

Podpora otevřených dat, sdílení dat a datové články

LIBOR ANSORGE

Výzkumný ústav vodohospodářský T. G. Masaryka jakožto vydavatel časopisu Vodohospodářské technicko-ekonomické informace (VTEI) a zároveň veřejná výzkumná organizace v rezortu Ministerstva životního prostředí dlouhodobě podporuje otevřený přístup k datům a sdílení datových sad pro využití dalšími uživateli. Již od roku 2001 jsou na internetu dostupné datové sady vodního hospodářství zahrnuté do Hydro-ekologického informačního systému (HEIS), který VÚV vyvíjí již od 90. let [1]. Naplňuje tak Národní strategii otevřeného přístupu ČR k vědeckým informacím na léta 2017–2020, kterou vláda České republiky schválila 14. června 2017 [2] a jež navazuje na už dříve deklarované potřeby a výhody sdílení vědeckých a odborných dat. Na úrovni EU byla přijata Směrnice 2019/1024 ze dne 20. června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru, která nahradila starší Směrnici 2003/98/ES. Směrnice 2019/1024 nově v čl. 10 upravuje (mimo jiné) otevřenost výzkumných dat pro opakované použití dalšími uživateli. Základní myšlenkou pro sdílení výzkumných dat je, že jejich zpřístupnění zvyšuje věrohodnost provedeného výzkumu. Navíc výzkumná data mohou být užitečná nejen pro řešitele, jenž data získal, ale též pro další vědce, což umožní opatřit si výsledky v jiném kontextu a pohledem někoho jiného. I proto byl v České republice v rámci hodnocení výsledků výzkumu zaveden výsledek typu S – veřejně dostupná datová sada.

Standardní vědecké publikace jsou zaměřeny primárně na analýzu a šíření výsledků výzkumu. Na druhou stranu vytvoření datové sady s výzkumnými daty spotřebuje často značné finanční, materiální i personální zdroje. Je důležité, aby autoři vzniklých datových sad s výzkumnými daty získali příslušné uznání za práci vynaloženou na pořízení datové sady [3]. Toho lze dosáhnout např. citací článku popisujícího datovou sadu, což vedlo ke vzniku nového typu článku „data paper“ [4]. Tyto články nejsou primárně zaměřeny na výsledky výzkumů, ale na popis datových sad, které v rámci výzkumu vznikly. Hlavním cílem těchto článků je usnadnit reprodukci a znovupoužití těchto datových sad, jež jsou publikovány pod některou ze svobodných licencí (např. Creative Commons). A o tento typ odborných článků se nyní rozšiřuje i časopis VTEI. Tím se připojuje k mnoha vědeckým nakladatelstvím, která publikují datové články, nebo dokonce vydávají specializované časopisy zaměřené na tento typ článků, jako např. Scientific data (ISSN 2052-4463) vydavatelství Nature Publishing Group, NCHS data brief (ISSN 1941-4927), který vydává U. S. National Center for Health Statistics, Data in Brief (ISSN 2352-3409) vydavatelství Elsevier BV, Geoscience Data Journal (ISSN 2049-6060) vydavatelství John Wiley and Sons Inc., nebo Data (ISSN 2306-5729) vydavatelství MDPI AG. Jedním z nejstarších časopisů je Journal of Chemical & Engineering Data (ISSN 0021-9568) nakladatelství ACS Publications, vycházející již od roku 1956.

Samozřejmě i datové články musejí splňovat určitá kritéria, aby je bylo možno považovat za výsledky vědecké činnosti, a naplňovat podmínky Národní strategie otevřeného přístupu ČR k vědeckým informacím. Jedním ze základních kritérií je splnění podmínek pro FAIR data [5], tj. dostupnost, přístupnost, interoperabilitu a opětovné použití dat. Na základě praxe jiných časopisů byly stanoveny následující podmínky pro datové články:

1. Časopis VTEI nezajišťuje ukládání datových sad. Datové sady musejí být hostovány v důvěryhodném datovém úložišti a musí být zajištěn stálý přístup k popisované datové sadě.
2. Datové úložiště, ve kterém jsou data uložena, by mělo nabídnout zaručené uchování a podle principů FAIR (Findable, Accessible, Interoperable and Re-usable) by mělo zajistit dostupnost, přístupnost, interoperabilitu a opětovné použití dat (vhodný repozitář lze nalézt v Mezinárodním registru repozitářů pro výzkumná data <https://www.re3data.org> nebo v databázi Open access repozitářů <https://v2.sherpa.ac.uk/pendoar/>; univerzálním repozitářem je např. ZENODO – <https://www.zenodo.org>).
3. Publikovaná datová sada musí být původní sadou informací, jež jsou získávány vědeckou metodou a mají hodnotu pro výzkumnou či odbornou komunitu.
4. Datová sada musí být důkladně popsána a také být přesná, opakovaně použitelná, reprodukovatelná a replikovatelná.
5. Datová sada musí být veřejně dostupná s licencí umožňující volné použití.
6. Pokud existují nezpracovaná data (označovaná také jako primární, zdrojová nebo nezpracovaná data) vztahující se k jakýmkoli tabulkám, grafům či obrázkům v rukopisu, musí být tato data veřejně dostupná, a to buď s datovým článkem (např. jako doplňkový soubor), nebo hostovaná na důvěryhodném úložišti dat (preferováno).
7. Všechny protokoly nebo metody používané k získávání údajů musejí být popsány v článku a podléhají recenznímu řízení. Tyto protokoly a metody musejí splňovat etické principy.
8. U údajů, jež z etických důvodů vyžadují kontrolu přístupu, musí být poskytnut mechanismus, aby redaktoři a recenzenti mohli přistupovat k těmto údajům bez odhalení jejich totožnosti autorům.
9. Sekundární data jsou výjimečně povolena, pokud jsou upravena pomocí postupů, díky nimž je datová sada jedinečná, autoři poskytují vědecký protokol pro sběr/vytváření dat, sběr sekundárních dat nebo analýza přidává těmto datům významnou hodnotu. Jsou-li popisována sekundární data, musí být poskytnut seznam použitých primárních zdrojů dat a veřejně zpřístupněna celá sekundární datová sada, a to buď s datovým článkem (např. jako doplňkový soubor), nebo hostovaná v důvěryhodném datovém úložišti.

Články o recenzi nebo doplňkové soubory z článku o recenzi se nepovažují za původní data a obvykle nejsou pro zveřejnění vhodné. Ve snaze usnadnit autorům tvorbu datových článků byla připravena jeho šablona, která je k dispozici v sekci Pokyn autorům coby soubor VTEI Data Article Template 1.2.docx.

Jednou z diskutovaných otázek ohledně datových článků je, zda a jak tyto články recenzovat. Recenze (Peer review) může zvýšit důvěru ve vědecké údaje i výsledky a umožnit hodnocení a certifikaci datových souborů z hlediska kvality [6]. Proto i datové články publikované ve VTEI procházejí recenzním řízením stejně jako odborné články.

V tomto čísle VTEI na str. 38 vychází první z datových článků. Věříme, že datové články budou motivující pro tvůrce dat, aby autoři důkladně ověřovali, dokumentovali, kontrolovali a šířili svou práci v souladu s principy FAIR.

Literatura

[1] PICEK, J. HEIS VÚV v roce 2006. *Vodohospodářské technicko-ekonomické informace* [on-line]. 2006, roč. 48, č. 3, s. 14–17. ISSN 0322-8916. Dostupné z: <https://www.vtei.cz/archiv/>

[2] VLÁDA ČR. *Národní strategie otevřeného přístupu ČR k vědeckým informacím na léta 2017–2020*. Praha: Úřad vlády České republiky. 2017

[3] WALTERS, W. H. Data journals: incentivizing data access and documentation within the scholarly communication system. *Insights* [on-line]. 2020, roč. 33, č. 1, s. 18. ISSN 2048-7754. Dostupné z: [doi:10.1629/uksg.510](https://doi.org/10.1629/uksg.510)

[4] KUNZE, J. A., P. CRUSE, R. HU, S. ABRAMS, K. HASTINGS, C. MITCHELL a L. R. SCHIFF. Practices, Trends, and Recommendations in Technical Appendix Usage for Selected Data-Intensive Disciplines [on-line]. 2011. Dostupné z: <https://escholarship.org/uc/item/9jw4964t>

[5] WILKINSON, M. D., M. DUMONTIER, I. J. AALBERSBERG, G. APPLETON, M. AXTON, A. BAAK, N. BLOMBERG, J.-W. BOITEN, L. B. DA SILVA SANTOS, P. E. BOURNE, J. BOUWMAN, A. J. BROOKES, T. CLARK, M. CROSAS, I. DILLO, O. DUMON, S. EDMUNDS, C. T. EVELO, R. FINKERS, A. GONZALEZ-BELTRAN, A. J. G. GRAY, P. GROTH, C. GOBLE, J. S. GRETHE, J. HERINGA, P. A. C. T. HOEN, R. HOOFT, T. KUHN, R. KOK, J. KOK, S. J. LUSHER, M. E. MARTONE, A. MONS, A. L. PACKER, B. PERSSON, P. ROCCA-SERRA, M. ROOS, R. VAN SCHAIK, S.-A. SANSONE, E. SCHULTES, T. SENGSTAG, T. SLATER, G. STRAWN, M. A. SWERTZ, M. THOMPSON, J. VAN DER LEI, E. VAN MULLIGEN, J. VELTEROP, A. WAAGMEESTER, P. WITTENBURG, K. WOLSTENCROFT, J. ZHAO a B. MONS. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* [on-line]. 2016, roč. 3, č. 1, s. 160018. ISSN 2052-4463. Dostupné z: [doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

[6] MAYERNIK, M. S., S. CALLAGHAN, R. LEIGH, J. TEDDS a S. WORLEY. Peer review of datasets: When, why, and how. *Bulletin of the American Meteorological Society* [on-line]. 2015, roč. 96, č. 2, s. 191–201 [vid. 30. květen 2021]. ISSN 0003-0007, 1520-0477. Dostupné z: [doi:10.1175/BAMS-D-13-00083.1](https://doi.org/10.1175/BAMS-D-13-00083.1)

Autor

Ing. Libor Ansoorge, Ph.D.
ORCID: 0000-0003-3963-8290