

OTVORENÉ VÝSKUMNÉ DÁTA A POŽIADAVKY NA NE Z POHĽADU OTVORENEJ VEDY

Mgr. Michaela Mikušková

Medzi kľúčové prvky taxonómie otvorenej vedy patria aj otvorené výskumné dáta. Požiadavky na otvorenosť dát, ktoré vznikli pri realizácii projektov financovaných z verejných prostriedkov, sa stávajú stabilnou súčasťou požiadaviek grantových agentúr. Podpora sprístupňovania výskumných dát a doplňujúcich materiálov je súčasťou Národnej stratégie pre otvorenú vedu na roky 2021 – 2028, ktorá nadväzuje na iniciatívy Európskej únie v tejto oblasti.

V súčasnom digitálnom veku sú nielen informácie, ale aj dáta veľmi dôležitou surovinou nevyhnutnou pre nové objavy a urýchlenie pokroku vo výskume. Na vedu a výskum sa vynakladajú nemalé finančné prostriedky a objem vyprodukovaných dát každý rok prudko narastá. Výskumné dáta sú však do značnej miery izolované (a často neusporiadané) v osobných zariadeniach vedcov a prístup k nim je blokovaný rôznymi technickými, právnymi a finančnými obmedzeniami. Aby sa financovatelia výskumu a grantové inštitúcie vyhli opakovanému vynakladaniu finančných prostriedkov na rovnaký výskum, vznikli iniciatívy na otvorené sprístupňovanie dát, minimálne tých, ktoré sú podkladom vydaných publikácií.

Ako možno charakterizovať otvorené dáta? **Otvorené dáta** sú údaje, ktoré môže ktokoľvek slobodne, bezplatne a opakovane využívať a šíriť pod podmienkou uvedenia autorstva a zachovania prístupových práv na ďalšie použitie [1]. Dáta sa zvyčajne vzťahujú na celý rad netextových materiálov a súborov údajov obsahujúcich laboratórne výskumy, testy a pozorovania, štatistické údaje, výsledky prieskumov, modely a pod. Otvorené dáta nie sú len doménou vedeckovýskumného procesu. Na Slovensku je dobre vybudovaná infraštruktúra pre otvorené dáta vo verejnej správe (portál <https://data.gov.sk>).

SPARC [2] – nezisková organizácia, ktorá podporuje otvorené systémy výskumu a vzdelávania – definuje **otvorené výskumné dáta** ako dáta, ktoré:

- sú voľne dostupné na internete,
- používateľ môže sťahovať, kopírovať, analyzovať, opätovne spracovávať a používať na akýkoľvek iný účel,
- sú dostupné bez finančných, právnych alebo technických prekážok okrem tých, ktoré sú neoddeliteľnou súčasťou prístupu k samotnému internetu.

Prečo sú výskumné dáta a ich sprístupňovanie také dôležité?

Výskumné dáta sú primárnym výstupom vedeckovýskumného procesu. Ich sprístupnenie na opätovné využívanie umožňuje overiť alebo reprodukovať výsledky výskumu na základe pôvodných údajov. Zvyšuje sa tým hodnota počítačovej investície, pretože výskumníci môžu stavať na predchádzajúcich výsledkoch bez toho, aby museli vynakladať finančné prostriedky na reprodukciu údajov. Otvorené dáta posilňujú integritu vedeckých a odborných záznamov a prispievajú k vyššej kontrole výstupov. Zároveň pomáhajú vytvárať obraz o pokroku výskumu v príslušnej vednej oblasti, čím urýchľujú tempo objavovania a pomáhajú zvyšovať dôveru verejnosti vo výsledky vedy a výskumu.

Princípy FAIR

Dáta musia byť kvalitné, presne a správne popísané a štruktúrované. S dátami je úzko spojený termín **FAIR**. Ide o akronym štyroch anglických slov **F**indable (nájditelný), **A**ccessible (dostupný), **I**nteroperable (interoperabilný) a **R**eusable (opakovane použiteľný), ktorých definícia bola prvýkrát uverejnená v roku 2016 v časopise *Scientific Data*. Tieto slová popisujú základné vlastnosti vedeckých dát a s nimi súvisiaceho manažmentu dát počas výskumného procesu. Princípy FAIR rešpektujú etické, obchodné a súkromné obmedzenia v prípade citlivých údajov alebo údajov, ktoré sú predmetom vlastníctva. Nestanovujú kritériá výlučne pre otvorené sprístupňovanie dát a mali by sa dodržiavať aj v prípade, že dáta nie sú verejne dostupné.

Na využívanie dát treba, aby boli strojovo čitateľné, pochopiteľné a vyhľadateľné pre človeka i počítač. Na popis dát sa používajú **metadáta**, ktoré možno charakterizovať ako dáta o dátach. Slúžia na popis dátových súborov, ich charakteristík, autorstva, obsahu, typu informácií a dát v nich uložených, kontextu ich vzniku a pod. Práve metadáta sú kľúčový prvok na opätovné využívanie dát.

Zásady FAIR sa vzťahujú na tri typy entít: údaje (alebo akýkoľvek digitálny objekt), metaúdaje (informácie o tomto digitálnom objekte) a infraštruktúru. Princípy FAIR [3] požiadavky na ne popisujú nasledovne:

Nájditelnosť – je základný prvok otvorených výskumných dát:

- (meta)dátam je priradený jedinečný trvalý identifikátor,
- dáta sú popísané metadátami,
- metadáta jasne špecifikujú identifikátor údajov, ktoré opisujú,
- metadáta sú registrované alebo indexované v zdroji, v ktorom sa dá vyhľadávať.

Prístupnosť – jasne definuje, kto a ako môže k dátam pristupovať, overovať si ich a autorizovať:

- (meta)dáta možno vyhľadať podľa ich identifikátora pomocou štandardizovaného komunikačného protokolu (http). Tento protokol:
 - je otvorený, bezplatný a univerzálne implementovateľný,
 - ak treba, umožňuje postup autentifikácie a autorizácie,

- metadáta by mali byť prístupné aj v čase, keď už nie sú k dispozícii dáta, ktoré popisujú (napríklad z dôvodu nečitateľnosti dát).

Interoperabilita – je z hľadiska využívania dát veľmi dôležitá. Dáta sa musia dať integrovať s inými údajmi, zároveň musia vedieť spolupracovať s aplikáciami na analýzu a spracovanie dát. (Meta)dáta by preto mali:

- používať formálny, prístupný a široko použiteľný jazyk na reprezentáciu znalostí,
- používať slovníky, ktoré dodržiavajú zásady FAIR,
- obsahovať odkazy na iné (meta)dáta.

Opakovaná použiteľnosť – cieľom FAIR princípov je optimalizácia opakovaného používania dát, ich replikovateľnosť a kombinovateľnosť prostredníctvom presnej dokumentácie:

- (meta)dáta by mali byť bohato opísané pomocou presných a relevantných atribútov,
 - (meta)dáta treba zverejniť za jasných licenčných podmienok, ktoré upravujú práva a povinnosti na používanie dát,
 - (meta)dáta uvádzajú svoj pôvod (autorstvo), ktorý umožňuje ich citovanie,
 - (meta)dáta spĺňajú štandardy relevantné pre danú oblasť výskumu.

Otvorené vedecké dáta spĺňajúce princípy FAIR patria medzi osem hlavných princípov Politiky otvorenej vedy Európskej únie [4].

Plán manažmentu dát

Nevyhnutnou súčasťou všetkých fáz vedeckého výskumu je plán manažmentu dát – Data Management Plan (DMP). Už pred začiatkom výskumného procesu si vedecí pracovníci musia odpovedať na niekoľko kľúčových otázok, ktoré ovplyvnia proces zberu, analýzy a spracovania, dlhodobého uchovávanía, zverejňovania, šírenia a opätovného použitia dát. Optimalizácia práce prostredníctvom DMP zjednodušuje prácu s dátami počas výskumného procesu a pomáha pri riešení strategických rozhodnutí, ktoré počas neho vznikajú. Mnohé výskumné inštitúcie a grantové agentúry vyžadujú vypracovanie DMP, prípadne ponúkajú interné šablóny na jeho tvorbu. K dispozícii sú aj rôzne nástroje na tvorbu DMP [5]:

- OpenAIRE – Argos: <https://argos.openaire.eu/splash/>,
- DCC (Digital Curation Centre, UK) DMP online: <https://dcc.ac.uk/dmponline>,
- DMPtool: <https://dmptool.org/>.

Dátové repozitáre

Najlepší spôsob na publikovanie dát s otvoreným prístupom je dátový repozitár. V súčasnosti je na výber niekoľko interdisciplinárnych i špecializovaných úložísk. Pri výbere repozitára treba zohľadniť kľúčové záujmy v troch oblastiach [6]:

1. Cieľová skupina

Dáta by mali byť prístupné a ľahko nájditel'né pre ľudí (alebo stroje), ktorí ich budú s najväčšou pravdepodobnosťou používať, tzn. určiť cieľovú skupinu. Ide prioritne o výskumníkov z vašej oblasti výskumu alebo aj mimo nej? Pri výbere úložiska je dôležité zohľadniť kritériá na zverejnenie dát, ktoré požaduje finančná agentúra podporujúca výskum.

2. Dlhodobá dostupnosť

Budú vedieť potenciálni používatelia ľahko nájsť dáta aj v repozitári špecializovanom pre vašu oblasť výskumu? Umožňuje repozitár uložiť dáta v jasnej štruktúre a v konzistentnom formáte, ktoré zabezpečia jeho strojovú čitateľnosť? Umožňuje repozitár obmedziť prístup k citlivým údajom?

3. Dodržiavanie predpisov

Pred výberom úložiska sa uistite, že ste skontrolovali nasledujúce skutočnosti:

- *Požiadavky svojho zamestnávateľa* – buduje inštitucionálny dátový repozitár, má vypracovaný plán manažmentu dát, prípadne interné usmernenia na zdieľanie výskumných dát?
- *Požiadavky poskytovateľov financovania výskumu* – aké požiadavky na sprístupňovanie dát má poskytovateľ financií na váš výskum?
- *Požiadavky časopisov* – aj keď váš zamestnávateľ alebo grantová agentúra nežiadujú sprístupňovanie výskumných dát, môže to žiadať časopis, v ktorom budete výsledky výskumu publikovať.

Pri výbere dátového úložiska treba zohľadniť aj spôsob zabezpečenia dlhodobého uchovávanía dátových súborov, použité certifikáty a štandardy, ponuku štatistík o využívaní sprístupnených dát a návodu, ako dáta citovať a i.

Registre obsahujúce dátové repozitáre [7]:

- <https://www.re3data.org/>,
- <https://v2.sherpa.ac.uk/opusdoar/about.html>,
- <https://fairsharing.org/>.

Problematika otvorených výskumných dát je veľmi široká a stále prináša nové výzvy. Plán manažmentu dát sa stáva nevyhnutnou súčasťou vedeckej práce a kladie zvýšené požiadavky na poznatky aj v tejto oblasti. Výskumné inštitúcie a financovatelia výskumu sú kľúčoví aktéri pri vytváraní podporných nástrojov pre vedeckovýskumných pracovníkov a pri budovaní kvalitnej infraštruktúry, ktorá zabezpečí správu, dlhodobé uchovávanie, dostupnosť

a interoperabilitu dát. V zahraničí sú jedným z kľúčových aktérov v tejto oblasti akademické knižnice, ktoré si našli uplatnenie vo vzdelávaní a poskytovaní služieb pri správe vedeckých a výskumných dát.

Použité zdroje:

[1] What is Open Data. Dietrich Daniel et al. *Open Data Handbook* [online]. Open Knowledge Foundation [cit. 2023-01-07].

Dostupné na: <http://opendatahandbook.org/guide/en/what-is-open-data/>

[2] Open Data. In *Sparc* [online]. Washington: SPARC, 2007-2022 [cit. 2023-01-07]. Dostupné na:

<https://sparcopen.org/open-data/>

[3] FAIR principles. In *Go Fair* [online]. GFISCO [cit. 2023-01-07]. Dostupné na: <https://www.go-fair.org/fair-principles/>

- [4] Open Science [online]. European Commission [cit. 2023-01-07]. Dostupné na: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en
- [5] STOŽICKÁ Zuzana. Archívacia a manažment výskumných dát [online]. Bratislava: Centrum vedecko-technických informácií SR, 2022 [cit. 2023-01-07]. Dostupné na: <https://doi.org/10.5281/zenodo.6320659>
- [6] How to Store and Manage Your Data [online]. San Francisco: Plos [cit. 2023-01-07]. Dostupné na: <https://plos.org/resource/how-to-store-and-manage-your-data/>
- [7] DOBBERSTEINOVÁ, Jitka. Zverejňovanie vedeckých dát z hľadiska vydavateľa [online]. Bratislava: Centrum vedecko-technických informácií SR, 2022 [cit. 2023-01-07]. Dostupné na: <https://doi.org/10.5281/zenodo.6420263>