

## NOVINKY ZAHRANIČNÍ KNIHOVNICKÉ LITERATURY

**ALBANI, Mirko, Rosemarie LEONE, Federica FOGLINI, Francesco DE LEO, Fulvio MARELLI a Iolanda MAGGIO. EVER-EST: The Platform Allowing Scientists to Cross-Fertilize and Cross-Validate Data [EVER-EST: Platforma umožňující vědcům obohacovat a potvrzovat data]. *Data Science Journal* [online]. 2020, vol. 21, article 30, s. 1–16 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-021>.**

V oblasti věd o Zemi hrají spolehlivá data a přístup k odpovídající výpočetní kapacitě a vizualizačním a analytickým nástrojům klíčovou roli. Stejně důležité je i sdílení dat s ostatními vědci, např. za účelem jejich validace a opětovného využití. V projektu EVER-EST (*European Virtual Environment for Research – Earth Science Themes: a solution*, <https://ever-est.eu/>) financovaném z programu Horizont 2020 bylo proto vyvinuto virtuální výzkumné prostředí, které tyto potřeby naplňuje a je orientováno na uživatele. Prostor umožňuje správu celého výzkumného cyklu. Vše je navrženo tak, aby bylo možné uplatňovat zásady FAIR (*Findability, Accessibility, Interoperability, and Reuse of digital assets*) a systematický přechod na otevřenou vědu.

Ústřední roli hrají výzkumné objekty, což jsou sémanticky bohatá seskupení zdrojů vztahujících se ke výzkumu (k datům, metodám a osobám). Výzkumný objekt obsahuje pracovní postup, vstupní data a výsledky a rovněž článek, který shrnuje výsledky a odkazuje na osoby odpovědné za výzkum. Nejdůležitější je právě vědecký pracovní postup, tedy řada strukturovaných činností a výpočtů, k nimž dochází při řešení vědeckých problémů. Jde o zaznamenání kroků komplexní analýzy, výpočetních experimentů včetně použitých nástrojů a platform a jejich sdílení s dalšími vědci. Zaznamenané pracovní postupy umožňují vyhodnocení reprodukovatelnosti výsledků, mohou být opětovně využity týmem nebo jiným vědcem (i za jiným účelem), sloužit k ověření metody a jako výukové pomůcky ukazující možnosti stávající datové infrastruktury.

Ze strany vědců to předpokládá určitou formalizaci procesů a využívání softwaru s otevřeným zdrojovým kódem (aby bylo možné sdílení a aby byla zajištěna reprodukovatelnost) a další dílčí dovednosti. Právě to může být – spolu s časovou náročností – značnou překážkou při využívání výzkumných objektů v praxi. Projekt EVER-EST proto přichází s prostředím (*Virtual Research Environment, VRE*), které řadu činností automatizuje. Pracuje s tzv. sférou spolupráce, což je uživatelské rozhraní pro vizualizaci korelací mezi podobnými objekty. Kolem uživatele se nacházejí tři sféry: vnitřní sféra (kontext zájmu uživatele), prostřední sféra (položky doporučené na základě kontextu zájmu uživatele) a vnější sféra (výzkumné objekty a uživatelé doporučení na základě předchozí aktivity uživatele).

Architektura infrastruktury EVER-EST je tvořena třemi vrstvami – vrstvou prezentační, servisní a datovou – a dalšími dvěma složkami, jimiž jsou hlavní uživatelské rozhraní a podniková servisní sběrnice. Platforma umožňuje vzdáleně přistupovat k datům, softwaru, výsledkům výzkumu i dokumentaci, zachytit pracovní postupy v jediném objektu, který je identifikovatelný prostřednictvím DOI, spolupracovat s kolegy v různých částech světa, dokumentovat vědeckou práci, zveřejňovat šedou literaturu a zajišťovat dlouhodobé uchování dat, softwaru, výsledků i jejich interpretace.

Samotné virtuální výzkumné prostředí je validováno čtyřmi virtuálními výzkumnými uskupeními z různých podoblastí věd o Zemi (sledování oceánu, přírodní rizika, sledování země a řízení rizik souvisejících se sopečnou činností a zemětřesením).

Funkční prostředí, které splňuje požadavky uživatelů zabývajících se výše uvedenými podoblastmi věd o Zemi, bylo vytvořeno v rámci tříletého projektu. Byl také vytvořen plán udržitelnosti, aby byly nadále dostupné výsledky projektu, a další úsilí směřuje k zajištění plnohodnotného provozu platformy i do budoucna. Projekt doložil, že standardizace

výzkumných objektů a interoperabilita přispívají k inovacím a otevřené vědě (dodržující zásady FAIR).

Mezi klíčová zjištění patří skutečnost, že bohatá a výstižná metadata jsou klíčovým faktorem pro sdílení a opětovné využívání dat, že vědecké výsledky musí být viditelné a snadno nalezitelné, že vědci potřebují za svou práci být řádně oceněni a že správa výzkumných objektů musí být začleněna do nástrojů, které vědci nyní používají.

**BAHIM, Christophe, Carlos CASORRÁN-AMILBURU, Makx DEKKERS, Edit HERCZOG, Nicolas LOOZEN, Konstantinos REPANAS, Keith RUSSELL a Shelley STALL.** **The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments [Model pro dosažení vyspělosti dat z hlediska zásad FAIR: možný způsob sladění hodnocení FAIR].** *Data Science Journal* [online]. 2020, vol. 19, article 41, s. 1–7 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-041>.

Vznikla již celá řada metodik a nástrojů sloužících ke zhodnocení míry, do níž data z výzkumu naplňují zásady FAIR (*Findability, Accessibility, Interoperability, and Reuse of digital assets*). Různé metodiky však různým způsobem interpretují samotné zásady FAIR. Byla proto vytvořena pracovní skupina, jejímž cílem bylo zpracovat model pro dosažení vyspělosti dat, který by měl sloužit jako prostředek pro srovnání hodnocení vzniklých na základě různých metodik a nástrojů.

Samotné zásady FAIR nestanovují, jakým způsobem se jich má dosáhnout. Ve Sdružení pro výzkumná data (*Research Data Alliance, RDA*) proto vznikla pracovní skupina, jejímž cílem bylo vytvořit sadu hodnotících kritérií, která by usnadnila srovnání různých přístupů k hodnocení a byla výsledkem shody mezi více než dvěma sty členy pracovní skupiny. Práce na modelu probíhaly od ledna 2019 do června 2020.

Příprava modelu byla rozdělena do čtyř etap. V první etapě došlo ke zpřesnění zadání a rozsahu úlohy a průzkumu terénu. Druhá etapa spočívala v tvorbě modelu; jejím výsledkem byl návrh sady ukazatelů s mírou jejich důležitosti a návrh využití ukazatelů v praxi. V třetí etapě proběhlo testování s cílem ověřit vhodnost, úplnost a použitelnost ukazatelů; výsledkem byla revize modelu a návrh doporučení RDA. V poslední (čtvrté) etapě proběhla finalizace modelu a v červnu 2020 jeho zveřejnění pod názvem *FAIR Data Maturity Model* (<https://doi.org/10.15497/RDA00050>). Model má formu doporučení RDA.

Vytvořené ukazatele mohou být využity k hodnocení datových zdrojů a metadat, která se k nim váží. Na jejich základě mohou být vytvořeny hodnotící metodiky. K ukazatelům je jednak přiřazena míra důležitosti, jednak jsou k dispozici dva způsoby hodnocení – na škále od jedné do pěti (měření pokroku při splňování požadavků ukazatele) a odpověď ano/ne (zjišťování splnění/nesplnění ukazatele).

Model je možné využít při zpracovávání plánů správy dat, tj. ještě předtím, než jakákoliv výzkumná data vůbec vzniknou. Předem tak lze stanovit, do jaké míry mají budoucí data splňovat zásady FAIR. S modelem lze pracovat i po vytvoření dat – jeho prostřednictvím lze ověřit, zda bylo žádoucí úrovně dosaženo. Tvůrcům dat může model posloužit i k tomu, aby zjistili, jak zlepšit míru splnění požadavků FAIR.

Důležitým vodítkem je doporučení, že hodnocení nemá mít význam samo o sobě, ale má být vnímáno jako prostředek pro zlepšení dat tak, aby se dostala na vyšší úroveň naplnění požadavků FAIR. Zásady FAIR je třeba chápat jako určitou metu, o jejíž dosažení by se tvůrci dat měli snažit. Navíc jde o určitý výhled do budoucna, o zlepšování možností opětovného využití dat. Zásady tedy není žádoucí interpretovat jako zcela přesná pravidla; při jejich uplatňování v praxi je vhodné brát ohled na specifika konkrétních vědeckých oborů a uskupení.

Při zpracování modelu dospěla pracovní skupina RDA k závěru, že podstatnou roli nehrají pouze data samotná, ale také metadata. Metadatům dokonce přikládá takovou

důležitost, že se domnívá, že by jim měl být přidělován samostatný trvalý identifikátor. To je však v rozporu se současnou praxí, kdy je obvykle používána vstupní stránka obsahující jak metadata, tak odkaz na vlastní data.

Osvědčila se také snaha o nalezení shody v pracovní skupině. Díky tomu se očekává, že model bude reálně použitelný v celé řadě vědeckých oborů a přispěje k lepší mezioborové spolupráci. Jde také o první krok, po němž může následovat vytvoření metodiky k hodnocení vědeckých výsledků financovaných z veřejných prostředků z hlediska míry splnění zásad FAIR. Významné kroky k zavedení zásad FAIR do praxe podnikla Geologická služba USA (*U.S. Geological Survey*, USGS), Národní úřad pro oceán a atmosféru (*National Oceanic and Atmospheric Administration*, NOAA) nebo Evropská komise v rámci politiky Otevřená věda. Zásady FAIR jsou také jádrem Evropského cloudu pro otevřenou vědu (*European Open Science Cloud*), což je iniciativa, jejímž cílem je vybudovat důvěryhodný, otevřený a distribuovaný systém pro přístup k datům splňujícím zásady FAIR a nabídnout nad nimi nadstavbové služby. Zásady FAIR hrají rovněž důležitou úlohu v rámcovém programu Horizont Evropa.

**CARROLL, Stephanie Russo, Ibrahim GARBA, Oscar L. FIGUEROA-RODRÍGUEZ, Jarita HOLBROOK, Raymond LOVETT, Simeon MATERECHERA, Mark PARSONS, Kay RASEROKA, Desi RODRIGUEZ-LONEBEAR, Robyn ROWE, Rodrigo SARA, Jennifer D. WALKER, Jane ANDERSON a Maui HUDSON. The CARE Principles for Indigenous Data Governance [Zásady CARE pro správu dat o původních obyvatelích]. *Data Science Journal* [online]. 2020, vol. 19, article 43, s. 1–12 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-043>.**

Původní, autochtonní obyvatelé (*Indigenous Peoples*) se vyskytují ve více než 90 zemích; celkem jde o více než 370 milionů osob reprezentujících přes pět tisíc různých kultur. Původní národy udržely kontinuitu s předkoloniálními společnostmi a je pro ně charakteristická snaha o využití vlastních společenských, politických a hospodářských systémů k uchování, rozvoji a přenosu své kultury, poznatků a vztahů s místem a zdroji budoucím generacím.

Data vztahující se k původním obyvatelům tvoří informace a poznatky o prostředí, zemi, nebi, zdrojích a jiných než lidských bytostech, dále informace o původních obyvatelích (např. administrativní, zdravotní, společenské, obchodní a firemní informace nebo informace ze sčítání lidu) a informace a poznatky o původních obyvatelích jako kolektivech, včetně tradičních a kulturních informací, ústně tradovaných příběhů, znalostí o předcích a klanech apod.

V souvislosti se zpracováním velkých objemů dat dochází k tomu, že většina dat o původních obyvatelích je v držení vlád, institucí a agentur. Ty často reprezentují jiné hodnoty, než které jsou důležité pro původní obyvatele.

Pro správu dat o původních obyvatelích jsou proto formulovány čtyři základní zásady, a to společný prospěch (*Collective Benefit*), oprávnění ke správě (*Authority to Control*), odpovědnost (*Responsibility*) a etika (*Ethics*). Jejich počáteční písmena vytvořila název CARE – *Collective Benefit, Authority to Control, Responsibility, and Ethics*. Návrh zásad byl vytvořen na workshopu v Botswaně v roce 2018. Jejich smyslem je umožnit původním obyvatelům rovný přístup k opětovnému využívání těchto dat. Zásady CARE jsou navrženy tak, že doplňují zásady FAIR (viz předchozí článek, pozn. red.). Je tedy možné je používat souběžně.

Mezi organizace, které se zásadami CARE již začaly pracovat, se řadí Sdružení pro výzkumná data (*Research Data Alliance*, <https://rd-alliance.org/>) v oblasti výzkumu, Smithsonianovský institut (*Smithsonian Institution*, <https://www.si.edu/>) v oblasti sbírek

a Charta otevřených dat (*Open Data Charter*, <https://opendatacharter.net/>) v oblasti státní správy.

Jsou-li zásady CARE uplatňovány souběžně se zásadami FAIR, jsou výsledkem data, jež odrážejí realitu původních obyvatel, jsou pro ně užitečná a zůstávají v jejich správě, ale zároveň podporují objevování nových poznatků a inovace.

**DAVID, Romain, Laurence MABILE, Alison SPECHT, Sarah STRYECK, Mogens THOMSEN, Mohamed YAHIA, Clement JONQUET, Laurent DOLLÉ, Daniel JACOB, Daniele BAILO, Elena BRAVO, Sophie GACHET, Hannah GUNDERMAN, Jean-Eudes HOLLEBECQ, Vassilios IOANNIDIS, Yvan LE BRAS, Emilie LERIGOLEUR a Anne CAMBON-THOMSEN. FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles [Znalost zásad FAIR: Achillova pata jejich využívání v praxi]. *Data Science Journal* [online]. 2020, vol. 19, article 32, s. 1–11 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-032>.**

V roce 2017 vznikla v rámci Sdružení pro výzkumná data (*Research Data Alliance*, RDA) Zájmová skupina pro odměny a zásluhy za sdílení (*SHARing Rewards and Credit Interest Group*, SHARC IG). V praxi se ukazuje, že příprava dat tak, aby splňovala zásady FAIR (*Findability, Accessibility, Interoperability, and Reuse of digital assets*, viz též předchozí články v rubrice, pozn. red.), vyžaduje značné množství času, energie a odborných znalostí a dovedností. Proto je zpřístupňování dat v souladu se zásadami FAIR zatím považováno za prioritu především v těch vědeckých disciplínách, které běžně zpracovávají velké objemy dat (těmi jsou např. jaderná fyzika, astronomie či genomika). Na druhou stranu grantové agentury nyní velmi podporují zařazování zásad FAIR do plánů správy dat; v některých případech dokonce jde o povinnost. Týká se to např. řady grantových výzev Evropské komise (včetně programu Horizont 2020).

Existují různé metody a nástroje, které podporují zásady FAIR (zejména pokud jde o strojovou využitelnost dat bez zásahu člověka); ty ale nejsou příliš srozumitelné pro vědce, kteří nejsou zvyklí rutinně pracovat s velkými objemy dat. Tyto vědce je proto nejprve třeba motivovat, aby měli sami zájem o zpřístupňování svých dat v souladu se zásadami FAIR. Z toho důvodu zájmová skupina SHARC (<https://www.rd-alliance.org/groups/sharing-rewards-and-credit-sharc-ig>) vyvinula nástroj, který je zaměřen na hodnocení míry splnění zásad FAIR a má podobu srozumitelnou širokému spektru vědců. V rámci činnosti skupiny vznikl i slovník, který obsahuje vymezení důležitých pojmů. Skupina SHARC zpracovala šablonu vycházející z otázek, které si vědci kladou při zpracovávání plánů správy dat. Šablonu si různé vědecké komunity mohou podle potřeby upravit a je navíc vytvořena tak, aby ji bylo možné využít při školení a k podpoře zlepšení míry souladu dat se zásadami FAIR. Hodnotící nástroj je k dispozici prostřednictvím identifikátoru <http://doi.org/10.5281/zenodo.3922069>.

Při zpracování nástroje se ukázalo, že je třeba počítat s postupným zaváděním zásad FAIR, přičemž je nezbytné výslovně zachytit dosažení každé úrovně. V různých disciplínách může zavádění zásad FAIR narazit na různé překážky, např. u tradičních disciplín na nedostatečný zájem o změnu zavedených zvyků nebo obvyklých nástrojů nebo v nových disciplínách na problém s přílišnou komplexností stávajících procesů. V každém případě hraje podstatnou úlohu vzdělávání vědců v oblasti zpřístupňování dat v souladu se zásadami FAIR. Samotné přípravné práce vyžadují souhru mezi grantovými agenturami, tvůrci politik a vydavateli (ti by měli data podle zásad FAIR požadovat), institucemi (ty by měly prostřednictvím svých organizačních složek, např. knihoven, poskytovat infrastrukturu, školení a podporu a stanovovat postupy) a vědeckými komunitami (ty by měly vytvářet vlastní normy).

Tzv. FAIRifikace by měla zahrnovat čtyři samostatné kroky, a to přípravu a školení zúčastněných stran a přípravu a vlastní zpracování dat do podoby, která bude v souladu se zásadami FAIR. Konkrétní postupy při tzv. FAIRifikaci dat se poměrně rychle vyvíjejí; vznikají např. nová data v nových formátech. Zpřístupnění dat v souladu se zásadami FAIR je proto možné pouze v případě, že je mu ve výzkumném projektu věnována dostatečná pozornost.

Dosud také platí, že vědci za úsilí věnované zpřístupňování dat nejsou nijak ohodnoceni. Předpokladem pro zavedení hodnocení je existence podpůrných mechanismů, které by jej umožnily, např. zavedení identifikátorů pro vědce a jejich data. Dobrou zprávou je, že v současné době již řada takových identifikátorů (pro vědecké výstupy, vědce i výzkumné organizace) existuje; je tedy na čem stavět.

**HETTNE, Kristina Maria, Peter VERHAAR, Erik SCHULTES a Laurents SESINK. From FAIR Leading Practices to FAIR Implementation and Back: An Inclusive Approach to FAIR at Leiden University Libraries [Od převažující praxe v oblasti FAIR k implementaci a zpět: celkový přístup k zásadám FAIR v Knihovnách Univerzity v Leidenu]. *Data Science Journal* [online]. 2020, vol. 19, article 40, s. 1–7 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-040>.**

Univerzita v Leidenu (<https://www.universiteitleiden.nl/en>) přijala opatření, které se vztahuje ke správě dat a zahrnuje zásady FAIR (viz též předchozí články v rubrice, pozn. red.), již v roce 2016 (samotné zásady byly formulovány o dva roky dříve). Za účelem jeho zavedení do praxe připravila také program, jehož cílem bylo zvýšit povědomí o zásadách FAIR, zavést potřebné služby a vzdělat výzkumné pracovníky v oblasti správy dat. V roce 2019 program vstoupil do druhé fáze, která byla výrazněji orientována na interoperabilitu a opětovnou využitelnost dat a na jejich strojovou zpracovatelnost. Na realizaci programu se podíleli zaměstnanci univerzity, kteří se inspirovali podněty z mezinárodních organizací typu GO FAIR (*GO FAIR Initiative*, <https://www.go-fair.org/go-fair-initiative/>), RDA (*Research Data Alliance*, <https://www.rd-alliance.org/>) nebo CODATA (*Committee on Data*, <https://codata.org/>).

Opatření na Univerzitě v Leidenu se týká tří kroků, které jsou součástí správy dat. Jde o kroky prováděné před výzkumem, v jeho průběhu a po jeho dokončení. V návaznosti na přijetí celouniverzitního opatření byly jednotlivé fakulty požádány o převedení obecných zásad opatření do podoby konkrétních doporučení (pro jednotlivé disciplíny). Za účelem podpory těchto kroků vznikl projektový tým, jehož členy se stali odborníci na správu dat ze Střediska pro digitální vědu (*Centre for Digital Scholarship, CDS*; spadá pod Knihovny Univerzity v Leidenu), poradci pro akademické záležitosti, IT odborníci a odborníci na správu dat z jednotlivých fakult. Mezi realizované aktivity patří např. pravidelné workshopy (konají se každých šest týdnů) zaměřené na tvorbu plánů správy dat.

Na konci roku 2018 se CDS aktivně zapojilo do mezinárodních aktivit směřujících k přípravě řešení pro data podle zásad FAIR. Šlo o činnosti iniciované a společně vedené organizacemi GO FAIR a RDA. V červnu 2020 vznikl tříbodový rámec pro tzv. FAIRifikaci dat (<https://www.go-fair.org/2020/07/08/a-three-point-framework-for-fairification/>). Při vzdělávacích aktivitách se ukázalo jako potřebné vytvořit přehled kompetencí (dovedností a znalostí), které jsou potřebné k provádění činností směřujících k uplatnění zásad FAIR v konkrétních disciplínách. Zástupci Univerzity v Leidenu se proto zapojili do mezinárodních aktivit (např. workshopů), jejichž cílem bylo vytvořit pro vědce z různých oborů právě takové vzdělávací pomůcky (<https://doi.org/10.5281/zenodo.2555497> či <https://osf.io/sjzc8/>). Ty mohou vedle vědců využít i knihovníci usilující o zvýšení povědomí o zásadách FAIR na svých univerzitách.

Univerzita v Leidenu nabízí vědcům kromě školení i individuální konzultace vztahující se ke konkrétním tématům z oblasti správy dat. Většina otázek, na které se hledají odpovědi, se týká návrhu databáze a modelování dat, a také zveřejňování dat v souladu se zásadami FAIR. CDS se také jako partner účastnilo vzdělávacího projektu, jehož cílem bylo zpřístupnění existujícího vědeckého archivu s digitálními dokumenty podle zásad FAIR. Podílelo se rovněž na řešení projektu, jehož cílem bylo v souladu se zásadami FAIR zpřístupnit starší data Střediska pro lingvistiku (*Centre for Linguistics*).

Je nezbytné počítat s tím, že správce dat musí disponovat řadou dovedností sahajících od informačních a komunikačních technologií až po dovednosti umožňující inovovat a efektivně spolupracovat. Žádoucí je se rovněž zapojovat do mezinárodních aktivit vztahujících se ke zpracovávání a zpřístupňování vědeckých dat podle zásad FAIR.

**HRYNASZKIEWICZ, Iain, Natasha SIMONS, Azhar HUSSAIN, Rebecca GRANT a Simon GOUDIE. Developing a Research Data Policy Framework for All Journals and Publishers [Návrh základních požadavků na výzkumná data určený pro všechny časopisy a vydavatele]. *Data Science Journal* [online]. 2020, vol. 19, article 5, s. 1–15 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-005>.**

Stále více vydavatelů a časopisů vyžaduje od autorů článků výzkumná data nebo přinejmenším doporučuje jejich zveřejnění. Jde o důležitý krok k reprodukovatelným vědeckým výsledkům. Požadavky jednotlivých vydavatelů a časopisů se od sebe ale mnohdy značně liší a je obtížné se v nich orientovat. Proto je vyvíjena snaha o jejich standardizaci.

Aby bylo možné navrhnout společný základ, byly nejprve shromážděny a analyzovány požadavky významných vydavatelů (Springer Nature, Elsevier, Wiley, PLOS). Řada z nich, a také některé grantové agentury, např. Evropská komise, se v požadavcích odvolávají na zásady FAIR (*Findability, Accessibility, Interoperability, and Reuse of digital assets*, viz též předchozí články v rubrice, pozn. red.). Vlastní návrh společných základních požadavků vznikl v průběhu dvou let a připravila jej zájmová skupina Sdružení pro výzkumná data (RDA) zaměřená na standardizaci požadavků na data a jejich zavádění do praxe. Již první verze návrhu zahrnovala zpětnou vazbu získanou na zasedáních Sdružení pro výzkumná data.

Výsledkem je soupis čtrnácti prvků uspořádaný do šesti úrovní. Prvky jsou: vymezení výzkumných dat, vymezení výjimek, embarga, doplňkové materiály, úložiště pro data, citace dat, licence pro data, podpora vědců/autorů, prohlášení o dostupnosti dat, formáty dat a standardy pro data, povinné sdílení dat (u specifických publikací), povinné sdílení dat (u všech publikací), recenzování dat a plány správy dat.

Výzkumná data musí být vždy vymezena jako data podporující zjištění nebo tvrzení uvedená ve vydaném článku. Požadavky také musí upřesňovat, kterých dat se týkají (tj. zda jde např. o tabulky, kód, obrázky, zvuk, video, mapy nebo nezpracovaná a/nebo zpracovaná data). Musí být rovněž přesně vymezeno, u kterých dat se veřejné sdílení neočekává. Důležitá je návaznost na proces uchování dat; předpokladem je seznam doporučených důvěryhodných nebo podporovaných úložišť. Může jít o seznam, který si vytvoří časopis či vydavatel samostatně, o seznam, s nímž se obecně pracuje v příslušném vědeckém oboru, nebo důvěryhodný seznam třetí strany, jímž je např. seznam dostupný z *FAIRsharing.org*, či vyhledávač úložišť (např. <https://repositoryfinder.datacite.org/>). Požadavky by měly upřednostňovat oborová úložiště před úložišti obecného rázu. Měly by také stanovit prohlášení o dostupnosti dat a jeho umístění v rukopise. Samozřejmostí by měla být podpora metadatových standardů a formátů používaných v příslušném oboru.

V roce 2019 byla zahájena partnerská spolupráce mezi zájmovou skupinou Sdružení pro výzkumná data, která požadavky připravila, a sdružením akademických vydavatelů STM (*STM Association*, <https://www.stm-assoc.org/>). Cílem je zvýšit míru využití výzkumných dat mezi vydavateli (blíže viz <https://www.stm-researchdata.org/>).

**PLOMP, Esther. Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments [Přechod do digitálního světa: trvalé identifikátory pro výzkumné vzorky, zdroje a přístroje]. *Data Science Journal* [online]. 2020, vol. 19, article 46, s. 1–8 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-046>.**

V posledních letech jsme svědky značných pokroků v oblasti správy výzkumných dat. Díky tomu, že data jsou přidávána k publikacím jako tzv. doplňkové materiály, jsou pro zájemce dostupnější než data dostupná pouze na vyžádání. Data přiložená k publikacím jsou však často k dispozici v nevyhovující podobě (např. ve formátu PDF) nebo odkazy na ně nejsou funkční. Stále častěji jsou tato data zpřístupňována v úložištích, která jsou zřízena přímo k tomuto účelu a splňují zásady FAIR (viz též další články v rubrice, pozn. red.). Podle nich by data měla být naležitelná (*Findable*), přístupná (*Accessible*), interoperabilní (*Interoperable*) a opětovně využitelná (*Reusable*). Zásady se nevztahují pouze k výsledným datům v tradičním pojetí, ale také k algoritmům, nástrojům nebo pracovním postupům, díky nimž data vznikla. Za data jsou pokládány také vzorky (např. vzorky rostlin, minerálů, vody, tkání, mapy či texty v hmotné podobě). Proto se jeví jako potřebné rozšířit zásady FAIR i na fyzické vzorky, artefakty, činidla a analytické přístroje.

Informace vzniklé na základě fyzických vzorků by měly být řádně zdokumentovány a trvale dostupné tak, aby další vědečtí pracovníci mohli tato data získat, ověřit a znovu s nimi pracovat. Shromažďování a správa vzorků a artefaktů je časově velmi náročná, a proto je vhodné tuto činnost řádně oceňovat, např. tak, že vzorky bude možné citovat. Mělo by se to odrazit i na hodnocení vědců, např. při kariérním postupu.

Předpokladem je existence trvalých identifikátorů. Ty se již po dvě desetiletí úspěšně používají pro digitální objekty. Jde např. o identifikátor DOI, který se hojně uplatňuje především u vědeckých článků. V posledních letech se stále více setkáváme i s dalšími typy identifikátorů, např. s identifikátory vědců (mj. ORCID; *ORCID – Open Researchers and Contributor Identifier*, <https://orcid.org/>), vědeckých aktivit (mj. RAiD – *Research Activity Identifier*, <https://www.raid.org.au/>) nebo grantových agentur (mj. FundRef – *Funder Registry*, <https://www.crossref.org/services/funder-registry/>). Trvalé identifikátory umožňují vzájemné propojování dat, jak názorně ukázal např. evropský projekt FREYA (<https://www.project-freya.eu/en>).

Pro hmotné složky výzkumu jsou trvalé identifikátory již nyní používány a propagovány několika iniciativami. Identifikátor IGSN (*International Geo Sample Number*, tj. mezinárodní číslo pro geologické vzorky, <https://www.igsn.org/>) je využíván od roku 2007, a to především v oblasti věd o Zemi. Identifikátor RRID (*Research Resource Identifier*, tj. identifikátor výzkumných zdrojů, <https://scicrunch.org/resources>) byl zaveden v roce 2014 a je využíván ve více než 120 časopisech pro zdroje z biomedicínského výzkumu (činidla, materiály a nástroje). Byl dokonce vyvinut nástroj SciBot, který umožňuje poloautomatizovanou kontrolu správnosti RRID v publikacích (<https://scicrunch.org/resources/about/scibot>). V roce 2015 byly položeny základy Distribuovaného systému vědeckých sbírek (*Distributed System of Scientific Collections*, *DiSSCo*, <https://www.dissco.eu/>), jehož cílem bylo zvýšit soulad dat o biodiverzitě se zásadami FAIR. Od roku 2017 začal experimentálně (zejména v oblasti věd o Zemi) využívat identifikátor PIDINST (*Persistent Identification of Instruments*, tj. trvalý identifikátor přístrojů; <https://www.rd-alliance.org/groups/persistent-identification-instruments-wg>).

Infrastruktura pro zavedení trvalých identifikátorů hmotných složek výzkumu je dostupná, zatím je však využívána především v oblasti věd o Zemi a o životě. Širší využití (mj. řízených slovníků či ontologií) komplikují specifika jednotlivých disciplín. Je proto nutné zpracovat společná základní metadata. Na tom by se měly podílet mj. vědecké společnosti; jsou však vítány i příspěvky „zdola“.

Již nyní existují vědecká pracoviště, která poskytují ke svým fyzickým vzorkům přístup zájemcům zvenčí. Příkladem je NASA zpřístupňující vzorky z povrchu Měsíce.

**SCHWARDMANN, Ulrich. Digital Objects – FAIR Digital Objects: Which Services Are Required? [Digitální objekty – digitální objekty podle zásad FAIR: Které služby jsou třeba?]. *Data Science Journal* [online]. 2020, vol. 19, article 15, s. 1–6 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-015>.**

Největší překážkou pro automatizaci práce s daty je jejich různorodost a komplexnost. Proto je vhodné pracovat s abstrakcemi – se zapouzdřením a virtualizací. Díky zapouzdření se dají skrýt podrobnosti, které na dané úrovni nejsou třeba. Díky virtualizaci mohou být objekty nahrazeny svými zástupci, nejčastěji odkazy (ukazateli) na daný objekt.

To, jak správně vyjádřit logickou strukturu digitálních objektů, je stále předmětem diskusí. Důležitou roli v každém případě hrají odkazy. Nejjednodušší možností jsou URL, nicméně ta jsou mnohdy nestabilní v čase. Proto je nezbytné pracovat s další úrovní odkazů, a to s trvalými identifikátory. U nich by mělo platit, že vždy vedou k aktuálnímu umístění digitálního objektu. V tomto směru se velmi osvědčil a osvědčuje systém Handle (ostatně i identifikátor DOI je implementací tohoto systému).

Rovněž zásady FAIR (*Findability, Accessibility, Interoperability, and Reuse of digital assets*, viz též předchozí články v rubrice, pozn. red.) jsou založeny na vztahu mezi metadaty a daty či samotným digitálním objektem, přičemž klíčovou roli hrají trvalé identifikátory. Zatímco zásady FAIR je možné vnímat jako určitá pravidla, digitální objekty jsou abstrakce v technickém smyslu. Jeví se proto jako vhodné pracovat s digitálními objekty, které naplňují zásady FAIR.

Užitečné je pracovat se systémem Handle, protože umožňuje velmi široké využití a nastavení. To je důležité zvláště při práci s daty, kdy vědci potřebují sdílet velmi podrobné informace (pouhé bibliografické údaje nestačí). Je rovněž nezbytné popisovat objekty v podobě srozumitelné strojům. Základem jsou typy MIME (*Multipurpose Internet Mail Extensions*, víceúčelová rozšíření internetové pošty), ale opět je třeba pracovat s většími podrobnostmi. Do hry tak vstupují datové typy. Při přidávání dalších metadataových prvků je však třeba zároveň brát ohled na skutečnost, že čím více dalších polí bude přidáváno, tím bude celá infrastruktura pomalejší. Základní metadataové prvky, které by se měly používat, doporučila pracovní skupina Sdružení pro výzkumná data (RDA). Navržený profil může být rozšířen (rozšiřování by se mělo řídit stanovenými doporučeními). V každém případě by datové typy měly být určitým způsobem standardizovány, aby bylo dosaženo alespoň minimální interoperability (ta je jednou ze zásad FAIR). Mohly by proto vzniknout spolehlivé registry definic datových typů ve strojově čitelné a interpretované podobě, které by byly opět identifikovány prostřednictvím trvalých identifikátorů. Prototyp takového registru je dostupný v infrastruktuře ePIC (ePIC – *European Persistent Identifiers Consortium*, <http://dtr.pidconsortium.net/>) – jde o implementaci otevřeného softwaru Cordra, který je určen pro správu digitálních objektů. Datové typy je v tomto případě možné definovat hierarchicky.

Aby vše fungovalo, bude třeba vytvořit řadu služeb (nejenom pro zjišťování datových typů podle trvalých identifikátorů; důležitým předpokladem je rovněž existence úložišť, která budou poskytovat spolehlivý přístup k vlastním digitálním objektům). Úložiště pro data jako taková existují, ale velmi zřídka k nahraným datům přidávají trvalé identifiká-

tory. Očekává se, že plnohodnotný přechod na digitální objekty podle zásad FAIR bude vyžadovat značné úsilí a potrvá delší dobu.

**SMITS, Daen Adriaan Ben a Marta TEPEREK. Research Data Management for Master's Students: From Awareness to Action [Správa výzkumných dat u studentů magisterského studia: od povědomí k praktické realizaci]. *Data Science Journal* [online]. 2020, vol. 19, article 30, s. 1–11 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-030>.**

Věda se v současné době potýká s krizí reprodukovatelnosti. V roce 2016 byly publikovány výsledky výzkumu Monye Bakerové, jehož se zúčastnilo 1500 vědců z oborů přírodních věd. Ukázalo se, že v některých disciplínách se až 80 % respondentů setkalo s problémy při ověřování výsledků jiných vědců, s tím, že původní laboratorní data již nebyla dostupná apod. (<https://doi.org/10.1038/533452a>). Na to zareagovaly grantové agentury, které po vědcích začaly vyžadovat předkládání plánů správy dat hned v počátcích výzkumných projektů, ale i časopisy, vlády jednotlivých zemí a konkrétní výzkumné instituce. Zavedená pravidla jsou však funkční pouze v případě, že nejsou příliš vzdálena od běžné vědecké práce a jsou v souladu s tím, co sami vědci pokládají za dobrou praxi při správě dat.

Výzkum více než 700 respondentů na Technické univerzitě v Delftu, který proběhl v letech 2017 a 2018 (<https://doi.org/10.18352/lq-10287>), ale doložil, že vědci nejsou příliš obeznámeni se zásadami FAIR (*Findability, Accessibility, Interoperability, and Re-use of digital assets*, viz též předchozí články v rubrice, pozn. red.), podle nichž by se efektivní správa dat měla řídit. Výzkum byl zaměřen na vědecké pracovníky, nikoliv na studenty na magisterském stupni, přestože je zřejmé, že schopnost správné práce s daty je i pro tuto skupinu velmi důležitá.

Byl proto připraven další výzkum, jehož cílem bylo zjistit postoje studentů magisterského stupně ke správě dat. Proběhl v září až říjnu 2019 formou polostrukturovaných rozhovorů se šestnácti absolventy nizozemských univerzit (všichni ukončili studium po roce 2015). Při magisterském studiu se klade poměrně velký důraz na získání dovedností potřebných pro vědeckou práci; všichni účastníci výzkumu rovněž zpracovávali diplomovou práci a získali tedy praktickou zkušenost s realizací výzkumu.

Všech šestnáct účastníků odpovídalo na jedenáct otázek rozdělených do tří skupin – na zkušenosti se správou dat, pozornost věnovanou správě dat v rámci studijního programu a na data v dnešním světě. Úplná zpracovaná data z rozhovorů jsou k dispozici ke stažení z úložiště 4TU.ResearchData (<https://doi.org/10.4121/uuid:cb3bce67-64f8-4def-a34c-9001e86109d1>).

Bylo zjištěno, že během svého výzkumu s daty pracovali všichni studenti. Pět respondentů naznačilo, že používalo pouze kvalitativní metody, devět jich pracovalo pouze s metodami kvantitativními a dva z nich používali oba typy metod. Konkrétně např. analyzovali existující datové sady v programu SPSS (*Statistical Package for the Social Sciences*), sbírali data o pacientech, realizovali rozhovory, analyzovali dokumenty nebo prováděli dotazníková šetření.

Respondenti byli nejprve požádáni o vymezení pojmu správa výzkumných dat (*research data management*). Někteří dobře zachytili vybrané aspekty pojmu, ale řada z nich význam posunula směrem k metodologii nebo návrhu výzkumu. Když respondenti sdíleli své zkušenosti se správou dat, většina z nich zmiňovala především otázky ochrany soukromí (ochranu soukromí neuvědlnil pouze jeden respondent). Ochranu soukromí znovu zmiňovali v souvislosti s úlohou dat v dnešním světě. K různým problémům při práci s daty přistupovali respondenti spíše intuitivně, ať už s využitím rad vedoucího diplomové práce nebo bez nich. Pouze tři respondenti měli vytvořen určitý plán správy

dat; ostatní postupovali bez něj. Polovina respondentů už nevěděla, jak se ke svým vlastním výzkumným datům dostat; jejich výzkum je proto nereprodukovatelný. Ani jeden ze zúčastněných data nepokládá za důležitý vědecký výstup nebo za informaci podstatnou pro vědecký svět. Žádný z respondentů svá data nezveřejnil. Ačkoliv tak sami prakticky nečinili, respondenti jednomyslně pokládali dobré postupy při správě dat během výzkumu za podstatné pro jeho budoucí opakovatelnost. V rozhovorech byla zmíněna také rizika plynoucí ze zveřejnění výzkumných dat – možné zneužití citlivých dat nebo skutečnost, že uživatelé dat by nemuseli disponovat dostatečnými dovednostmi k jejich správné interpretaci.

Jako studenti neprošli žádným speciálním školením o správě dat; dílčí aspekty byly pojednány v kurzech zaměřených na výzkumné metody nebo např. v diplomových seminářích. Většina respondentů však vyjádřila zájem se o správě dat dozvědět více podrobností.

Lze shrnout, že až na otázku ochrany soukromí respondenti nedisponovali příliš hlubokými znalostmi o správě dat. Jeví se proto jako žádoucí, aby se ve studijních programech univerzit tomuto tématu věnovala větší pozornost (bez ohledu na konkrétní obor). Na Technické univerzitě v Delftu již vstup správce dat z jedné fakulty zařadili do výuky. Tamtéž funguje uskupení DelftOpenHardware (<https://delftopenhardware.nl/>). Do jeho činnosti se zapojují i studenti a tím se učí dobře zpracovávat dokumentaci a seznamují se s tím, že je potřebné sdílet i návrhy hardwaru. Na univerzitě je dále k dispozici online portál pro vytváření plánů správy dat (<https://dmponline.tudelft.nl/>). Ten je zaměřen především na výzkumné pracovníky a doktorandy; tato studie však ukázala, že by bylo vhodné myslet i na potřeby studentů magisterského stupně.

**STOCKER, Markus, Louise DARROCH, Rolf KRAHL, Ted HABERMANN, Anusuriya DEVARAJU, Ulrich SCHWARDMANN, Claudio D'ONOFRIO a Ingemar HÄGGSTRÖM. Persistent Identification of Instruments [Trvalá identifikace přístrojů]. *Data Science Journal* [online]. 2020, vol. 19, article 18, s. 1–12 [cit. 2021-03-15]. Dostupné z: <http://doi.org/10.5334/dsj-2020-018>.**

Při vytváření vědeckých dat hrají velmi důležitou úlohu přístroje. Přístroje jsou konkrétně míněny měřicí přístroje, tedy zařízení využívaná pro měření, ať již samostatně nebo společně s jedním či více doplňkovými zařízeními. Mohou být statické (např. meteostanice) nebo mobilní (např. drony). Mohou být využívány k pozorování nebo provádění experimentů a mohou je vlastnit a pracovat s nimi jednotliví vědci, výzkumné skupiny, národní, mezinárodní nebo globální výzkumné infrastruktury nebo další typy organizací.

Ukazuje se, že k tomu, aby bylo možné interpretovat digitální datovou sadu, je třeba znát řadu podrobností o zařízení použitém k získání dat. Do hry tak vstupují metadata. Pracovní skupina PIDINST (*Persistent Identification of Instruments*), součást Sdružení pro výzkumná data (*Research Data Alliance, RDA*), si proto dala za cíl shromáždit konkrétní případy užití, identifikovat společná metadata, navrhnout a zveřejnit metadataové schéma a zpracovat zpětnou vazbu k jeho jednotlivým verzím, podporovat zavedení schématu do stávajícího systému trvalých identifikátorů, schéma prakticky vyzkoušet, vzbudit zájem o něj u ostatních členů RDA a každé dva týdny pořádat virtuální setkání členů pracovní skupiny.

Pracovní skupina shromáždila celkem 15 případů užití, z nichž ty, u nichž dali autoři své svolení, jsou zveřejněny na GitHubu (<https://github.com/rdawg-pidinst/use-cases>). Většina z nich se vztahuje k vědám o Zemi. V případech užití bylo celkem zjištěno 43 různorodých vlastností. Při porovnávání byly seskupeny do deseti kategorií: identifikace, přístroj, model, vlastník, výrobce, datum, způsobilost, výstup, související přístroj,

vydavatel. U více než poloviny případů se vyskytly vlastnosti trvalý identifikátor, název přístroje, popis přístroje, typ přístroje, vlastník přístroje, výrobce a datum.

Řešení vychází ze stávající infrastruktury a je navrženo tak, aby se s ním dalo snadno pracovat. Jeho součástí je trvalý identifikátor a metadatové schéma obsahující hlavní vlastnosti umožňující přesnou identifikaci přístroje a jeho uvedení do potřebných souvislostí. Metadatové schéma svým pojetím navazuje na obecně uznávanou dobrou praxi a vhodně ji doplňuje. Řada vlastností je koncipována tak, že umožňuje vkládat hodnoty v podobě volně formulovaného textu nebo oborových slovníků. Výsledkem jsou sice data, u nichž nelze mít vysoká očekávání, pokud jde o interoperabilitu, ale je to nutný ústupek při práci s heterogenními daty. Důležitou vlastností schématu je také skutečnost, že prostřednictvím souvisejícího identifikátoru lze upozornit např. na časopisecké články, datové sady nebo jiné objekty, které se k přístroji váží.

Ačkoliv je již doložena praktická použitelnost schématu, zatím nebylo do detailu finalizováno. Na GitHubu je zveřejněno jako živý dokument (<https://github.com/rdawg-pidinst/schema>).

Po zveřejnění schématu byly zahájeny diskuse směřující k zavedení identifikace v infrastrukturách ePIC (<https://www.pidconsortium.net/>) a DataCite (<https://datacite.org/>). Spolupráce vyústila v zavedení schématu PIDINST do infrastruktury ePIC v podobě prototypu. Společné vlastnosti byly také přiřazeny k vlastnostem ze schématu DataCite. Ačkoliv většinu vlastností bylo možné přiřadit, u některých (např. u názvu modelu) tomu tak nebylo. Proto bylo navrženo příslušné obohacení schématu DataCite (zprístupněného na platformě GitHub).

Se schématem pracuje nebo se chystá pracovat Jednotný systém pro sledování uhlíku (*Integrated Carbon Observation System, ICOS*), který je zaměřen na měření skleníkových plynů v Evropě, systém PANGAEA (<https://pangaea.de/>), jenž archivuje a zveřejňuje datové sady z oblasti věd o Zemi a environmentálních studií, a vznikající systém EISCAT3D (<https://eiscat.se/eiscat3d-information/>), který je orientován na studium atmosféry a prostoru v blízkosti Země a související výzkum.

Zpracovala: PhDr. Linda Jansová, Ph.D.

Redakčně připravila: PhDr. Anna Machová

JANSOVÁ, Linda. Novinky zahraniční knihovnické literatury. *Knihovna: knihovnická revue*. 2020, **32**(1), 90–100. ISSN 1801-3252.