováním/zpřístupněním vědeckých dat. Také je potřeba určit, jak této osobě pomůže zázemí instituce, resp. nastavit míru součinnosti jedince a lokální podpory pro tuto specifickou činnost.

Je nejpraktičtější, pokud primární osobou určenou k uzavírání publikačních dohod v rámci instituce je ten, kdo je datům nejbližší. Ideálně tedy jejich tvůrce (vědec) nebo vedoucí výzkumné skupiny, který má k dispozici součinnost s právní podporou dané instituce. V případě meziinstitucionální spolupráce je vhodné již na počátku (ideálně smluvně) stanovit režim nakládání s daty.

## R1.2 (Meta)data jsou spojena se svým původem (data provenance)

Data provenance (také data lineage) jsou informace o entitách, aktivitách a lidech zapojených do tvorby dat, které mohou být použity k posouzení kvality, spolehlivosti nebo důvěryhodnosti. Jedná se o součást akademického výzkumu a během let bylo vyvinuto několik modelů, které tuto oblast pokrývají. Ve své podstatě jde o metadata spojená se záznamy, které podrobně popisují původ, změny a informace podporující důvěryhodnost nebo platnost dat. Původ dat je důležitý mimo jiné také pro dohledávání chyb v datech a při reportování.

Zjednodušeně řečeno, původ dat pomáhá odpovědět na otázky jako:

- proč byla data vytvořena;
- jak byla data vytvořena;
- kde byla data vytvořena;
- kdy byla data vytvořena;
- kým byla data vytvořena;
- koho a jak citovat;
- zda data obsahují data od někoho jiného, která byla transformována nebo rozšířena.

Tyto informace by měly být popsány ve strojově čitelném formátu.<sup>[67]</sup>

Nejobecnějším modelem je W3C Provenance Data Model využitelný v mnoha oborech a oblastech. Detailní specifikace jsou k dispozici na oficiálních stránkách W3C skupiny: <https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/>.

V současné době vzniká ISO standard ošetřující práci s provenance informací v biotechnologickém prostředí založený právě na principu W3C Provenance Data Model. První část tohoto ISO standardu vyšla v roce 2023 a je zaměřena na provenance biologického materiálu a přidružených dat: ISO/TS 23494-1:2023 – Biotechnology — Provenance information model for biological material and data — Part 1: Design concepts and general requirements.

### R1.3 (Meta)data splňují standardy vědecké komunity pro daný obor

Existuje několik standardů pro různé typy dat, od obecných popisů datových sad (např. Dublin Core) až po specifické druhy dat. Z toho důvodu je dobré od začátku projektu přemýšlet o jejich správném použití:

- na začátku projektu rozhodnout, jaké jsou databáze a repozitáře, které pro dané typy dat budou využívány;
- rozlišit mezi generickými repozitáři (např. [Zenodo](#)) a oborovými repozitáři (např. [CLARIN](#), [ČSDA](#), [Github](#)) pro určitý typ dat;
- navštívit webové stránky repozitáře – vyhledaného např. prostřednictvím [www.re3data.org](http://www.re3data.org) – a zkontrolovat informace o požadovaných metadatech;
- projít proces odeslání kvůli identifikaci požadavků na metadata;
- mít na paměti, že specifická úložiště pro určitý typ dat obvykle mají validátory pro metadata.

Pokud ještě není jasné, jaké úložiště použít, je vhodné:

- zjistit, jaké je doporučené minimální množství informací, pomocí např.:
  - <https://www.fged.org/projects/miame/>
- použít požadované pro náš typ dat v naší komunitě nebo jiná metadata doporučené v následujících zdrojích:
  - RDA <https://rd-alliance.github.io/metadata-directory/standards/>
  - FAIRSHARING <https://fairsharing.org/> v sekci „Standards“ and „Collections“
  - DCC <https://www.dcc.ac.uk/guidance/standards/metadata/list>

Nemusí se vždy jednat o oficiální standardy, protože řada komunit/oborů svoje standardy teprve tvoří. Komunity tedy mohou mít standardy méně formální, které ale přesto zvyšují „férovost“ publikování (meta)dat způsobem zvyšujícím možnost jejich opětovného využití pro komunitu/obor.

V některých situacích může mít tvůrce dat platné a specifikované důvody, proč se odchýlit od standardní dobré praxe v daném oboru. Popis těchto odchylek by také měl být součástí metadat o příslušné datové sadě.

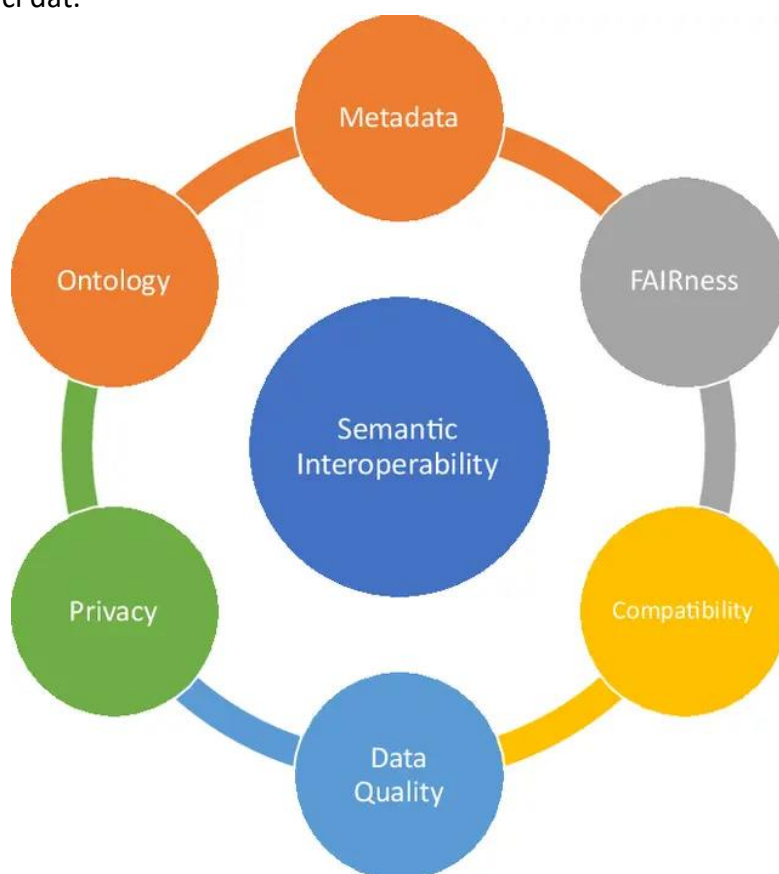
## Sémantická interoperabilita

Klíč k úspěšné výměně dat spočívá ve vzájemném porozumění sdíleným datům. Pod pojmem sémantická interoperabilita je možné si představit stav, kdy se u dvou osob studujících konkrétní artefakt z domény, v níž pracují, spustí stejné kognitivní procesy, tj. že daný artefakt chápou co možno nejpodobněji.

Pokud se bude bavit biolog s hercem o kultuře, jeden může myslet výsledek svého posledního pokusu s bakteriemi a druhý bude přemýšlet nad vývojem divadelní scény. Sémantická interoperabilita by měla zajistit, že oba pochopí, co ten druhý zrovna myslí, když zmíní protokol nebo pěstování kultury.

Sémantická interoperabilita pomáhá popsat postupy a zvyklosti při nakládání s daty ve specifických oborech nebo komunitách. Definování takového postupu může výrazně zefektivnit celý proces sběru, vyhodnocení a publikování dat, a například zprehlednit nebo definovat použití workflow či nástrojů, které efektivně využijí nebo dodržují FAIR principy.

Zvláště důležité je, aby sémantická interoperabilita byla definována pro mezioborovou spolupráci, kdy může dojít k nedorozumění při používání specifických termínů nebo postupů. Obecně ale může efektivně přispět k přehlednosti a automatizaci postupů správy dat nebo i integraci dat z různých přístrojů/nástrojů při sběru dat. Prakticky pomůže zajistit kompatibilitu např. při agregaci dat.

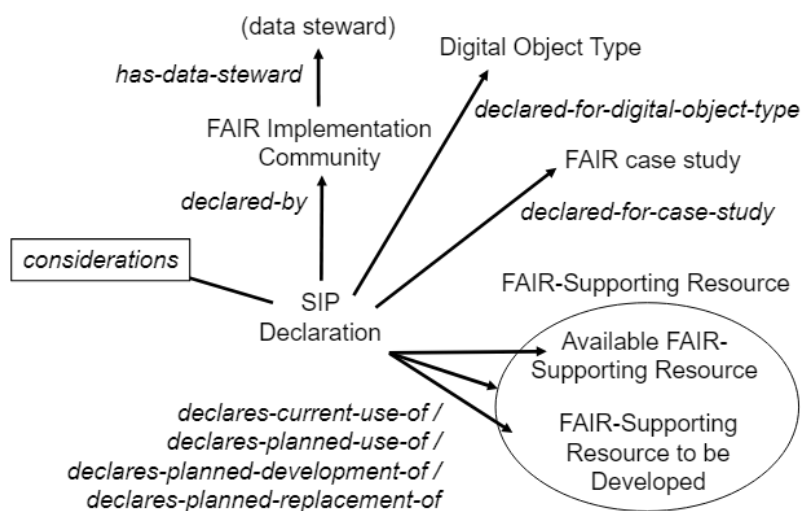


**Obrázek č. 4** Sémantická interoperabilita v datových prostorech je komplexní problém zahrnující více aspektů.<sup>[68]</sup>

Existuje nástroj SIP Wizard (<https://sip-wizard.ds-wizard.org/wizard/>) vyvinutý ve spolupráci s EOSC Semantic Interoperability Task Force (podrobnosti a návod: [https://osf.io/fn2wj/?view\\_only=](https://osf.io/fn2wj/?view_only=)) používající uživatelské rozhraní a prostředí Data Stewardship Wizardu (<https://ds-wizard.org/>).

V nástroji lze definovat profil sémantické interoperability (semantic interoperability profile, SIP, obrázek č. 6), což je seznam deklarovaných možností implementace se zaměřením na aspekty interoperability hlavních zásad FAIR. Zahrnuje sémantické artefakty a jejich podpůrné služby vybrané komunitou pro konkrétní případovou studii a typ dat.

Průvodce zachycuje SIP pomocí dotazníku, který vyžaduje poskytnutí odpovědí, jež explicitně profilují přístup k sémantické interoperabilitě dané komunity. SIP jsou publikovány SIP Wizard jako FAIR (strojově čitelné) a open data (nanopublikace<sup>2</sup>), **které pak mohou sloužit jako reference pro praktické činnosti FAIR správy dat prováděné členy této komunity**. Publikování SIP také podporuje opětovné použití a přeměnu SIP jinými komunitami, **což šetří čas při „znovuobjevování kola“** a současně podporuje sjednocování v možnostech implementace FAIR SIPa použité FAIR SupportingResources lze vyhledávat ve FAIR CONNECT.<sup>[69]</sup>



**Obrázek č. 5** Specifikace semantic interoperability profile (SIP).<sup>[70]</sup>

<sup>2</sup> Nanopublikace jsou vyjádřeny formou znalostního grafu s metadaty, která jsou formální a strojově interpretovatelná. Vzhledem k tomu, že nanopublikace lze citovat, poskytují výzkumným pracovníkům pobídky, aby svá data zpřístupnili ve standardních formátech podporujících dostupnost dat a interoperabilitu.

## Strojová funkčnost

Pro pochopení, co to je strojová funkčnost (machine actionability), je nejdříve potřeba definovat strojově čitelná (machine readable) data, které lze rozdělit do dvou skupin:

- Data čitelná člověkem, která jsou označena tak, aby je mohly číst i stroje (např. mikroformáty, RDFa, HTML).
- Formáty datových souborů určené především pro strojové zpracování (CSV, RDF, XML, JSON). Tyto formáty jsou strojově čitelné pouze v případě, že data v nich obsažená jsou formálně strukturovaná; export souboru CSV ze špatně strukturované tabulky definici nesplňuje.

Strojová funkčnost se týká informací, které jsou konzistentně strukturované tak, aby bylo podle takové struktury možné nastavit nebo naprogramovat automatické zpracování nebo vytěžení dat. V poslední době je zdůrazňována v oblasti plánů správy dat, metadat a analýzy dat, protože nástroje je musí efektivně zpracovávat a vyhodnocovat.

Strojově funkční/čitelná data musí být strukturovaná.

Řada výzkumných ústavů již zavedla nebo se chystá zavést používání elektronického laboratorního deníku (ELN), systému pro správu laboratorních informací (LIMS) nebo podobných systémů pro správu vzorků, činidel, metadat a dat během výzkumného projektu. Tento software by totiž mohl strukturovaně uspořádat informace a umožnit „lepší“ strojové zpracování (meta)dat ve srovnání s tradičními laboratorními knihami nebo jednotlivými složkami a soubory v počítači. Použití strojově zpracovatelných (meta)dat umožňuje škálovatelná řešení, která lze aplikovat v průběhu trvání projektu, což zvyšuje efektivitu a zajišťuje, že poznatky a příspěvky zůstanou v rámci grantu či výzkumné skupiny relevantní.

Stejně tak poskytovatelé financí a instituce žádají výzkumné pracovníky, aby svá (meta)data zpřístupňovali v souladu s FAIR principy a strojově zpracovatelným způsobem. To znamená, že (meta)data by měla být v databázích, v nichž mohou být vystavena takovým způsobem, aby je vyhledávače a sklízecí servery (data harvesting) mohly objevit, indexovat a propojit s dalšími relevantními kontextovými informacemi, čímž se výrazně zvýší pravděpodobnost opětovného využití dat.

Výzkumní pracovníci mají značný prospěch ze strukturování metadat a dat podle zavedených standardů v databázích, což usnadňuje zjednodušené vyhledávání, filtrování a reprodukovatelnost experimentů v různých parametrech a experimentálních podmínkách. Tento přístup umožňuje snadnější integraci datových souborů, automatizaci manipulačních úloh pomocí běžných softwarových nástrojů, jako jsou R a OpenRefine, a využití vizualizačních a průzkumných nástrojů. Kromě toho lze díky systému provádět bezproblémový import, export a výměnu metadat mezi platformami (při splnění sémantické interoperability v komunitě/oboru). Strojově zpracovatelná metadata navíc zlepšují vyhledatelnost referenčních dat a existujících datových sad prostřednictvím vyhledávačů a specializovaných datových katalogů a portálů, čímž zvyšují efektivitu výzkumu a spolupráce.

Mimo oborově nebo komunitně specifické nástroje, které umí pomocí metadat strukturovat, organizovat nebo prohledávat, jsou k dispozici některé obecně využitelné nástroje.

### **Machine-Actionable DMP (maDMP)**

Nástroj Data Stewardship Wizard je přímo vyvinutý s ideou používat strojovou funkčnost pro tvorbu Data Management Planu. Ať už se jedná o specifický výběr, zvýraznění důležitých otázek pro specifickou fázi projektu, nebo o exportování vyplněného znalostního modelu do „papírové“ pro člověka čitelné verze DMP. Funkcionalita celého maDMP se do budoucna bude rozšiřovat skrze API a plánuje se přímé napojení jak na SIP Wizard, tak další externí služby (CESNET). To, co ale může využít každý uživatel, je například automatické vyhodnocení FAIR metrik, dobré praxe tvorby DMP a otevřenosti. Tyto metriky jsou pak zobrazeny v podobě grafu a uživatelé poskytují důležitou formu zpětné vazby.

Pro data stewarda, který DMP kontroluje/připomínkuje, graf metrik poskytuje přehled o konzistenci vyplnění nebo přehled o tom, zda jsou metriky v souladu se záměrem grantu a nastavených interních pravidel pro správu a publikování dat.

### **Machine-Actionable FAIR (maFAIR)**

Dalším užitečným nástrojem pro využití strojové funkčnosti je automatické hodnocení „férovosti“ prostřednictvím perzistentního identifikátoru pro data. Jedná se o nástroj F-UJI <https://www.f-ujl.net/>, což je webová služba pro programové hodnocení „férovosti“ výzkumných datových sad na základě metrik vyvinutých v rámci projektu FAIRsFAIR. Pomocí webového formuláře lze zadat identifikátor (např. DOI, URL) datové sady k posouzení. Volitelně lze zadat také URI koncového bodu metadatové služby (OAI-PMH, SPARQL, CSW), který může F-UJI použít k identifikaci dalších informací.

Nástroj je zvláště užitečný při prohledávání publikovaných dat, která chceme znovu použít. Nejvyšší využitelnost ovšem bude mít pro datové sady v oborově specifických repozitářích, kde se jednodušeji dají najít data vyhovující vědecké komunitě/oboru.<sup>[71]</sup>

## Assessment Results:

### Evaluated Resource:

Úvod do Jupyter Book

✓ Save

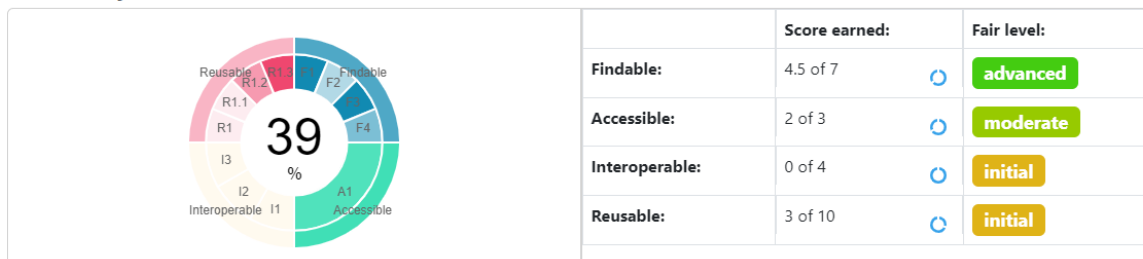
↓ {JSON}

📄 New

FAIR level: ? **moderate**

|                                    |                                                                                                 |
|------------------------------------|-------------------------------------------------------------------------------------------------|
| Resource PID/URL:                  | <a href="https://hdl.handle.net/20.500.14391/3024">https://hdl.handle.net/20.500.14391/3024</a> |
| DataCite support:                  | enabled                                                                                         |
| Metric Version:                    | metrics_v0.5                                                                                    |
| Metric Specification:              | <a href="https://doi.org/10.5281/zenodo.6461229">https://doi.org/10.5281/zenodo.6461229</a>     |
| Software version:                  | 3.1.0                                                                                           |
| Download assessment results:       | <a href="#">{JSON}</a>                                                                          |
| Save and share assessment results: |                                                                                                 |

### Summary:



**Obrázek č. 6** Výsledek hodnocení datové sady prostřednictvím nástroje F-UJI.

Zdroj: <https://www.f-ujl.net/>

## Užitečné odkazy

|                                                       |                                                                                                                                     |
|-------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">UNESCO Recommendation on Open Science</a> | UNESCO doporučení k open science                                                                                                    |
| <a href="#">130/2002 Sb.</a>                          | zákon o podpoře výzkumu a vývoje z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje) |
| <a href="#">A FAIRy tale</a>                          | průvodce zásadami FAIR pro výzkumná data                                                                                            |
| <a href="#">AMNESIA</a>                               | nástroj pro anonymizaci dat                                                                                                         |
| <a href="#">ARGOS</a>                                 | DMP nástroj                                                                                                                         |
| <a href="#">COAR</a>                                  | vyhledávač repozitářů                                                                                                               |
| <a href="#">CoreTrustSeal</a>                         | certifikace důvěryhodných repozitářů                                                                                                |
| <a href="#">Creative Commons</a>                      | otevřené licence pro výzkumné objekty                                                                                               |
| <a href="#">DCC DMP checklist</a>                     | DMP checklist                                                                                                                       |
| <a href="#">DMP online</a>                            | DMP nástroj                                                                                                                         |
| <a href="#">DMP Tool</a>                              | DMP nástroj                                                                                                                         |
| <a href="#">DSW</a>                                   | DMP nástroj                                                                                                                         |
| <a href="#">ELIXIR</a>                                | online návod                                                                                                                        |
| <a href="#">EOSC</a>                                  | European Open Science Cloud                                                                                                         |
| <a href="#">EOSC-CZ</a>                               | česká iniciativa EOSC                                                                                                               |
| <a href="#">EU Open Science</a>                       | nařízení a zdroje Evropské komise                                                                                                   |
| <a href="#">FAIR Cookbook</a>                         | online návod                                                                                                                        |
| <a href="#">FAIR Data Self Assessment Tool</a>        | nástroj na hodnocení míry FAIR výzkumných dat                                                                                       |
| <a href="#">FAIR DataSet Maturity</a>                 | nástroj na hodnocení míry FAIR výzkumných dat                                                                                       |
| <a href="#">FAIR checker</a>                          | nástroj na hodnocení míry FAIR výzkumných dat                                                                                       |
| <a href="#">FAIRAware</a>                             | rozhodovací nástroj pro FAIR data                                                                                                   |
| <a href="#">FAIRification framework</a>               | online návod                                                                                                                        |
| <a href="#">FAIRification Process</a>                 | online návod                                                                                                                        |
| <a href="#">FAIRplus</a>                              | online návod                                                                                                                        |
| <a href="#">FAIRsFAIR</a>                             | praktická řešení pro používání zásad FAIR                                                                                           |
| <a href="#">FAIRsharing</a>                           | registr terminologických artefaktů, modelů/formátů, pokynů pro vykazování a schémat identifikátorů                                  |
| <a href="#">FAIRToolkit</a>                           | online návod                                                                                                                        |
| <a href="#">FOSTER</a>                                | školicí materiály k Open Science                                                                                                    |
| <a href="#">F-UJI</a>                                 | nástroj na hodnocení míry FAIR výzkumných dat                                                                                       |
| <a href="#">GDPR decision tree</a>                    | GDPR rozhodovací strom                                                                                                              |
| <a href="#">GO FAIR</a>                               | komunita pracující na implementaci hlavních zásad FAIR                                                                              |
| <a href="#">HOW TO FAIR</a>                           | uplatňování zásad FAIR v praxi, pomoc s vytvořením plánu správy dat a šířením výsledků výzkumného projektu                          |
| <a href="#">Choose an open source license</a>         | nástroj pro výběr licence                                                                                                           |
| <a href="#">IDENTIFIKATORY.CZ</a>                     | české stránky o perzistentních identifikátorech                                                                                     |

|                                                                      |                                                                               |
|----------------------------------------------------------------------|-------------------------------------------------------------------------------|
| <a href="#">Národní politika výzkumu, vývoje a inovací 2021+</a>     | Národní politika výzkumu, vývoje a inovací České republiky 2021+              |
| <a href="#">OA checker</a>                                           | nástroj na kontrolu Open Access časopisů                                      |
| <a href="#">OpenAIRE</a>                                             | infrastruktura pro otevřený přístup k výzkumu (nástroje a služby)             |
| <a href="#">OpenDOAR</a>                                             | adresář Open Access repozitářů                                                |
| <a href="#">RDM/data steward training</a>                            | vzdělávací portál                                                             |
| <a href="#">Re3data</a>                                              | vyhledávač repozitářů                                                         |
| <a href="#">Science Europe guidance document</a>                     | pokyny DMP pro jednotlivé obory                                               |
| <a href="#">Sherpa Romeo</a>                                         | vyhledávač Open Access časopisů                                               |
| <a href="#">Směrnice Evropského parlamentu a Rady (EU) 2019/1024</a> | Směrnice Evropského parlamentu a Rady (EU) 2019/1024                          |
| <a href="#">Turning FAIR into reality</a>                            | závěrečná zpráva a akční plán skupiny odborníků Evropské komise pro FAIR data |

## Použité zdroje

- [1] WILKINSON, Mark D., Michel DUMONTIER, IJsbrand Jan AALBERSBERG et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* [online]. 2016, **3**(1), 160018. ISSN 2052-4463. Dostupné z: doi:10.1038/sdata.2016.18
- [2] FRIESIKE, Sascha a Thomas SCHILDHAUER. Open Science: Many Good Resolutions, Very Few Incentives, Yet. In: Isabell M. WELPE, Jutta WOLLERSHEIM, Stefanie RINGELHAN a Margit OSTERLOH, ed. *Incentives and Performance: Governance of Research Organizations* [online]. Cham: Springer International Publishing, 2015 [vid. 2024-06-27], s. 277–289. ISBN 978-3-319-09785-5. Dostupné z: doi:10.1007/978-3-319-09785-5\_17
- [3] Open Science. *European Commission* [online]. 3. listopad 2024 [vid. 2024-11-03]. Dostupné z: [https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en)
- [4] Open innovation, open science, open to the world. Shaping Europe's digital future. *European Commission* [online]. 16. červen 2016 [vid. 2024-06-27]. Dostupné z: <https://digital-strategy.ec.europa.eu/en/library/open-innovation-open-science-open-world>
- [5] HNÁTKOVÁ, Eva, Eva DIBUSZOVÁ a Martin SVOBODA. *Otevřená věda: Analýza mezinárodního prostředí* [online]. 2022 [vid. 2024-06-27]. Dostupné z: <https://hdl.handle.net/20.500.14391/2894>
- [6] PONTIKA, Nancy, Petr KNOTH, Matteo CANCELLIERI a Samuel PEARCE. Fostering open science to research using a taxonomy and an eLearning portal. In: *i-KNOW '15: 15th International Conference on Knowledge Technologies and Data-Driven Business: Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business* [online]. Graz: ACM, 2015, s. 1–8. ISBN 978-1-4503-3721-2. Dostupné z: doi:10.1145/2809563.2809571
- [7] MENDEZ, Eva, Rebecca LAWRENCE, Catriona J. MACCALLUM, Eva Moar et al. *Progress on open science: towards a shared research knowledge system: final report of the open science policy platform* [online]. Luxembourg: Publications Office of the European Union, 2020 [vid. 2025-02-07]. ISBN 978-92-76-18882-7. Dostupné z: <https://data.europa.eu/doi/10.2777/00139>
- [8] DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION (EUROPEAN COMMISSION). *OSPP-REC: Open Science Policy Platform Recommendations* [online]. Luxembourg: Publications Office of the European Union, 2018 [vid. 2024-06-27]. ISBN 978-92-79-88333-0. Dostupné z: <https://doi.org/10.2777/958647>

- [9] BERTRAM, Michael G., Josefin SUNDIN, Dominique G. ROCHE, Alfredo SÁNCHEZ-TÓJAR, Eli S. J. THORÉ a Tomas BRODIN. Open science. *Current Biology* [online]. 2023, **33**(15), R792–R797. ISSN 09609822. Dostupné z: doi:10.1016/j.cub.2023.05.036
- [10] *Open Knowledge. Open Definition* [online]. [vid. 2024-08-24]. Dostupné z: <https://opendefinition.org/>
- [11] *Doporučení Komise (EU) 2018/790 ze dne 25. dubna 2018 o přístupu k vědeckým informacím a jejich uchovávání* [online]. [vid. 2024-08-24]. Dostupné z: <https://eur-lex.europa.eu/eli/reco/2018/790/oj/eng>
- [12] *How To FAIR* [online]. [vid. 2024-06-28]. Dostupné z: <https://howtofair.dk/>
- [13] ENGELHARDT, Claudia, Katarzyna BIERNACKA, Aoife COFFEY et al. D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions. *Zenodo* [online]. 21. prosince 2021 [vid. 2024-06-27]. Dostupné z: doi:10.5281/zenodo.5905866
- [14] MONS, Barend, Cameron NEYLON, Jan VELTEROP et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* [online]. 2017, **37**(1), 49–56. ISSN 0167-5265. Dostupné z: doi:10.3233/ISU-170824
- [15] HIGMAN, Rosie, Daniel BANGERT a Sarah JONES. Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights* [online]. 2019, **32**(1) [vid. 2024-06-27]. ISSN 2048-7754. Dostupné z: doi:10.1629/uksg.468
- [16] HERINGOVÁ, Hana a Petra ČERNOHLÁVKOVÁ. *Úvod do problematiky perzistentních identifikátorů* [online]. 22. březen 2024 [vid. 2024-07-17]. Dostupné z: <https://hdl.handle.net/20.500.14391/3002>
- [17] FAIR Principles. *GO FAIR* [online]. [vid. 2024-07-19]. Dostupné z: <https://www.go-fair.org/fair-principles/>
- [18] FENNER, Martin a ARYANI, AMIR. Introducing the PID Graph. *DataCite* [online]. 28. březen 2019 [vid. 2024-07-30]. Dostupné z: <https://doi.org/10.5438/jwvf-8a66>
- [19] MIERZ, Sandra. Project TAPIR: Harvesting the power of PIDs. *TIB-Blog* [online]. 1. březen 2022 [vid. 2024-07-19]. Dostupné z: <https://blog.tib.eu/2022/03/01/project-tapir-harvesting-the-power-of-pids/>
- [20] ČERNOHLÁVKOVÁ, Petra. Národní podpora implementace perzistentních identifikátorů [online]. 2022 [vid. 2024-07-19]. Dostupné z: <https://doi.org/10.48813/k2xs-y923>

- [21] Persistent identifiers. *Digital Preservation Handbook* [online]. [vid. 2024-06-28]. Dostupné z: <https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>
- [22] VALE, Patrick. Creating a landing page. *Crossref* [online]. [vid. 2024-06-28]. Dostupné z: <https://www.crossref.org/documentation/member-setup/creating-a-landing-page/>
- [23] JAKUBOVÁ, Veronika. FTP, SFTP, SMB a další protokoly pro přenos souborů: který vybrat? *MasterDC* [online]. [vid. 2024-06-28]. Dostupné z: <https://www.master.cz/blog/ftp-sftp-smb-protokoly-pro-prenos-souboru-ktery-vybrat/>
- [24] *What is SAML vs OAuth? Find out what's different* [online]. [vid. 2024-06-28]. Dostupné z: <https://auth0.com/intro-to-iam/saml-vs-oauth>
- [25] A1: (Meta)data are retrievable by their identifier using a standardised communication protocol. *GO FAIR* [online]. [vid. 2024-06-28]. Dostupné z: <https://www.go-fair.org/fair-principles/metadata-retrievable-identifier-standardised-communication-protocol/>
- [26] *Zenodo – Research. Shared* [online]. [vid. 2025-02-07]. Dostupné z: <https://about.zenodo.org/policies/>
- [27] DataverseNO Preservation Policy. *info: DataverseNO* [online]. [vid. 2025-02-07]. Dostupné z: <https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/>
- [28] *Interoperability* [online]. [vid. 2024-05-15]. Dostupné z: <https://fairplus.github.io/the-fair-cookbook/content/recipes/interoperability.html>
- [29] HANZLÍKOVÁ, Dagmar. Jak FAIR jsou vaše výzkumná data? *Zenodo* [online]. 3. duben 2020 [vid. 2024-05-15]. Dostupné z: <http://doi.org/10.5281/zenodo.3739188>
- [30] Standardy pro metadata. *Národní digitální knihovna* [online]. [vid. 2024-05-15]. Dostupné z: <https://standardy.ndk.cz/ndk/standardy-digitalizace/metadata>
- [31] KRANTZ, Peter. A lightweight semantic interoperability framework for countries and large organizations (and small ones). *Peter Krantz* [online]. 15. prosinec 2010 [vid. 2024-05-15]. Dostupné z: <https://www.peterkrantz.com/2010/a-lightweight-semantic-interoperability-framework-for-countries-and-large-organizations-and-small-ones/>
- [32] DIRECTORATE-GENERAL FOR DIGITAL SERVICES (EUROPEAN COMMISSION). *New European interoperability framework: promoting seamless services and data flows for European public administrations* [online]. Luxembourg: Publications Office of the European Union, 2017 [vid. 2025-02-07]. ISBN 978-92-79-63756-8. Dostupné z: <https://data.europa.eu/doi/10.2799/78681>

- [33] MARÍN-ARRAIZA, Paloma. Interoperability and data reuse Module 3: FAIR Research Data in the Life Cycle. University of Vienna. 2023.
- [34] Semantic Interoperability. Joinup. *European Commission* [online]. [vid. 2024-05-15]. Dostupné z: <https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/glossary/term/semantic-interoperability>
- [35] EOSC EXECUTIVE BOARD, GENERÁLNÍ ŘEDITELSTVÍ PRO VÝZKUM A INOVACE (EVROPSKÁ KOMISE), Oscar CORCHO, Magnus ERIKSSON, Krzysztof KUROWSKI et al. *EOSC interoperability framework: report from the EOSC Executive Board Working Groups FAIR and Architecture* [online]. Luxembourg: Publications Office of the European Union, 2021 [vid. 2025-02-07]. ISBN 978-92-76-28949-4. Dostupné z: <https://data.europa.eu/doi/10.2777/620649>
- [36] Cross-Domain Interoperability Framework (CDIF). *The WorldFAIR Project* [online]. 8. září 2022 [vid. 2025-02-07]. Dostupné z: <https://worldfair-project.eu/cross-domain-interoperability-framework/>
- [37] CSV, *Comma Separated Values (RFC 4180)* [online]. 7. květen 2024 [vid. 2025-02-07]. Dostupné z: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml>
- [38] ISO/IEC 21778:2017. *ISO* [online]. [vid. 2025-02-07]. Dostupné z: <https://www.iso.org/standard/71616.html>
- [39] OASIS Open Document Format for Office Applications (OpenDocument) TC. *OASIS* [online]. [vid. 2025-02-07]. Dostupné z: <https://groups.oasis-open.org/communities/tc-community-home2?CommunityKey=4bf06d41-79ad-4c58-9e8e-018dc7d05da8>
- [40] I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. *GO FAIR* [online]. [vid. 2024-05-15]. Dostupné z: <https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/>
- [41] Otevřená data. *Architektura eGovernmentu ČR* [online]. [vid. 2024-05-15]. Dostupné z: [https://archi.gov.cz/nap:otevrena\\_data](https://archi.gov.cz/nap:otevrena_data)
- [42] *CSV Schema* [online]. [vid. 2024-05-15]. Dostupné z: <https://digital-preservation.github.io/csv-schema/>
- [43] TENNISON, Jeni. *CSV on the Web: A Primer* [online]. 25. únor 2016 [vid. 2024-05-15]. Dostupné z: <https://www.w3.org/TR/tabular-data-primer/>
- [44] *XML Schema* [online]. [vid. 2024-05-15]. Dostupné z: [https://www.w3schools.com/xml/xml\\_schema.asp#:~:text=An%20XML%20Schema%20describes%20the,Formed%22%20and%20%22Valid%22](https://www.w3schools.com/xml/xml_schema.asp#:~:text=An%20XML%20Schema%20describes%20the,Formed%22%20and%20%22Valid%22)

- [45] *JSON Schema* [online]. [vid. 2024-08-01]. Dostupné z: <https://json-schema.org/>
- [46] *RDF Schema 1.1* [online]. [vid. 2024-05-15]. Dostupné z: <https://www.w3.org/TR/rdf-schema/>
- [47] *OWL. Semantic Web Standards* [online]. [vid. 2024-08-01]. Dostupné z: <https://www.w3.org/OWL/>
- [48] *Shapes Constraint Language (SHACL)* [online]. 20. červenec 2017 [vid. 2024-05-15]. Dostupné z: <https://www.w3.org/TR/shacl/>
- [49] *FAIRsharing* [online]. [vid. 2024-05-15]. Dostupné z: <https://fairsharing.org/>
- [50] *Linked Open Vocabularies (LOV)* [online]. [vid. 2024-05-15]. Dostupné z: <https://lov.linkeddata.es/dataset/lov>
- [51] *BARTOC.org* [online]. [vid. 2024-05-15]. Dostupné z: <https://bartoc.org/>
- [52] AVANÇO, Karla. FAIR Principles and Linked Open Data. *The road to FAIR* [online]. 30. červenec 2021 [vid. 2022-04-28]. Dostupné z: doi:10.58079/trrt
- [53] BLANEY, Jonathan. Introduction to the Principles of Linked Open Data. *Programming Historian* [online]. 7. květen 2017 [vid. 2024-05-15]. Dostupné z: <https://programminghistorian.org/en/lessons/intro-to-linked-data>
- [54] *Shvėzdičková otevřená data* [online]. [vid. 2024-08-01]. Dostupné z: <http://5stardata.info/cs/>
- [55] BERNERS-LEE, Tim. *Linked Data – Design Issues* [online]. 27. červenec 2006 [vid. 2024-05-15]. Dostupné z: <https://www.w3.org/DesignIssues/LinkedData.html>
- [56] Stanovení podmínek užití otevřených dat. *Otevřená data* [online]. [vid. 2022-04-28]. Dostupné z: <https://opendata.gov.cz/cinnost:stanoveni-podminek-uziti>
- [57] Licenses. *Open Source Initiative* [online]. 16. září 2022 [vid. 2024-08-01]. Dostupné z: <https://opensource.org/licenses>
- [58] R1.1: (Meta)data are released with a clear and accessible data usage license. *GO FAIR* [online]. [vid. 2024-06-21]. Dostupné z: <https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/>
- [59] R1.2: (Meta)data are associated with detailed provenance. *GO FAIR* [online]. [vid. 2024-06-21]. Dostupné z: <https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance/>
- [60] R1.3: (Meta)data meet domain-relevant community standards. *GO FAIR* [online]. [vid. 2024-06-21]. Dostupné z: <https://www.go-fair.org/fair-principles/r1-3-metadata-meet-domain-relevant-community-standards/>

- [61] FAIRSHARING TEAM. *FAIRsharing record for: Brain Imaging Data Structure* [online]. FAIRsharing. 2018 [vid. 2024-06-21]. Dostupné z: doi:10.25504/FAIRSHARING.RD1J6T
- [62] *Metadata review guidelines* [online]. [vid. 2024-08-01]. Dostupné z: [https://data.4tu.nl/s/documents/Metadata\\_review\\_guidelines\\_June\\_2021.pdf](https://data.4tu.nl/s/documents/Metadata_review_guidelines_June_2021.pdf)
- [63] *Data licenses* [online]. [vid. 2024-08-01]. Dostupné z: [https://fairplus.github.io/the-fair-cookbook/content/recipes/reusability/ATI\\_licensing\\_data.html](https://fairplus.github.io/the-fair-cookbook/content/recipes/reusability/ATI_licensing_data.html)
- [64] EL EMAM, Khaled a Luk ARBUCKLE. *Anonymizing Health Data*. Sebastopol: O'Reilly Media, 2013. ISBN 978-1-4493-6307-9.
- [65] *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*. 2016.
- [66] EL EMAM, Khaled. *Guide to the de-identification of personal health information*. Boca Raton: Taylor & Francis, 2013. ISBN 978-1-4665-7906-4.
- [67] FAIR Data Principles. *NFDI4Chem Knowledge Base* [online]. [vid. 2024-06-21]. Dostupné z: [https://knowledgebase.nfdi4chem.de/knowledge\\_base/docs/fair/](https://knowledgebase.nfdi4chem.de/knowledge_base/docs/fair/)
- [68] BOUKHERS, Zeyd, Christoph LANGE a Oya BEYAN. Enhancing Data Space Semantic Interoperability through Machine Learning: a Visionary Perspective. *arXiv* [online]. 15. březen 2023 [vid. 2024-05-15]. Dostupné z: doi:10.48550/arXiv.2303.08932.
- [69] GONZALEZ, Esteban a Anne-Sofie FINK. *Synchronisation Workshop 2023*.
- [70] MAGAGNA, Barbara, Erik SCHULTES a Tobias KUHN. *FAIR Implementation Profile (FIP) Ontology* [online]. [vid. 2024-05-15]. Dostupné z: <https://peta-pico.github.io/FAIR-nanopubs/fip/index-en.html>
- [71] Machine-Actionability. *Data Stewardship Wizard* [online]. [vid. 2024-05-15]. Dostupné z: <https://ds-wizard.org/machine-actionability>
- [72] *Open data, FAIR data and RDM: the ugly duckling* [online]. [vid. 2025-03-14]. Dostupné z: <https://www.uhasselt.be/en/university-library/research/research-data-management/fair>