



Analýza mezinárodních metadatových schémat pro popis výzkumných dat

Verze 1.0
19. 12. 2023



**Spolufinancováno
Evropskou unií**

Zpracovali:

Hana Vyčítalová (<https://orcid.org/0000-0002-8323-5790>)

Petra Černošlávková (<https://orcid.org/0000-0001-8683-6156>)

Jakub Klímek (<https://orcid.org/0000-0001-7234-3051>)

Analýza byla revidována pracovní skupinou EOSC CZ Metadata, jejíž tajemnicí je Michaela Capandová (<https://orcid.org/0000-0001-7606-4084>).

DOI: <https://doi.org/10.48813/6kix-cr80>

Praha, 2023



Užití tohoto díla se řídí mezinárodní licencí Creative Commons Attribution License 4.0 která umožňuje neomezené využití, distribuci a kopírování díla pomocí jakéhokoliv média, za podmínky řádného uvedení původních autorů a zdroje.

Obsah

1 Úvod	4
2 Mapování metadatových schémat pro obecný popis výzkumných dat.....	6
2.1 Přehled univerzálních metadatových schémat.....	6
2.2 Nejčastěji uváděná schémata v systému re3data.org	9
2.3 Schémata vybraná k další analýze	11
2.3.1 Kontext otevřených dat	11
3 Popis vybraných univerzálních schémat	13
3.1 Dublin Core Metadata Initiative (DCMI) Metadata Terms.....	13
3.2 DataCite Metadata Schema.....	14
3.3 DCAT (Data Catalog Vocabulary) a DCAT-AP	15
3.4 Data Documentation Initiative (DDI)	16
3.5 Mapování mezi analyzovanými schématy.....	17
4 Závěr	18
Příloha č. 1 – Tabulkový přehled vybraných schémat	20
Použité zdroje	21

1 Úvod

V mezinárodním prostředí existuje velké množství metadatových schémat sloužících pro popis výzkumných dat. Některá jsou určena výhradně pro výzkumná data, jiná mohou být použita pro popis různorodých digitálních objektů. Dají se rozlišit schémata univerzální, použitelná obecně pro popis digitálních objektů, a schémata oborová sloužící jednomu případně skupině příbuzných oborů. Zrovna tak se dají oddělit schémata, hodící se spíše pro popis textových dokumentů, schémata vhodná a užívaná pro popis výzkumných dat nebo schémata určená třeba pro popis softwarů.

Tato analýza se zabývá univerzálními metadatovými schématy pro popis výzkumných dat užívanými v mezinárodním prostředí a jejím cílem je definovat užší výběr schémat, která budou následně posuzována za účelem stanovení základního metadatového modelu pro popis výzkumných dat, který zároveň poslouží jako výměnný formát¹ mezi repozitáři v nově vznikajícím českém prostředí Národní datové infrastruktury/Národní repozitářové platformy a Národním metadatovým adresářem a dalšími (mezinárodními) agregátory. Základní model by měl stanovit minimální popisné prvky pro datové sady dostatečně obecné, aby byl model použitelný ve všech oborech a zajistil interoperabilitu základních metadat napříč obory. Základní model bude sloužit pro jednotné vyhledávání nad metadaty výzkumných dat v Národním metadatovém adresáři. Na základní model pak budou navazovat nejrůznější oborová rozšíření, která by měla zjistit dostatečně detailní popis výzkumných dat a přispět tak k znovuvyužití dat. Vybrané univerzální metadatové schéma nebo schémata mohou být použita buď v zavedené standardní podobě nebo s drobnými úpravami či doplněnými prvky dle potřeb české komunity VaVal.

Při dalším posuzování vybraných metadatových schémat je třeba vzít v úvahu tyto faktory:

- potřeby české komunity VaVal,
- současné zavedené standardy a praxi při správě výzkumných dat (zejména jejich ukládání v repozitářích),
- cíl a funkce Národního metadatového adresáře,
- přidělování identifikátoru DOI,
- potřeby poskytovatelů prostředků na výzkum, vývoj a inovace,
- komunikace a interoperabilita repozitářů s dalšími systémy (zejména Národní katalog otevřených dat – bude dále vysvětleno, IS VaVal, OpenAIRE).

Vytváření přehledu nejrozšířenějších schémat a i následnou analýzu znesnadňuje fakt, že existují rodiny metadatových schémat (např. OpenAIRE Guidelines nebo Data Package), které vycházejí

¹ Standardizovaný **formát** struktury bibliografického, katalogizačního záznamu, autoritního záznamu, klasifikačních údajů, sloužící pro výměnu dat mezi jednotlivými informačními institucemi anebo systémy. Stanoví povinné údaje pro výměnu, pořadí údajů, jejich odpovídající kódová označení a pravidla použití. Zdroj: VODIČKOVÁ, Hana. Výměnný formát. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha : Národní knihovna ČR, 2003- [cit. 2023-08-30]. Dostupné z: https://aleph.nkp.cz/F/?func=direct&doc_number=000001405&local_base=KTD.

ze stejného základu, ale liší se úpravami buď pro geografickou oblast, jazykovou oblast, typ dat nebo repozitářů nebo obor. V souvislosti s tímto problémem se používá také termín aplikační profil, zejména v oblasti otevřených dat (open data). Je ovšem otázka, zda je vhodné posuzovat takovou rodinu schémat/modelů jako celek nebo každou zvlášť. Přestože vycházejí ze stejného základu, mohou se lišit co do rozsáhlosti (počet popisných prvků, hierarchické členění), tak i v použití výměnných formátů nebo rozšíření v určitých geografických oblastech. Při detailnějším analyzování jednotlivých schémat proto posuzujeme spíše jednotlivé aplikační profily/konkrétní úpravu schémat.

2 Mapování metadatových schémat pro obecný popis výzkumných dat

Tato kapitola mapuje existující univerzální či multioborová schémata pro popis výzkumných dat, která jsou zaevidována v mezinárodních zdrojích. Při mapování nejznámějších schémat bylo využito informací a podkladů zejména z těchto zdrojů: The RDA Metadata Standards Catalog (dále také MSC)², web Digital Curation Centre³, systém FAIRsharing⁴ (2.1 Přehled univerzálních metadatových schémat). Následně bylo v systému re3data.org⁵ (Registry of research data repositories) ověřeno, zda se k nalezeným metadatovým schématům hlásí nějaké datové repozitáře. Zároveň systém re3data.org posloužil jako další referenční zdroj o nejužívanějších schématech v mezinárodním prostředí (2.2 Nejčastěji uváděná schémata v systému re3data.org). Poslední část této kapitoly (2.3 Schémata vybraná k další analýze) určuje na základě informací z kapitol 2.1 a 2.2 užší výběr metadatových schémat k bližšímu popisu.

2.1 Přehled univerzálních metadatových schémat

Následující seznam je kompilací univerzálních nebo multioborových metadatových schémat pro popis výzkumných dat uváděných v The RDA Metadata Standards Catalog a na webu Digital Curation Centre. u každého schématu uvádíme jeho krátký popis, odkaz na záznam v MSC (pokud MSC schéma uvádí) a odkaz na dokumentaci. Dále jsou doplněny ještě informace ze systému FAIRsharing (DOI daného schématu).

- **AGLS Metadata Profile** (Australian Government Locator Service; Australian Standard AS 5044-2010) – metadatový standard, založeno na Dublin Core, standardizace dle Australian Standard AS 5044-2010.
Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m44>
DOI (FAIRsharing): 10.25504/FAIRsharing.35db1b.
- Dokumentace: <https://agls.gov.au/>

² *Metadata Standards Catalog* [online]. 2023, [cit. 2023-03-06]. Dostupné z: <https://rdamsc.bath.ac.uk/>. Jedná se o otevřený adresář metadatových standardů, kterými lze popisovat výzkumná data. Do adresáře mohou přispívat zaregistrovaní uživatelé. Katalog byl vytvořen v rámci činnosti Metadata Standards Directory Working Group při RDA (Research Data Alliance), která se stará i o jeho údržbu.

³ General Research Data. *Digital Curation Centre* [online]. 2023 [cit. 2023-03-06]. Dostupné z: <https://www.dcc.ac.uk/resources/subject-areas/general-research-data>. Digital Curation Centre poskytuje konzultační služby a vzdělávací aktivity k problematice správy výzkumných dat, digitálního kurátorství atd.

⁴ FAIRsharing [online]. 2023 [cit. 2023-03-06]. Dostupné z: <https://fairsharing.org/>. FAIRsharing je registr metadatových standardů, samotných repozitářů zpřístupňujících výzkumná data i politik týkajících se managementu, sdílení a dlouhodobého ukládání dat. O registr a podpůrné služby se stará tým odborníků z Data Readiness Group, která je součástí University of Oxford.

⁵ *Re3data.org - Registry of Research Data Repositories* [online]. 2023 [cit. 2023-05-16]. Dostupné z: <https://www.re3data.org/>. DOI: 10.17616/R3D. Re3data je globální registr repozitářů pro výzkumná data z různých akademických oborů. Registr vznikl v rámci projektu podpořeného German Research Foundation. Partnery projektu jsou Berlin School of Library and Information Science, DataCite, Helmholtz Open Science Office působící v rámci GFZ German Research Centre for Geosciences, KIT Library coby součást Karlsruhe Institute of Technology (KIT) a knihovny Purdue University.

- **CERIF** (Common European Research Information Format) – schéma pro popis výstupů z výzkumu (výzkumných aktivit) doporučené Evropskou unií pro členské země EU. Bývá užíván v CRIS systémech pro evidenci výsledků VaVal.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m4>
 DOI (FAIRsharing): 10.25504/FAIRsharing.6945dq
 Dokumentace: <https://eurocris.org/Uploads/Web%20pages/CERIF-1.6/documentation/MInfo.html>.
- **DDI** (Data Documentation Initiative) – schéma pro popis výzkumných dat, zejména z oblasti sociálních, ekonomických věd a věd o chování.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m13>
 - DOI DDI Codebook (FAIRsharing): 10.25504/FAIRsharing.EZCpPd
 Dokumentace DDI Codebook: <https://ddialliance.org/Specification/DDI-Codebook/2.5/>
- **Data Package** – formát pro popis a výměnu kolekcí dat.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m10>
 DOI (FAIRsharing): 10.25504/FAIRsharing.3b861d.
 Dokumentace: <https://specs.frictionlessdata.io/data-package/>.
- **DataCite Metadata Schema** – univerzální schéma základních metadat pro popis jakýchkoliv výzkumných dat, set povinných polí je vyžadován kvůli registraci identifikátorů DOI přidělovaných agenturou DataCite.
 - Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m11>
 DOI DCME (FAIRsharing): 10.25504/FAIRsharing.me4qwe
 Dokumentace DCME: <https://schema.datacite.org/>
- **DCAT – Data Catalog Vocabulary** – slovník v RDF, který umožňuje zajistit správnou výměnu dat mezi katalogy na webu, může být použit pro popis datových sad.
 - Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m12>
 DOI (FAIRsharing): 10.25504/FAIRsharing.h4j3qm
 Dokumentace: <https://www.w3.org/TR/vocab-dcat/>
- **DCAT-AP – DCAT Application profile for data portals in Europe** – aplikační profil DCAT pro popis dat v evropských datových portálech
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m124>
 Dokumentace: <https://op.europa.eu/en/web/eu-vocabularies/dcat-ap>
- **Dryad Metadata Application Profile** – schéma pro generický repozitář Dryad, aplikační profil Dublin Core, sloužící pro popis výzkumných dat dokládajících recenzovanou vědeckou literaturu.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m55>
 Dokumentace: https://github.com/CDL-Dryad/dryad-app/blob/main/documentation/v1_wiki_content.md.
- **OAI-ORE** (Open Archives Initiative Object Reuse and Exchange) – standard pro popis a výměnu digitálních objektů, standard IEEE 1484.13.6-2015.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m26>
 DOI (FAIRsharing): 10.25504/FAIRsharing.3mtaee
 Dokumentace: <http://www.openarchives.org/ore/1.0/primer>
- **DCMI Metadata Terms** – univerzální schéma pro popis různých objektů, ISO Standard 15836 (únor 2009).
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m15>

- DOI DCME (FAIRsharing): 10.25504/FAIRsharing.3nx7t
 Dokumentace: <https://www.dublincore.org/specifications/dublin-core/dces/>
- **Observations and Measurements** – XML implementace pro model OGC (Open Geospatial Consortium) a ISO Observations and Measurements (O&M) pro geoprostorová data, OGC Observations and Measurements v2.0 také publikované jako ISO/DIS 19156, v nově zrevidované verzi se jedná o normu ISO 19156:2023.
 Záznam v MSC: nemá
 DOI (FAIRsharing): 10.25504/FAIRsharing.q7kr4s
 Dokumentace: <https://www.ogc.org/standard/om/>
 - **OpenAIRE Guidelines** – rodina specifikací pro datové archivy (založeno na DataCite), dokumentové repozitáře a pro manažery institucionálních a oborových repozitářů (založeno na Dublin Core), pro CRIS manažery (založeno na CERIF).
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m92>
 DOI OpenAIRE Guidelines for Data Archives (FAIRsharing): 10.25504/FAIRsharing.123197
 Dokumentace OpenAIRE Guidelines for Data Archives: <https://guidelines.openaire.eu/en/latest/data/introduction.html>
 - **PREMIS** (Premis Data Dictionary for the Preservation Metadata) – metadatové schéma zajišťující údaje nutné pro archivaci digitálních objektů.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m31>
 DOI: 10.25504/FAIRsharing.27a0f4.
 Dokumentace: <http://www.loc.gov/standards/premis/>
 - **PROV** – rodina specifikací PROV definuje model umožňující interoperabilní výměnu informací o původu objektů v heterogenních prostředích.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m33>
 DOI PROV-DM (FAIRsharing): 10.25504/FAIRsharing.36c7ea
 Dokumentace PROV-DM: <http://www.w3.org/TR/prov-dm/>
 - **QuDEx** (Qualitative Data Exchange Format) – schéma pro popis a výměnu kolekcí dat.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m34>
 DOI (FAIRsharing): 10.25504/FAIRsharing.d42cb9.
 Dokumentace: <https://www.data-archive.ac.uk/managing-data/standards-and-procedures/metadata-standards/qudex/>
 - **RDF Data Cube Vocabulary** – standard pro popis a publikování statistických dat na webu – popisuje strukturu datových kostek a jejich obsah, kompatibilní s SDMX (Statistical Data and Metadata eXchange), ISO/TS 17369:2005 standard for exchanging and sharing statistical data and metadata among organizations.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m35>
 DOI (FAIRsharing): 10.25504/FAIRsharing.c3b573.
 Dokumentace: <https://www.w3.org/TR/vocab-data-cube/>.
 - **RIF-CS** (Registry Interchange Format – Collections and Services) – schéma pro výměnu dat, založeno na standardu ISO 2146:2010 „Information and documentation – Registry services for libraries and related organisations.
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m98>
 DOI (Fairsharing): 10.25504/FAIRsharing.reybx2
 Dokumentace: <https://documentation.ardc.edu.au/display/DOC/About+RIF-CS>
 - **CSV on the Web** – Doporučení W3C pro popis CSV souborů na webu
 Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m125>
 Dokumentace: <https://www.w3.org/TR/tabular-data-primer/>

- **Tabular Data Package** – profil Data Package pro výměnu dat ve formátu CSV.
Záznam v MSC: <https://rdamsc.bath.ac.uk/msc/m77>
DOI (Fairsharing): 10.25504/FAIRsharing.082881.
Dokumentace: <https://specs.frictionlessdata.io/tabular-data-package/>.

Zdroje:

Metadata Standard Catalogue: <https://rdamsc.bath.ac.uk/subject/Multidisciplinary>;

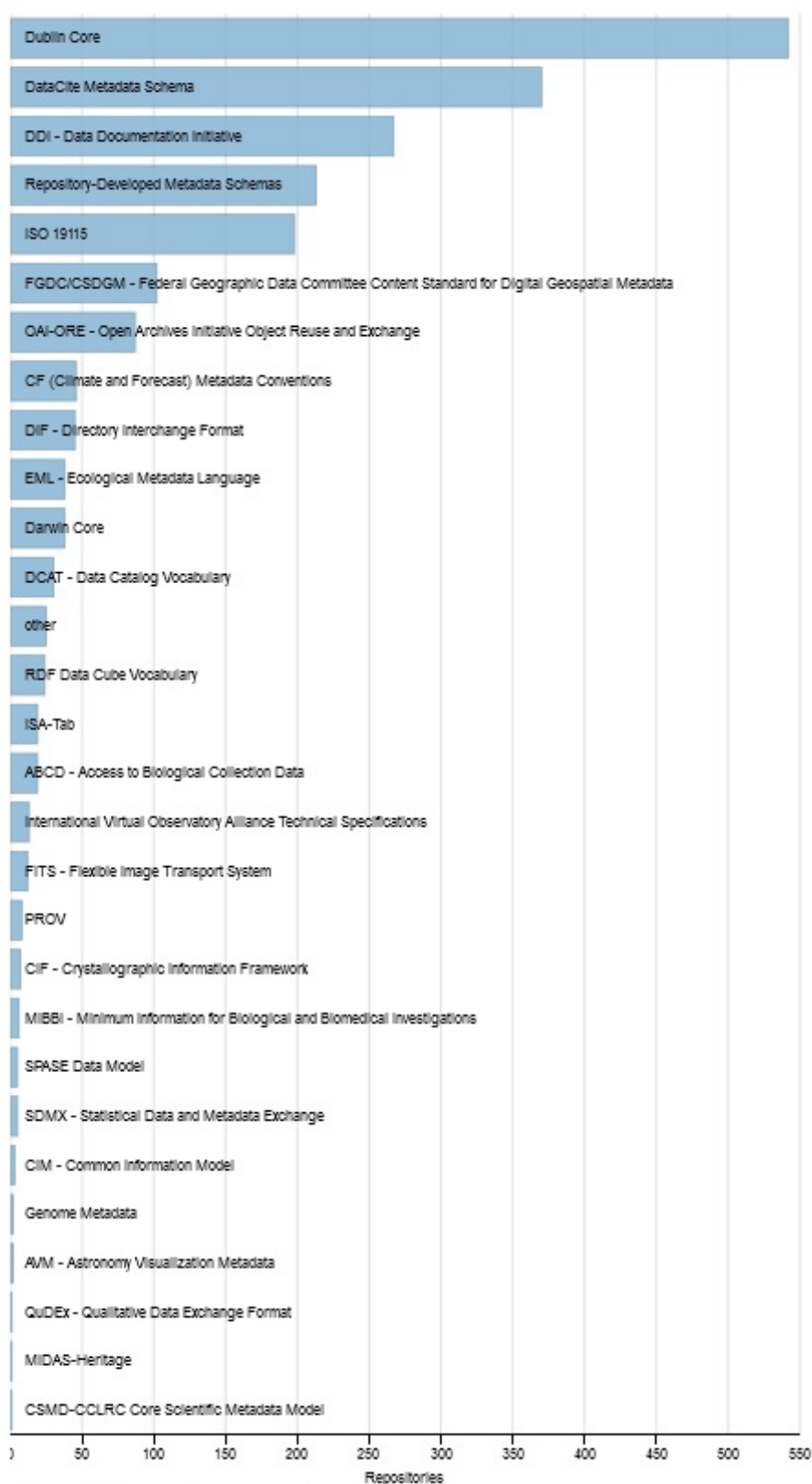
Digital Curation Centre: <https://www.dcc.ac.uk/resources/subject-areas/general-research-data>;

FAIRsharing: <https://fairsharing.org/>

2.2 Nejčastěji uváděná schémata v systému re3data.org

Mezinárodní systém pro evidenci datových repozitářů re3data.org jako jedno z popisných kritérií zaznamenává rovněž použitá metadatová schémata. Tento systém tedy posloužil jako referenční zdroj o nejčastěji užitých schématech v evidovaných datových repozitářích.

Metadata standards



Obr. č. 1 Graf četnosti schémat užívaných v datových repozitářích, stav k 11. 8. 2023, zdroj Re3data.org (<https://www.re3data.org/metrics/metadataStandards>)

Nejčastějším uvedeným schématem je DublinCore (540 záznamů), DataCite Metadata Schema (370 záznamů) je na druhém místě, jako třetí nejčastější schéma se objevuje schéma DDI (267 záznamů). Na čtvrtém místě je skupina Repository-Developed Metadata Schemas, která ale zahrnuje různá schémata, která si provozovatelé repozitářů vytvořili nebo upravili přímo pro potřebu svého repozitáře, ta pro potřeby analýzy nejsou relevantní. Je nutné podotknout, že čísla nemusejí komplexně vystihovat celou oblast repozitářů. Jednak údaj o použitých metadatových schématech není nutné vyplnit, jednak jeden repozitář může využívat, a tedy i uvádět více metadatových schémat. Kromě vnitřního (interního) metadatového schématu, které je často určeno užitím konkrétního softwaru (např. Dublin Core v DSpace), repozitáře často vystavují svá (meta)data přes API či specifitěji přes protokol OAI-PMH v různých metadatových schématech a standardech. Nicméně tento fakt nijak nezpochybnuje statistiku nejvyužívanějších schémat (Re3data.org – Registry of Research Data Repositories, 2023). Četnost schémat byla sledována od konce února do 11. srpna 2023 a je také nutné zmínit, že největší nárůst užívání byl zaznamenán opět u schémat Dublin Core, DataCite (nejvýznamnější změna – tomuto schématu přibylo 30 záznamů), DDI, významnější změna byla i u schématu OAI-ORE (nárůst o 18 záznamů).

2.3 Schémata vybraná k další analýze

Nejčastěji využívaná univerzální metadatová schémata Dublin Core, DataCite a DDI jsou dále podrobena detailnějšímu průzkumu. K těmto třem schématům bylo ještě zařazeno schéma DCAT, které sice nefiguruje mezi nejčastěji užívanými schématy v re3data.org, ale jedná se o schéma užívané obvykle k metadatovému popisu (katalogizaci) otevřených dat ve státní a veřejné správě, jejichž koncepce se bude úzce dotýkat i dat výzkumných.

2.3.1 Kontext otevřených dat

Otevřená data jsou dle Zákona č. 106/1999 Sb., o svobodném přístupu k informacím „... informace zveřejňované způsobem umožňujícím dálkový přístup v otevřeném a strojově čitelném formátu, jejichž způsob ani účel následného využití není povinným subjektem, který je zveřejňuje, omezen a které jsou evidovány v národním katalogu otevřených dat.“ (Česká republika, 2023a, § 3 odst. 5). Začátky otevřených dat byly spojené především se sférou veřejné správy. Jednalo se zejména o otevírání a zviditelnění dat z úředních desek úřadů a základních územních samosprávných celků, databází úřadů poskytujících veřejné informace atd. (Klímek, 2020). Ukazuje se ale, že i data jiného původu (včetně dat z výzkumu podpořeného z veřejných prostředků) budou řazena mezi data, u kterých bude vyžadován otevřený formát. Navíc přístupnost a snadná naležitelnost výzkumných dat co největším počtem různorodých uživatelů (i mimo výzkumnou komunitu) může být v zájmu i samotných autorů.

Směrnice Evropského parlamentu a rady (EU) 2019/1024 ze dne 20. června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru⁶ (dále jen Směrnice EP) výslovně zmiňuje výzkumná data (dle terminologie této zmíněné směrnice „údaje z výzkumu“). Tato

⁶ SMĚRNICE EVROPSKÉHO PARLAMENTU a RADY (EU) 2019/1024 ze dne 20. června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru, více na <https://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:32019L1024&from=CS>.

směrnice apeluje na členské státy, aby i údaje z výzkumu financovaných z veřejných prostředků byly zveřejňovány coby otevřená data, podle FAIR [principů](#)⁷ a aby v každém státě byla nastavena politika otevřeného přístupu. Směrnice uvádí řadu možných výjimek, na které je třeba brát ohled, tj. problematika soukromí, ochrany osobních údajů, důvěrnosti, národní bezpečnosti, veřejné bezpečnosti, oprávněných obchodních zájmů, jako jsou obchodní tajemství, a práv duševního vlastnictví třetích stran atd. Otevřený přístup by měl fungovat v souladu se zásadou „co nejotevřenější, co nejméně uzavřené“ („as open as possible, as closed as necessary“) (Směrnice Evropského parlamentu a rady (EU) 2019/1024 ze dne 20. června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru (přepracované znění), 2019). (European Commission, 2023).

Novelizovaný Zákon č. 130/2002 Sb. z roku 2022⁸ v § 12a (Přístup k výzkumným datům) rovněž nařizuje zpřístupňovat výzkumná data z výzkumu či vývoje, který byl hrazen z veřejných prostředků, v otevřeném a strojově čitelném formátu. Zákon opět umožňuje dříve zmíněné výjimky. Údaje o výzkumných datech bude také nově nutné předávat do Informačního systému výzkumu, vývoje a inovací, což může přinést ještě další požadavky na popis a formát dat. (Česká republika, 2023b).

Z těchto důvodů by se na výzkumná data měla vztahovat stejná pravidla jako na otevřená data, tj. otevřený formát, strojová čitelnost, otevřené podmínky užití a registrace v Národním katalogu otevřených dat, čímž by skutečně naplňovala definici otevřených dat. (Klímek, 2021). Definici a účel Národního katalogu otevřených dat stanovuje Zákon č. 106/1999 Sb., o svobodném přístupu k informacím: „Národní katalog otevřených dat je informační systém veřejné správy přístupný způsobem umožňujícím dálkový přístup sloužící k evidování informací zveřejňovaných jako otevřená data.“ Správcem tohoto katalogu je od 1. dubna 2023 Digitální a informační agentura, která převzala celou problematiku otevřených dat (Česká republika, 2023b). Data mohou být uložena v lokálním katalogu otevřených dat (LKOD), čili v repozitáři, odkud jsou jejich metadata harvestována do Národního katalogu otevřených dat (NKOD). (Klímek, 2021). Lokální katalog/repoziťář ovšem musí zvládat výměnu dat s NKOD podle *Otevřené formální normy⁹ pro Rozhraní katalogů otevřených dat: DCAT-AP-CZ¹⁰*, která obsahuje i definici rozhraní lokálních katalogů pro jejich harvestování. Právě možnost harvestování repozitářů ze strany Národního katalogu otevřených dat je jedním z důvodů pro zařazení formátu DCAT do této analýzy. Dalším důvodem je, že DCAT, specificky DCAT-AP, se jakožto standard pro datové portály v Evropě vydaný Evropskou komisí začíná prosazovat ve všech odvětvích, o které má Evropská komise zájem, včetně EOSC jako jednoho z potenciálních společných evropských datových prostorů (Data Spaces). DCAT-AP pak tvoří součást tzv. blueprintu pro stavbu těchto datových prostorů. Lze tedy očekávat, že implementace DCAT-AP bude po repozitářích, které mají být součástí EOSC, vyžadována, nebo alespoň silně doporučována jako prostředek pro zajištění jejich interoperability s ostatními datovými portály v Evropě. (Data Spaces Support Centre, 2023)

⁷ <https://www.go-fair.org/fair-principles/>

⁸ Zákon č. 130/2002 Sb. Zákon o podpoře výzkumu a vývoje z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje), více na <https://www.zakonyprolidi.cz/cs/2002-130>.

⁹ Otevřené formální normy (OFN) definují technická doporučení pro určité datové sady, aby byla zajištěna interoperabilita pro dané prostředí. (Otevřené formální normy, 2023).

¹⁰ DCAT-AP-CZ je národní aplikační profil DCAT-AP pro Českou republiku.

3 Popis vybraných univerzálních schémat

Detailnější analýze byly podrobeny tato schémata:

Dublin Core Metadata Initiative Metadata Terms

Dokumentace: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

DataCite

Dokumentace: https://schema.datacite.org/meta/kernel-4.4/doc/DataCite-MetadataKernel_v4.4.pdf

DCAT (DCAT-AP, DCAT-AP-CZ)

Dokumentace DCAT, verze 3: <https://www.w3.org/TR/vocab-dcat-3/>

Dokumentace DCAT-AP: <https://semiceu.github.io/DCAT-AP/releases/3.0.0/>

Dokumentace DCAT-AP-CZ: <https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/>

DDI

Dokumentace: https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation.html

3.1 Dublin Core Metadata Initiative (DCMI) Metadata Terms

Schéma DCMI Metadata Terms vzniklo už v roce 1995 jako univerzální schéma pro popis digitálních objektů, zdrojů na webu (DCMI, 2023b). Nejedná se tedy o schéma vytvořené přímo pro popis výzkumných dat, ale je možné jej využít i pro tento účel. Jedná se o jedno z nejrozšířenějších schémat, jeho použití tedy poskytuje jistotu snadného mapování na jiná schémata a zaručuje interoperabilitu.

Schéma vyvíjí Dublin Core Metadata Initiative, organizace podporující inovace v oblasti metadat a sdílení „best practices“. Kromě rozvoje a aktualizace DCMI Metadata Terms tato iniciativa pořádá konference, školení a další odborné akce. Současná verze DCMI je z roku 2020. (DCMI, 2023a).

Schéma je složeno z 15 základních prvků – elementů (Dublin Core Metadata Element Set, známé také pouze jako Dublin Core nebo DC). Schéma tedy není původně vytvořené jako hierarchické, obsahovalo pouze zmíněnou sadu základních prvků. Časem ale bylo třeba přijmout rozšíření (DCMI Metadata Terms), většina prvků (elementů) byla rozšířena o kvalifikátory (např. prvek dc.contributor pro přispěvatele může mít kvalifikátory, které vyjadřují role jednotlivých osob – dc.contributor.author, dc.contributor.editor aj.), které stojí na stejné úrovni a zpřesňují typ nebo obsah jednotlivých elementů. Každý z termínů je opatřen vlastním identifikátorem URI. DCMI je vyjádřený jako RDF slovník. Je možné ho zobrazovat ve formátu XML, JSON nebo UML. Ve schématu DCMI záměrně nejsou určeny povinnosti prvků, ani opakovatelnosti, jelikož definice těchto prvků slouží jako sjednocení základních pojmů používaných při popisu různorodých objektů. Ve schématu je definováno několik řízených slovníků, např. typů dokumentů. (DCMI,

2023c). DCMI je využíván jako schéma pro popisná metadata v softwaru DSpace využívaném pro tvorbu digitálních knihoven a repozitářů. (Krejčíř, 2006).

Více na: <https://www.dublincore.org/>

3.2 DataCite Metadata Schema

DataCite je jednou z registračních agentur, které poskytují perzistentní identifikátory DOI. DataCite se specializuje na přidělování DOI pro datové sady a další výstupy výzkumu a další služby spojené s přidělováním tohoto identifikátoru. (DataCite, 2023b).

DataCite Metadata Schema je univerzální schéma pro popis výzkumných dat nebo dalších výstupů z výzkumu, které vytvořila pro potřeby dobré identifikace zdrojů (co nejpresnějšího vyhledávání a zajištění konzistence metadatového popisu) mezinárodní nezisková organizace DataCite. (DataCite Metadata Working Group, 2021).

Ke každému identifikátoru DOI se váže sada popisných metadat, proto tato organizace vyvinula univerzální schéma DataCite Metadata Schema a stará se o jeho rozvoj a aktualizaci prostřednictvím The Metadata Working Group. Do této skupiny jsou přijímáni jako členové i metadatoví specialisté z různých institucí či zástupci významných stakeholderů v oblasti výzkumných dat. Do rozvoje se může zapojit i širší komunita prostřednictvím návrhů či komentářů nově připravovaných verzí (DataCite, 2023a). Aktuální je momentálně verze 4.4 z roku 2021 (DataCite, 2023a), na podzim roku 2022 byla představena verze 4.5 a byla otevřena pro komentování a revizi. Zatím je k dispozici pouze draft verze 4.5. Na rok 2023 bylo avizováno vydání nové verze 5 (tzv. major version), která by měla zahrnovat významnější změny. (DataCite Metadata Working Group, 2022)

Schéma DataCite je hierarchické, založené na XML a JSON. Skládá se zatím z 20 hlavních prvků (properties), které mohou mít další podřízené prvky (elementy a atributy). Jedná se např. o prvek *Název*, který má vlastnost *typ názvu* (přeložený název, podnázev, alternativní název a další název). Ve schématu je určeno 6 povinných prvků, u všech prvků na všech úrovních je určena opakovatelnost. Ve schématu je definováno také několik seznamů preferovaných hodnot pro různé prvky nebo jejich vlastnosti (tzv. Controlled List Definitions), např. pro typy dokumentů, role přispěvatelů nebo typy datací. (DataCite Metadata Working Group, 2021).

Schéma DataCite je velmi důležité brát v potaz právě kvůli přidělování identifikátoru DOI datovým sadám. Požadavek identifikace výzkumných dat prostřednictvím identifikátoru DOI stále roste, tyto požadavky přicházejí ze strany samotných výzkumníků, kteří vidí výhody spojené se snadnou identifikací dat, poskytovatelů podpory (např. Horizon Europe) i vydavatelů. Na používání globálního unikátního identifikátoru apelují také FAIR principy. Při žádosti o získání identifikátoru DOI je nutné buď ručně nebo automatizovaně předat metadata, která budou obsahovat alespoň minimální povinné údaje stanovené schématem DataCite a která zůstanou s přiděleným DOI trvale spjatá. Bez těchto metadat není možné identifikátor DOI pro datovou sadu získat.

Více na: <https://datacite.org/> nebo <https://support.datacite.org/>

3.3 DCAT (Data Catalog Vocabulary) a DCAT-AP

DCAT je standard pro popis datových sad, případně datových služeb, který má zajišťovat konzistenci popisu dat publikovaných ve webovém prostředí a umožňovat interoperabilitu mezi jednotlivými webovými katalogy dat. O rozvoj a aktualizace standardu se stará pracovní skupina Dataset Exchange Working Group (DXWG), působící pod mezinárodním konsorciem World Wide Web Consortium (W3C). (Albertoni, Browning, Cox, Gonzales Beltran, Perego a Winstanley, 2020). První verze je z roku 2014, aktuální verze DCATu je verze 2 z roku 2020a je k dispozici už i draft verze 3 ze 7. března 2023 (Albertoni, Browning, Cox, Gonzales Beltran, Perego a Winstanley, 2023), přičemž specifikace bude ještě v roce 2023 posunuta do stavu Candidate Recommendation s předpokládaným datem vydání finálního doporučení v roce 2024 (W3C, 2023).

Schéma je vyjádřeno jako RDF slovník. Nevěnuje se pouze popisu datasetů, verze 2 uvádí celkem 15 tříd (classes), z nichž každá třída se věnuje popisu určité entity (datasetu jako takového, katalogu, datové služby atd.), každá třída má dále určitý počet popisných prvků (properties). Vzhledem k třídám a prvkům v nich je schéma hierarchické, ale v jednotlivých třídách jsou prvky vždy na stejné úrovni a dále se nečlení. DCAT jako takový je definován jako slovník sjednocující základní pojmy používané při katalogizaci dat na webu. Není určen k přímému použití, nýbrž má sloužit jako základ pro aplikační profily, které teprve specifikují, jak má být DCAT použit v nějakém konkrétním kontextu. Proto DCAT jako takový nestanovuje povinnosti, opakovatelnosti, ani řízené slovníky pro použití s konkrétními definovanými hodnotami. (Albertoni, Browning, Cox, Gonzales Beltran, Perego a Winstanley, 2020).

Více na: <https://www.w3.org/TR/vocab-dcat-3/>

DCAT používají systémy a portály pro zveřejňování otevřených dat z oblasti veřejného sektoru a státní správy (např. European Union Open Data Portal, Govdata – das Datenportal für Deutschland, Open Government Canada a další), ale i systémy pro ukládání a zveřejňování výzkumných dat (např. Health Data Research Innovation Gateway z Velké Británie, Fairdata IDA z Finska nebo Norwegian Polar Data Centre). (Re3data.org – Registry of Research Data Repositories, 2023).

DCAT slouží jako základ pro aplikační profily, které specifikují, jak DCAT použít v konkrétním případě. Aplikační profil staví na již existujících standardech, ale je více specifický. Určuje povinnosti a opakovatelnosti u existujících tříd a jejich vlastností, v menší míře pak může specifikovat nové třídy a vlastnosti pro použití v konkrétním kontextu, včetně užívání řízených slovníků. Pro Evropu je Evropskou komisí vydán aplikační profil **DCAT-AP**. Aktuálně se blíží vydání verze DCAT-AP 3.0.0, jejíž draft je z letošního roku. Schéma obsahuje základních 14 entit¹¹, které lze doplnit o 19 podpůrných entit. Pro každou entitu jsou opět specifikovány další prvky (properties). Ve schématu jsou rovněž určeny povinnosti těchto prvků. Ke schématu patří také řízené slovníky (pro jazyky, formáty atd.) vydávané Úřadem pro publikace Evropské unie,

¹¹ Změna terminologie oproti dřívějšímu označení „třídy“ ve verzi 2.0.

které jsou strukturované dle standardu SKOS¹² (Simple Knowledge Organization System). (Fragkou, 2023 ; Van Nuffelen, 2021).

Aplikační profily DCATu jako DCAT-AP pak mohou být ještě dále profilovány pro více specifická prostředí či obory, pro implementace existuje řada doporučení a průvodců. Například DCAT-AP pak má rozšíření GeoDCAT-AP pro popis geoprostorových dat, StatDCAT-AP pro popisy souborů statistických údajů, MobilityDCAT-AP pro popis dat z dopravy, apod. (European Commission, 2023). Kromě oborových profilů jsou obvyklé profily národní, k těm patří i **DCAT-AP-CZ** spravovaný Digitální a informační agenturou (DIA). Tento národní profil využívá z DCAT-AP 4 třídy: pro popis katalogu, datové sady, distribuce datové sady a datové služby, a přebírá evropské řízené slovníky. (Klímek, 2021) Nad rámec DCAT-AP specifikuje strukturu zápisu podmínek užití dat, které jsou kompatibilní s českým autorským právem. Pro lokální katalogy otevřených dat v ČR je DCAT-AP-CZ povinný.

Více na: <https://opendata.gov.cz/>

3.4 Data Documentation Initiative (DDI)

DDI je mezinárodní standard pro popis dat, zejména však dat získaných z průzkumů, anket, dotazníků a statistik, obvykle z oblastí sociálních věd, ekonomie a zdravotnictví (DDI Alliance, 2023a). Standard DDI je v gesci Data Documentation Initiative Alliance (DDI Alliance), mezinárodní iniciativě, na které spolupracují členské instituce (DDI Alliance, 2023b). Členy jsou např. MIT, Stanford University a další univerzity a výzkumné instituty (DDI Alliance, 2023e).

Jedná se v podstatě o tři schémata, založená na XML schématu nebo UML:

- DDI Codebook (DDI-C) – pro popis jednotlivých datasetů, je použit jako základní schéma v softwaru pro digitální repozitáře Dataverse;
- DDI Lifecycle (DDI-L) – pro dokumentaci jednotlivých fází životního cyklu dat, popis longitudinálních dat, umožňuje popis více sad najednou;
- DDI Cross-Domain Integration (DDI-CDI) – model pro popis a integraci různých typů a formátů dat z různých vědních oborů, který je zatím ve vývoji

(DDI Alliance, 2023h, DDI Alliance, 2023f).

Dále se budeme věnovat schématu DDI Codebook, které slouží vysloveně pro popis jednotlivých datových sad.

Schéma DDI Codebook se skládá z 5 hlavních částí. První určuje popisné prvky pro celý dokument jako takový, druhá část prvky studií – kolekci dat (informace o sběru dat atd.), třetí část je věnovaná prvkům pro popis jednotlivých souborů dat dokumentu, čtvrtá část obsahuje soupis proměnných prvků a poslední pátá navrhuje prvky pro popis dalších doprovodných materiálů (dotazníky, pokyny pro tazatele, uživatelské příručky atp.). Schéma je hierarchické, každá část obsahuje další atributy a elementy, ve schématu jsou rovněž určeny povinnosti a opakovatelnosti

¹² SKOS publikuje rovněž konsorcium W3C.

prvků (DDI Alliance, 2023d). Ke schématu patří také řízené slovníky (např. pro typ dat, časové zóny nebo typologii poznámek. (DDI Alliance, 2023c, DDI Alliance, 2023g).

DDI Codebook má momentálně zveřejněnou aktuální verzi 2.5 z roku 2012, na loňský rok byla avizována verze 2.6, která byla otevřená k připomínkování do konce října. Neproběhlo však ještě oficiální zveřejnění této verze. (DDI Alliance, 2023d).

Více na: <https://ddialliance.org/>

3.5 Mapování¹³ mezi analyzovanými schématy

V rámci podpory interoperability metadat z různých repozitářů dochází k vytváření generických mapování mezi rozšířenými schématy, tzv. crosswalks. Vzhledem k tomu, že se analýza zaměřila na nejčastěji užívaná schémata, tak mezi nimi již zmiňované crosswalks většinou existují. U schématu DataCite existuje mapování jak na schéma [Dublin Core](#)¹⁴, tak na [DCAT-AP](#)¹⁵. Pro další zkoumání bude zásadní zejména mapování DataCite na DCAT-AP, tzv. CiteDCAT-AP, který vytvořilo Joint Research Center (JRC) při Evropské komisi. Jedná se zatím o neoficiální verzi mapování mezi verzí DataCite 4.4 a DCAT-AP 2.0. Nicméně podle vyjádření člena týmu JRC je v plánu publikování oficiální verze a také aktualizace na verzi DCAT-AP 3.0 a případně DataCite 4.5 (zatím nevydaná). (Perego, 2021 ; Perego, 2023)

DDI Alliance oficiálně na svém webu uvádí, že DDI Codebook ve verzi 2 má vytvořené [mapování](#) na Dublin Core¹⁶. V rámci projektu Social Sciences and Humanities Open Cloud (SSHOC) byly navíc v loňském roce vytvořeny [konverze](#) DDI Codebook ve verzi 2.5 na DCAT-AP, Dublin Core i DataCite.

V podstatě tedy schází pouze mapování mezi Dublin Core a DCAT-AP, nicméně DCAT prvky DCMI Metadata Terms přímo využívá, mapování užitých prvků by tedy nemělo být příliš složité.

Nicméně tato analýza se nezabývá mapováním mezi danými schématy a zmíněné podklady pro konverze nebyly dále zkoumány a hodnoceny co do kvality. Je nutné počítat s tím, že po detailnější analýze jednotlivých mapování může dojít ke zjištění, že by bylo nutné zpracovat mnohem detailnější podklady a implementace konverzí mezi jednotlivými schématy bude náročnější.

¹³ Schéma nebo tabulka, která představuje sémantické mapování polí nebo datových prvků v jednom datovém standardu k polím nebo datovým prvkům ve druhém datovém standardu, který má podobnou funkci nebo význam. Mapuje vztahy a ekvivalence mezi dvěma nebo více metadatovými formáty. Tyto křížové převody dat umožňují vyhledávacím strojům současné prohledávání v heterogenních databázích zadáním jediného dotazu, jako kdyby byly jedinou databází (sémantická interoperabilita) a efektivně převádět data z jednoho metadatového standardu do jiného. Zdroj: CELBOVÁ, Ludmila. Mapování metadat. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha : Národní knihovna ČR, 2003- [cit. 2023-08-30]. Dostupné z:

https://aleph.nkp.cz/F/?func=direct&doc_number=000000542&local_base=KTD.

¹⁴ https://schema.datacite.org/meta/kernel-4.4/doc/DataCite_DublinCore_Mapping.pdf

¹⁵ <https://ec-jrc.github.io/datacite-to-dcat-ap/>

¹⁶ <https://ddialliance.org/resources/ddi-profiles/dc>

4 Závěr

Analýza vybraných univerzálních schémat užívaných pro popis výzkumných dat ukázala, že všechna analyzovaná schémata mají záštitu významných organizací, společností nebo iniciativ. U všech probíhají revize a aktualizace a řeší se požadavky uživatelů. Pouze u schématu DDI nebyla vydána delší dobu nová verze, ostatní schémata byla všechna aktualizována v posledních třech letech. U všech schémat existují odkazy na dokumentace a pomůcky pro jejich implementaci.

Pro doporučení určitého schématu je tedy nutné se zaměřit zejména na schémata jako taková (jejich architekturu, množství popisných prvků a podporované formáty) a okolnosti jejich užití. Vybraná schémata zároveň není jednoduché srovnávat ani z těchto hledisek. Zatímco DataCite, DCMI Metadata Terms a DDI Codebook jsou určena přímo pro popis jednotlivých objektů, DCAT umožňuje popis více entit, nejen jednotlivých objektů, ale i nadřazených celků. Navíc je ze své podstaty RDF slovníku, stejně jako DCMI Metadata Terms, volně kombinovatelný a rozšiřitelný. Srovnávání jejich hierarchií nebo počtu prvků tedy není zcela vypovídající. Při analýze schématu DCAT, popřípadě DCAT-AP a DCAT-AP-CZ, byla zkoumána především třída pro popis datové sady (dataset).

DCMI Metadata Terms je sám o sobě pro popis výzkumných dat poměrně obecný, což je dáno jeho zaměřením na definici pouze základních metadatových prvků, nejsou v něm tedy např. určené povinnosti. Schéma nepracuje dostatečně přehledně a detailně s popisem souvisejících zdrojů (tzv. related items), které mohou být pro popis výzkumných dat poměrně klíčové (provázání záznamu datové sady na článek, jinou datovou sadu, použitý software aj.). DCMI Metadata Terms také neobsahuje specifické údaje o financování, tj. číslo projektu, název poskytovatele, název projektu atd.

Schéma DDI je sice zmíněno v seznamu univerzálních metadatových schémat, ale jak bylo výše uvedeno, je spíše interdisciplinární se zaměřením na sociální vědy, ekonomiku a zdravotnictví a je užíváno pro popis statistických dat, dat z průzkumů, anket atd. Pro základní schéma se zároveň zdá být příliš detailní a pojmenování jednotlivých elementů (údajů) není vždy intuitivní.

DataCite je skutečně univerzální schéma zaměřené na obecné údaje o datových sadách (a případně i jiných výstupů z výzkumu). Zároveň jde do dostatečného detailu, co se týče údajů o financování výzkumu, popisu datové sady, identifikace autorů a přispěvatelů i popisu souvisejícího zdroje. Možná je zbytečně detailně rozveden popis geolokace, nicméně se jedná o nepovinný prvek. Kromě toho je nutné toto schéma respektovat kvůli přidělování identifikátorů DOI, které je často vyžadováno pro trvalou identifikaci datových sad. Identifikátory DOI nemohou být přiděleny bez balíčku povinných metadatových údajů, proto tyto údaje musí být vyžadovány pro popis datových sad i v ČR.

DCAT bude potřeba vzít v úvahu jako standard pro interoperabilitu záznamů na webu, jelikož zejména DCAT-AP se v evropském kontextu prosazuje v rámci společných evropských datových prostorů. DCAT-AP se používá jak pro popis otevřených dat, tak i dat, která otevřená nejsou. (Data Spaces Support Centre, 2023). Data z výzkumu jsou nově uznávaným výstupem výzkumu

a vývoje a v novele zákona č. 130/2002 Sb. bylo nově uzákoněno, že příjemce podpory na výzkum a vývoj je povinen (až na uvedené výjimky) poskytovat přístup k výzkumným datům. Paragraf 12a pak uvádí, že by výzkumná data měla být poskytována v otevřeném a strojově čitelném formátu, tj. jako otevřená data. Bude tedy nutné, aby repozitáře pro výzkumná data byly harvestovatelné Národním katalogem otevřených dat.

Jak bylo již zmíněno, dvě ze čtyř blíže analyzovaných schémat (DataCite a DCAT-AP) nelze opomenout při definování základního modelu vzhledem k jejich kontextu. V případě DataCite jde o jeho rozšíření ve stávajících repozitářích výzkumných dat a relativní věcná úplnost vzhledem k potřebám vědeckých komunit. Navíc dodržování schématu DataCite je podmínkou pro přidělování identifikátoru DOI výzkumným datům. V případě DCAT-AP/DCAT-AP-CZ se jedná o interoperabilitu s ostatními datovými portály na webu, o užití v rámci společných evropských datových prostorů, a v případě otevřených výzkumných dat také o legislativní záštitu. Profil DCAT-AP a jeho profil DCAT-AP-CZ byl vyvinut pro jinou oblast (veřejná a státní správa) a tedy i jiný druh dat. Z výše uvedeného je jasné, že základní model pro popis výzkumných dat v českém prostředí bude muset obsahovat prvky potřebné jak pro reprezentaci metadat ve schématu DataCite, tak pro reprezentaci metadat dle DCAT-AP a DCAT-AP-CZ, aby mohla být zajištěna automatizovaná výměna metadatových záznamů dle těchto schémat s potřebnými lokálními i mezinárodními systémy. V následující fázi bude třeba identifikovat, které konkrétní prvky DataCite, DCAT-AP a DCAT-AP-CZ budou potřeba v české komunitě VaVal a zda bude potřeba schémata upravit (např. změnou povinností jednotlivých prvků nebo rozšířením o další prvky). Po stanovení základního modelu pro popis výzkumných dat v ČR doporučujeme zajistit konverze na původní schémata DataCite za účelem získávání DOI pro datové sady a DCAT-AP-CZ za účelem napojení na Národní katalog otevřených dat a skrze něj na Oficiální portál evropských dat (data.europa.eu). Do budoucna doporučujeme zvážit konverzi na Dublin Core vzhledem k jeho častému užití v repozitářích pro výzkumná data a využití schématu DDI jako multioborové rozšíření pro oblast statistických dat.

Příloha č. 1 – Tabulkový přehled vybraných schémat

Schéma	DataCite	DCAT	DCAT-AP	DCAT-AP CZ	DDI Codebook	Dublin Core
Tvůrce	DataCite	Dataset Exchange Working Group (DXWG)	European Commission - SEMIC team		Data Documentation Initiative Alliance (DDI Alliance)	Dublin Core Metadata Initiative
Rok vzniku	2009	2014	2013?	2019?	2000	1995
Aktuální verze	4.4 (avizována 4.5 a 5)	2 (verze 3 draft)	3.0.0	2021-01-11	2.5 (avizována 2.6)	2020-01-20
Rok vzniku aktuální verze	2021	2020	2023	2021	2014	2020
Dokumentace	ano (https://schema.datacite.org/)	ano (https://www.w3.org/TR/vocab-dcat-2/)	ano (https://github.com/SEMICeu/DCAT-AP/tree/master/releases/3.0.0)	ano (https://ofn.gov.cz/rozhran%C3%AD-katalog%C5%AF-otev%C5%99en%C3%BDch-dat/2021-01-11/)	ano (https://ddialliance.org/Specification/DDI-Codebook/2.1/DTD/Documentation/version2-1-all.html#5.0)	ano (https://www.dublincore.org/specifications/dublin-core/dcmi-terms/)
Licence/podmínky užití		W3C Document License (https://www.w3.org/Consortium/Legal/2015/doc-license)	CC BY 4.0		GNU General Public License, version 3 of the License, or (at your option) any later version.	CC BY
Počet hlavních prvků	20 prvků (properties)	15 tříd	14 hlavních entit, 19 podpůrných	4 třídy	5 částí (celkový počet prvků 351)	15 prvků
Hierarchické schéma	ano	ano	ano	ano	ano	ne
Vyjádření povinnosti polí	ano	pouze doporučení u některých prvků	ano	ano	ano	ne
Vyjádření opakovatelnosti polí	ano	ne	ano	ne	ano	ne
Reprezentace/zobrazení	XML, JSON	RDF, RDF/XML, JSON-LD, Turtle, HTML-RDFa, N3	RDF, RDF/XML, UML	RDF, JSON	XML	RDF, XML, JSON, UML
Doporučené slovníky	ano	odkazuje na rejstříky či slovníky	ano	odkazuje na rejstříky či slovníky	ano	ano
Formát slovníků	list	schéma	schéma	schéma	schéma	list
Je povinné pro užití?	přidělení DOI přes DataCite	katalogy otevřených dat	katalogy otevřených dat	katalogy otevřených dat		

Použité zdroje

ALBERTONI, Riccardo, BROWNING, David, COX, Simon, GONZALEZ BELTRAN, Alejandra, PEREGO, Andrea a WINSTANLEY, Peter, ed., 2020. *Data Catalog Vocabulary (DCAT) - Version 2: W3C Recommendation 04 February 2020* [online]. W3C [cit. 2023-05-15]. Dostupné z: <https://www.w3.org/TR/vocab-dcat-2/>

ALBERTONI, Riccardo, BROWNING, David, COX, Simon, GONZALEZ BELTRAN, Alejandra, PEREGO, Andrea a WINSTANLEY, Peter, ed., 2023. *Data Catalog Vocabulary (DCAT) - Version 3: W3C Working Draft 07 March 2023* [online]. W3C [cit. 2023-03-06]. Dostupné z: <https://www.w3.org/TR/vocab-dcat-3/#abstract>

ČESKÁ REPUBLIKA, 2023a. Zákon č. 106/1999 Sb., o svobodném přístupu k informacím. In: *Sbírka zákonů ČR*. 39/199. [cit. 2023-05-15]. Dostupné také z: <https://www.zakonyprolidi.cz/cs/1999-106>

ČESKÁ REPUBLIKA, 2023b. Zákon č. 130/2002 Sb.: Zákon o podpoře výzkumu a vývoje z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje). In: *Sbírka zákonů ČR*. 56/2002. [cit. 2023-05-15]. Dostupné také z: <https://www.zakonyprolidi.cz/cs/2002-130>

DataCite Metadata Working Group, 2021. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. <https://doi.org/10.14454/3w3z-sa82>

DataCite Metadata Working Group, 2022. DataCite Metadata Schema 4.5: Request for Comments. Dostupné z: https://datacite.org/documents/DataCite_Metadata_Schema_4.5_RFC.pdf

DataCite, 2023a. *DataCite Schema* [online]. [cit. 2023-03-06]. Dostupné z: <https://schema.datacite.org/>

DataCite, 2023b. DataCite's Value. *DataCite* [online]. 2023, 2012-01-17 [cit. 2023-03-06]. Dostupné z: <https://datacite.org/value.html>

DATA SPACES SUPPORT CENTRE, 2023. *Data Spaces Blueprint Version 0.5* [online]. Munich: Data Spaces Support Centre [cit. 2023-12-07]. Dostupné z: <https://dssc.eu/space/BPE/179175433/Data+Spaces+Blueprint+%7C+Version+0.5+%7C+September+2023>

DCMI, 2023a. About DCMI. *DublinCore* [online]. [cit. 2023-03-06]. Dostupné z: <https://www.dublincore.org/about/>

DCMI, 2023b. DCMI History. *DublinCore* [online]. [cit. 2023-05-15]. Dostupné z: <https://www.dublincore.org/about/history/>

DCMI, 2023c. DCMI Metadata Terms. *DublinCore* [online]. [cit. 2023-03-06]. Dostupné z: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

DDI Alliance, 2023a. *DDI* [online]. [cit. 2023-03-06]. Dostupné z: <https://ddialliance.org/>

DDI Alliance, 2023b. About the Alliance. *DDI* [online]. [cit. 2023-03-06]. Dostupné z: <https://ddialliance.org/>

DDI Alliance, 2023c. Controlled vocabularies. *DDI* [online]. [cit. 2023-03-06]. Dostupné z: <https://ddialliance.org/>

DDI Alliance, 2023d. DDI-Codebook 2.5. *DDI* [online]. [cit. 2023-03-06]. Dostupné z: <https://ddialliance.org/Specification/DDI-Codebook/2.5/>

DDI Alliance, 2023e. DDI Membership. *DDI* [online]. [cit. 2023-03-06]. Dostupné z: <https://ddialliance.org/membership/overview>

DDI Alliance, 2023f. Overview of Current Products. *DDI* [online]. 2023 [cit. 2023-03-06]. Dostupné z: <https://ddialliance.org/products/overview-of-current-products>

DDI Alliance, 2023g. XML Schema Documentation. *DDI* [online]. [cit. 2023-03-06]. Dostupné z: https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation.html

DDI Alliance, 2023h. What is DDI. *DDI* [online]. [cit. 2023-03-06]. Dostupné z: <https://ddialliance.org/learn/what-is-ddi>

EUROPEAN COMMISSION, 2023. DCAT Application Profile for data portals in Europe: About DCAT Application Profile for data portals in Europe. *Joinup* [online]. [cit. 2023-05-15]. Dostupné z: <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/about>

FRAGKOU, Pavlina, 2023. DCAT-AP 3.0 [online]. [cit. 2023-08-08]. Dostupné z: <https://semiceu.github.io/DCAT-AP/releases/3.0.0/>

General Research Data, 2023. *Digital Curation Centre* [online]. [cit. 2023-03-06]. Dostupné z: <https://www.dcc.ac.uk/resources/subject-areas/general-research-data>

KLÍMEK, Jakub, 2020. Kontext otevřených dat. *Otevřená data* [online]. [cit. 2023-03-06]. Dostupné z: <https://opendata.gov.cz/informace/kontext:start>

KLÍMEK, Jakub, ed, 2021. *Rozhraní katalogů otevřených dat: DCAT-AP-CZ: Doporučení 11. ledna 2021* [online]. [cit. 2023-05-15]. Dostupné z: <https://ofn.gov.cz/rozhran%C3%AD->

[katalog%C5%AF-otev%C5%99en%C3%BDch-dat/2021-01-11/#polo%C5%BEky-distribuce-form%C3%A1t](#)

KREJČÍŘ, Vlastimil, 2006. Systémy pro tvorbu digitálních knihoven. In: *INFORUM 2006: 12. ročník konference o profesionálních informačních zdrojích, Praha 23.-25. května 2006* [online]. Praha: Albertina icome Praha, [cit. 2023-06-09]. ISSN 1801-2213. Dostupné z: <https://www.inforum.cz/archiv/inforum2006/prispevek.php-prispevek=24.htm>

Metadata Standards Catalog [online], 2023. [cit. 2023-03-06]. Dostupné z: <https://rdamsc.bath.ac.uk/>

Otevřené formální normy (OFN), 2023. *Portál otevřených dat* [online]. [cit. 2023-05-16]. Dostupné z: <https://data.gov.cz/ofn/>

PEREGO, Andrea, 2023. CiteDCAT-AP Governance information.Online. In: European Commission, Joint Research Centre (JRC) GitHub. 27. 9. 2023. Dostupné z: <https://github.com/ec-jrc/datacite-to-dcat-ap/issues/6> [cit. 2023-11-16]

PEREGO, Andrea, AUSTIN, Timothy, FRIIS-CHRISTENSEN, Anders, VACCARI, Lorenzino a TSINARAKI, Chrisa, 2021. DataCite to DCAT-AP Mapping: Unofficial Draft 31 March 2021 [online]. [cit. 2023-05-15]. Dostupné z: <https://ec-jrc.github.io/datacite-to-dcat-ap/>

Re3data.org - Registry of Research Data Repositories [online], 2023. [cit. 2023-05-16]. Dostupné z: <https://www.re3data.org/>. DOI: 10.17616/R3D

SMĚRNICE EVROPSKÉHO PARLAMENTU a RADY (EU) 2019/1024 ze dne 20. června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru (přepracované znění), 2019. In: *Úřední věstník Evropské unie*. [cit. 2023-05-15]. Dostupné také z: <https://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:32019L1024&from=CS>

VAN NUFFELEN, Bert, 2021. *DCAT Application Profile for data portals in Europe: Version 2.1.0* [online]. [cit. 2023-05-15]. Dostupné z: https://joinup.ec.europa.eu/sites/default/files/distribution/access_url/2021-12/5bf41792-1a2f-4851-ae2-6ecf43815bc1/dcat-ap_2.1.0.pdf

W3C, 2023. *Dataset Exchange Working Group (DXWG)* [online]. [cit. 2023-12-07]. Dostupné z: https://www.w3.org/2017/dxwg/wiki/Main_Page#Dates_for_CR