
Repozitáře a datové časopisy

Obdobně jako v případě vědeckých publikací, je několik možností, jak můžete svá data zveřejnit:

1. Jako doplňující materiály (Supplementary Materials) k vědeckému článku
2. Prostřednictvím repozitáře
3. V datovém časopise

Doplňující materiály k vědeckému časopisu

Některá vydavatelství nabízí možnost přiložit k článku **doplňující materiály** (Supplementary Materials). Autoři takto mohou k článku přiložit například podkladová data nebo skript, který použili ke zpracování dat, a nakladatel tyto materiály následně publikuje online a k vašemu článku přidá na tyto materiály odkaz. Rozhodnete-li pro tuto možnost, mějte na paměti, že pokud se nejedná o open access publikaci, pak v takovém případě většinou **majetková práva přechází na vydavatele**. Další nevýhodou tohoto modelu je, že čtenáři nebudou moci data dále využít a nebude možné dataset citovat nezávisle na publikaci.

Datové repozitáře

Nejvhodnějším způsobem, jak uchovat svá data – ať už se rozhodnete je otevřít či nikoli – je jejich uložení do **datového repozitáře**.

Abyste zvýšili pravděpodobnost, že vědci z oboru vaše data najdou, je vhodné uložit svá data do **oborového repozitáře**. Výhodou oborových repozitářů je, že sdružují výzkumné pracovníky konkrétního vědního oboru, kteří mezi sebou svá data sdílejí, a také, že se snáze přizpůsobí potřebám daného oboru, například pokud se jedná o typ dat, která se v daném oboru produkují. Příkladem oborového repozitáře je [LINDAT/CLARIAH-CZ](#) pro lingvistická data a nástroje, který je provozovaný na Ústavu formální a aplikované lingvistiky na MFF UK.

Při výběru vhodného oborového repozitáře je dobré dbát oborových zvyklostí – můžete se zeptat svých kolegů, kam data ukládají, nebo využít **mezinárodní registr datových repozitářů** [re3data.org](#) (Registry of Research Data Repositories).

Pokud se vám nepodaří najít vhodný oborový repozitář, můžete využít obecné repozitáře (také tzv. sirotčí), které ukládají data bez ohledu na oborové zaměření. Mezi nejpoužívanější obecné repozitáře patří [Zenodo](#), [Figshare](#) nebo [Dryad](#). Pro porovnání obecných repozitářů můžete využít také [tuto srovnávací tabulku](#).

Při výběru vhodného repozitáře si také zjistěte následující:

- Přidělí repozitář vašim datům **perzistentní identifikátor** (např. DOI)? Díky perzistentním identifikátorům jsou vaše data snáze vyhledatelná a citovatelná.
- Je repozitář **důvěryhodný/certifikovaný**? U certifikovaných repozitářů máte větší jistotu, že je o vaše data dobře postaráno.
- Poskytuje vybraný repozitář **otevřený přístup** k uloženým datům? Pokud chcete svá data sdílet otevřeně, pak je toto klíčová informace.
- Opatří repozitář vaše data **licencí**? Uvádí jasné podmínky, za jakých lze data uložená v repozitáři využívat? Je důležité, aby uživatelé vašich dat věděli, jak s nimi mohou nakládat.
- Poskytne repozitář pro vaše data **úvodní stránku s metadaty**? Metadata pomohou ostatním data najít, zjistit, co je to za data, a také jak je citovat.
- Umožňuje repozitář **verzování**? Pokud svůj dataset aktualizujete, můžete jej nahrát jako novou verzi k původnímu datasetu. Nový dataset dostane přidělený svůj vlastní identifikátor a uživatelé snadno zjistí, jaká je nejnovější verze nebo jaká verze byla použita v původní studii.

Na portálu [re3data.org](#) můžete některé z těchto informací snadno ověřit. U každého záznamu najdete v pravém horním rohu sérii piktogramů, které vám prozradí například, zda-li repozitář používá perzistentní identifikátory, je otevřený, nebo zda-li je certifikovaný. Bližší informace pak najdete po rozkliknutí záznamu nebo přímo na stránkách daného repozitáře.

LINDAT/CLARIN repository

Subject(s) Humanities Linguistics Artificial Intelligence, Image and Language Processing Humanities and Social Sciences Computer Science
Computer Science, Electrical and System Engineering Engineering Sciences

Content type(s) Standard office documents Audiovisual data Plain text Structured text Archived data Source code

Country Czech Republic European Union

LINDAT/CLARIN is designed as a Czech "node" of Clarin ERIC (Common Language Resources and Technology Infrastructure). It also supports the goals of the META-NET language technology network. Both networks aim at collection, annotation, development and free sharing of language data and basic technologies between institutions and individuals both in science and in all types of research. The Clarin ERIC infrastructural project is more focused on humanities, while META-NET aims at the development of language technologies and applications. The data stored in the repository are already being used in scientific publications in the Czech Republic.

Datové časopisy

Datové časopisy **odráží model klasického vědeckého publikování** a podporují publikování datových článků (data papers), které lze definovat jako "vědecké publikace s prohledatelnými metadaty, které popisují určitý online dostupný dataset nebo jejich skupinu, a které jsou publikovány v souladu se standardními akademickými postupy" (definice [Chavan & Penev, 2011](#)). Některé časopisy publikují pouze datové články (různí vydavatelé mohou tento typ článků označovat různými názvy, např. database article nebo data descriptor), jiné jsou smíšené a publikují jak datové články, tak tradiční výzkumné články. Datový článek je analogický s tradičními vědeckými publikacemi, lze jej citovat a také jej můžete vykazat v OBD (systém pro vykazování tvůrčí činnosti na UK).

Různé datové časopisy mohou mít různé politiky, ale obecným cílem je, aby datový článek **popisoval určitý dataset**, poskytoval informace o tom, co je to za data, kdy, kde a jakým způsobem vznikla atd., a také by měl obsahovat **odkaz na daný dataset** (nejlépe prostřednictvím trvalého identifikátoru jako je DOI). Datové články by neměly obsahovat vědecké analýzy, které byly na daném datasetu provedeny, pouze jejich popis. Obecně by vydavatelé neměli data samotná uchovávat, ta by měla být **uložena v důvěryhodném otevřeném datovém repozitáři** tak, aby data byla k dispozici i tehdy, je-li k článku omezený přístup. Většina datových časopisů také provádí **recenzní řízení**, ale mohou se lišit tím, co od recenzentů vyžadují a zda je recenzní řízení otevřené či nikoli.

Níže naleznete několik příkladů datových časopisů, další můžete najít např. [zde](#).

- [Scientific data](#) - přírodní vědy
- [Earth System Science Data](#) - vědy o Zemi
- [Journal of Open Archaeology Data](#) - archeologie
- [Biodiversity Data Journal](#) - biodiverzita
- [GigaScience](#) - biomedicína
- [Journal of Open Research Software](#) - výzkumný software

Užitečné zdroje:

Chavan, V., Penev, L. 2011. The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12, S2. <https://doi.org/10.1186/1471-2105-12-S15-S2>

Gould, Julie. 2014. How to publish your data in a data journal. Naturejobs Blog. <http://blogs.nature.com/naturejobs/2014/12/04/how-to-publish-your-data-in-a-data-journal/>

Stall, Shelley, Maryann E. Martone, et al. 2020. Generalist Repository Comparison Chart. Zenodo. <http://doi.org/10.5281/zenodo.3946720>