Geographic data citations: case study of DIBAVOD — systematic review

LIBOR ANSORGE, ANNA ANSORGEOVÁ

ABSTRACT

Society, mainly through state and public institutions, spends considerable funds on the acquisition, management and sharing of data acquired with public funds, including scientific data. Evaluating the success of an open data policy is very problematic. One possibility is to use the citations of these datasets to track the use of open data. Dataset citation is a relatively new field and still faces a number of methodological and technical challenges, including little awareness in the scientific community of the positives of dataset citation. Also problematic is the low level of skill in citing datasets, which generally leads to different forms and ways of citations. In this study, an analysis was performed of the citations of the geographic database DIBAVOD, which is managed by T. G. Masaryk Water Research Institute. In total, 122 citing documents were included in the study. The study showed that the forms and methods of citation vary widely and do not show any discernible trends over time. Only the number of citations shows a slightly increasing trend. Almost a quarter of the papers then only mentioned the use of DIBAVOD without indicating the source of this data or citing it in another form.

INTRODUCTION

The issue of data citation

In today's digital era, data – including geographic data – play a key role in science. Both government and scientific institutions invest significant resources in dataset creation, management, and access. As the digital environment has developed and the size of datasets increased in recent decades, the cost of this data has also steadily increased. Monitoring the use of datasets within the research community allows us to verify the value of the resources invested in creating datasets, their administration, and providing them to the public and other users.

Current public policies encourage or even stipulate that publicly funded datasets be shared and used for other types of analysis. Costello [1] has mapped a number of positive aspects of sharing research data, as well as the concerns and arguments of scientists who oppose it. The willingness to share research data varies across disciplines, and in addition to data repositories, personal websites of individual scientists are still widely used [2]. Zhao et al. [3] analysed 600 papers published in the journal PLoS One and concluded that scientists still prefer to create their own sets rather than analyse already collected data. In VTEI journal, one can also find very few articles that are based on the reuse of already published data [4, 5].

The requirement to make research data available in the Czech Republic is enshrined in Act No. 130/2002 Coll., on the support of research and development

from public funds and on amendments to certain related acts (the Act on the Support of Research and Development). Act No. 130/2002 Coll. introduced in Section 12a the obligation to provide research data, including research data that is an annex to or part of scientific publications, free of charge upon request. The data have to be provided no later than one year after the end of the project public funding. Information on acquired research data is recorded through the Research & Development Information System (R&D IS). The goal is that the considerable resources provided for the acquisition and management of research data are demonstrably spent "for the public good", i.e. so that other scientists can deal with other scientific tasks using already acquired datasets, whether based on individual datasets or combinations of multiple data sources [6]. Of course, Act No. 130/2002 Coll. also introduces certain exceptions that regulate when research data does not have to be provided.

Citations are a way of appreciating the work of cited authors in the scientific community; in the contemporary world they are used as a tool for evaluating science, which often serves for career advancement and as a basis for allocating funding for science. However, data citation is not intended to replace citation of relevant literature, but rather to provide verifiable and reusable information about the availability of research data that support published conclusions and claims. The lack of proper citation of datasets makes peer-reviewed publications less transparent, jeopardizes reproducibility, and hinders open science [7].

Citation of the dataset used is also necessary to comply with Act No. 121/2000 Coll., on copyright, rights related to copyright, and on amendments to certain acts (Copyright Act). Datasets fall under copyright works. According to Section 31 of the Copyright Act, the use of a copyrighted work for scientific purposes is permitted only "if possible, the name of the author, unless the work is anonymous, or the name of the person under whose name the work is made public, as well as the title of the work and the source are stated".

Data citation is therefore an important tool for acknowledging the work of data creators and curators and allowing them to track how their data is used. Data citation allows scientists and other users to easily find data that have been used in a particular piece of research, and to replicate that research and verify its results. Without proper data citation, it would be difficult to achieve the goals of open science, which seeks to share data and scientific knowledge to accelerate scientific progress. Finally, data citation helps to ensure that data sharing is fair and that the creators and curators of datasets receive due credit for their work [8].

Citing datasets faces a number of issues [9], such as the uniqueness and verifiability of the citation, i.e. how to cite datasets so that the citation allows for the precise identification of the dataset used and so that it can be verified that the data cited were used. Another issue is how to cite dynamic datasets that change over time, or whether to cite the dataset or the article that describes the dataset. Citing an article that describes the dataset contributes

to the author's H-index, which can be beneficial for their scientific career. Conversely, citing a dataset, even if relevant to research, usually does not directly affect the author's H-index. This disparity in impact on academic metrics may influence decisions about the recommended form of citation. Last but not least, the scientific community is addressing the question of how we can track and evaluate the use of datasets.

To address these issues, a number of standards and best practices for data citation have been developed. The most well-known are the Data Citation Principles, developed by FORCE11 [10]. Adherence to the proposed standards contributes to increasing the impact of both the cited and citing work [11]. However, the basic task of a data citation system is to guarantee the permanence of the cited data and the citations themselves [12], i.e., to ensure that the cited dataset remains available in the cited form in the future.

In recent years, a number of studies have been conducted to examine how datasets are cited. Gregory et al. [13] examined the practices, preferences, and motivations for citing data; they distinguish three types of dataset citation. The first type is data citation in the references. This means that datasets are cited like any other bibliometric source, with an abbreviated citation in the text of the article and a full citation in the reference list. This form of citation allows for easy tracking of citations using citation analysis tools and specialized citation services. The second type of data citation is a simple mention of the data used in the text of the publication. The last type is an indirect citation, where the reference to the data is given in the form of a citation of another related publication (e.g. a data article describing the data or a data document).

Smith et al. [14] point out another issue with citation of datasets, using the *Paleobiology Database* as an example. The problem is that collective works, such as large datasets composed of contributions from many authors (and articles based on them), are cited more often than the original data contributors to these large datasets.

Digital water management database

The *Fundamental Base of Geographic Data of the Czech Republic* (ZABAGED) is the primary geographic data set in the Czech Republic. The administrator of ZABAGED is the Land Survey Office, which administers and expands it in the public interest in accordance with Act No. 200/1994 Coll. The financing of the ZABAGED administration is thus ensured from the Czech state budget. In addition to ZABAGED, there are other geographic datasets. In the field of water management, this is mainly the *Digital Base of Water Management Data* (DIBAVOD). DIBAVOD is managed by the T. G. Masaryk Water Research Institute (TGM WRI) public research institution and its administration is ensured from the internal resources of this institution.

DIBAVOD is a reference geographic database created primarily from the corresponding ZABAGED layers. It is used to create thematic cartographic outputs in the field of water management and water protection over the base map of the Czech Republic 1 : 10,000. DIBAVOD is used, for example, for spatial analyses in the geographic information systems (GIS) environment and for processing reporting data under the Water Framework Directive 2000/60/EC in the field of water policy.

DIBAVOD can be characterized as a dynamic database containing 75 different objects that describe water management elements for the creation of basic water management maps. The objects are divided into ten focus groups:

A. Basic phenomena of surface and groundwater

B. Focus classification of surface and groundwater

C. Protected areas

D. Floodplains

E. Surface water gauging and monitoring points

F. Groundwater gauging and monitoring points

G. Water use subsystem objects

H. Abstraction and discharge points

I. Objects in streams

J. Meteorological observation objects

As part of sharing individual objects with the general public or interested parties, a web map application was created on the dibavod.cz website. This application is a synoptic interactive tool for publishing data and services with online access via a web browser. It can contain raster and vector datasets and allows the use of analytical and publication tools.

Currently, the system-wide stable financing of DIBAVOD is not ensured [15], which leads to some objects being unavailable or not updated for a long time. Information on the use of DIBAVOD is therefore an important aspect when deciding on further financing of DIBAVOD administration. TGM WRI does not have detailed information on the use of this database by the scientific community as the DIBAVOD data can be downloaded for free from the dibavod.cz portal. The aim of this study is therefore to map the citation rate of DIBAVOD and analyse the types of citations of this dataset. On the main page of the dibavod.cz application, the DIBAVOD authors themselves recommend citing the DIBAVOD dataset in the form of an indirect citation of the article *GIS and Cartography at the TGM WRI* [15], published in 2022 in the VTEI journal.

DATA AND METHODS

A systematic literature review was chosen as the primary research method for this study. A systematic literature review is a specific type of review that focuses on finding an answer to a pre-formulated research question by analysing the proof collected in the literature search [16]. The fundamental difference compared to so-called "narrative" literature reviews is the limitation of subjectivity through clearly defined rules for selecting and including literature in the review [17, 18]. Systematic literature reviews use the PRISMA methodology [19]. Due to their complexity, systematic literature reviews are suitable for cases where several dozen or a few hundred contributions are analysed.

The bibliometric databases Dimensions.AI [20], Scopus [21], and Web of Science [22] were selected for citation analysis. Data collection was carried out via the web interface of all three databases. Data collection was carried out on 7 March 2024 by searching for the string DIBAVOD in all fields and then repeated on 1 July 2024. A total of 216 scientific publications were found in the Dimensions.AI database, their metadata were exported in csv format and loaded into a spreadsheet. A total of 47 scientific publications were found in the Scopus database, which were again exported in csv format and loaded into a spreadsheet. Three articles were found in the Web of Science – Core collection database, and when the query was expanded to all databases in the Web of Science, five references were found to two datasets derived from DIBAVOD. Records for these datasets were not included in the analysis.

In the first step, duplicates were eliminated, resulting in a list of 231 scientific publications for screening. As part of the screening, each document found was checked to see if it actually contained a DIBAVOD citation. 104 records that did not cite DIBAVOD were excluded from further analysis, as were five records

for which it was not possible to verify whether they cited DIBAVOD (e.g. due to the unavailability of the paper for the authors).

The citation analysis included 122 papers citing DIBAVOD. A modified typology described by Gregory et al. [13] was used to monitor the types of citations. For each paper citing DIBAVOD, the form of citation and the method of citing the source were checked. The form of citation was classified into one of two categories – "citation in the text" or "citation in the list of references". In the case of the form of citation in the references, categories of the method of citing the source were created: "no source is cited", "a recommended article is cited" (i.e. the article *GIS and cartography at TGM WRI* [15] is cited), "TGM WRI is cited", "the website dibavod.cz is cited"). Based on the analysis of citations, a new category "TGM WRI Hydroecological Information System is cited" (alias HEIS TGM WRI) was added. HEIS TGM WRI is another information system operated by TGM WRI, which provides attribute data on water management in the Czech Republic. In the case of the "citation in text" form, the same categories of source citation methods were chosen, but logically, "a recommended article is cited" cannot appear in this DIBAVOD citation form. Citation analysis was performed by both authors of this study; the second author was in charge of the initial analyses, the first author checked the results and made decisions in the case of unclear classifications.

Subsequently, these data were statistically processed and the content analysed.

RESULTS AND DISCUSSION

Citations according to individual categories of citation form and method of citing the source are shown in *Fig. 1*. A total of 122 papers citing DIBAVOD were found. In DIBAVOD citations, citations in the form of links in the list of references slightly predominate; there are 64 (i.e. 52.5 %). This can be considered a good result, since in-text citations generally prevail over citations in the list of references [23]. However, the presented results are difficult to generalize because the number of citing articles is low. As Rogers et al. [24] point out, samples of 1,000 documents provide a good guide for relative (but not absolute) citation analyses; studies with fewer than 200 documents suffer from high variability in results.

Citing in the form of a link in the list of references is preferable from the perspective of dataset curators, as it allows for easier tracking of the use of the dataset using specialized bibliometric databases. However, this does not mean that citations of the dataset in the text of the citing document should be considered inappropriate. Data citation is still in its infancy and authors are still learning how to use it. It is therefore important that citing the datasets used becomes part of general "citation skills" and good publishing practices.

A somewhat unpleasant finding is the 29 papers (23.8 %) that only mention DIBAVOD in the text without providing any acknowledgement to the authors and curators of DIBAVOD, or a reference to this source. Although 23.8 % may not seem like a high percentage, it is still a manifestation of ignorance or non-compliance with citation rules. Inaccurate, incomplete or careless citation, where it is not possible to identify the cited source, is considered by most publication ethics manuals [e.g. 25] to be a violation of publication ethics, or plagiarism. Moreover, it is a de facto violation of Czech copyright law, which requires the author and source to be cited in addition to the title of the paper used.

In this context, the question arises whether currently not citing the exact source can be considered a violation of publication ethics if the (non-)citing work indicates that a specific dataset was used that can be easily found on the Internet. With regard to compliance with the FAIR principles [8], the citation of datasets should also contribute to finding the dataset used, its accessibility, interoperability and reusability. Citations of data play an important role in ensuring their findability and accessibility, especially when persistent

identifiers such as DOI are used in the citations. Groth et al. [26] discuss the benefits of citing datasets for their reuse. In our view, more frequent data citations will also have an impact on their interoperability, as data with higher interoperability should be used and cited more. Recognizing the importance of data citations, for example by including data citations in rating systems, will put pressure on data curators to ensure greater interoperability of the datasets they manage.

A total of 54 papers referenced the dibavod.cz website, which was the most common way of citing a source in the DIBAVOD dataset. Of these, 46 references to dibavod.cz were in the form of a link in the list of references, while eight references to dibavod.cz were listed directly in the text of the citing article. The descriptive article [15], which is recommended to be cited by the curators of the DIBAVOD database on the dibavod.cz website, was cited only five times, which is a very small number. This may be due to the fact that the article was written relatively recently (in 2022), and also to the fact that data articles are not yet widely used for citing datasets, but there is still a steady increase in citations of data articles [27]. However, the overall citation of datasets is still at a very low level, regardless of the data repository from which the data is uploaded [28].

Two articles cited DIBAVOD as part of the HEIS TGM WRI. In both cases, these were relatively old citations, the first from 2009, the second from 2021, but citing a source from 1965. Two other articles cited the HEIS TGM WRI; however, because both of these articles also cited TGM WRI or dibavod.cz, they were included in the categories citing these sources.

The last way of citing a source is represented by citations stating that DIBAVOD is managed by the TGM WRI. There were 32 such citations in total, 21 of which were in the form of in-text citations and 11 in the form of citations in the references. This method of citation cannot be considered optimal; however, it at least acknowledges the TGM WRI for the management of the DIBAVOD dataset.

Fig. 2 shows that a certain increasing trend can be seen in the total number of citations, but not in whether the share of citations in the form of in-text citations and citations in the references is changing. Similarly, *Fig. 3* shows that the way of citing the source does not indicate any noticeable trend either, and the individual categories are randomly represented in individual years. This suggests that dataset citation is not yet widely practiced in the Czech scientific community. However, education in the field of dataset citation is essential for supporting academic integrity, developing critical digital skills, and improving the ethical and effective use of data.
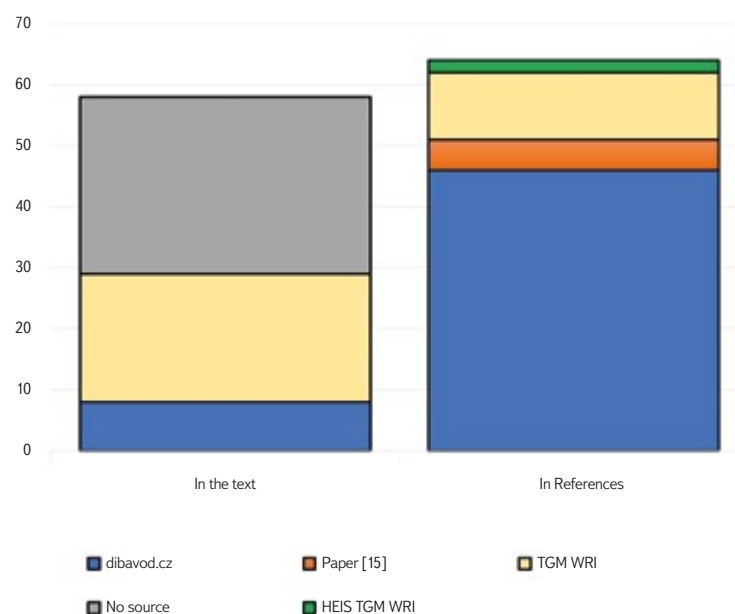

Fig. 1. Structure of DIBAVOD citation types

In this study, we focused on citations in scientific journals that are included in citation databases, which allowed for relatively simple data processing. However, geographic datasets such as DIBAVOD or ZABAGED are created primarily with the aim of providing data for the lay public. One of the challenges for systems for assessing the reuse of research data is the way in which these data are used by the general professional public outside the academic sphere. This includes in particular citations in the so-called grey literature, in strategic documents and policies, in decisions of administrative bodies, etc. However, these citation analyses are highly demanding because, unlike scientific publication production, there are no easily usable sources of information for these types of documents. The easiest way to do this is to use web search engines such as Google. However, the subsequent analysis of the search results is very difficult to automate. Citation of datasets may not be the only way to demonstrate the use of research data. Other options include download counts, usage agreements, etc. The biggest complication here is again the lack of readily available information on these types of indicators.
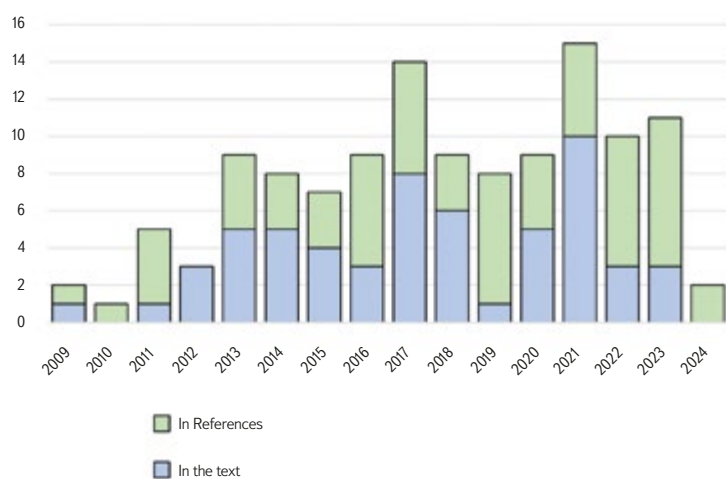

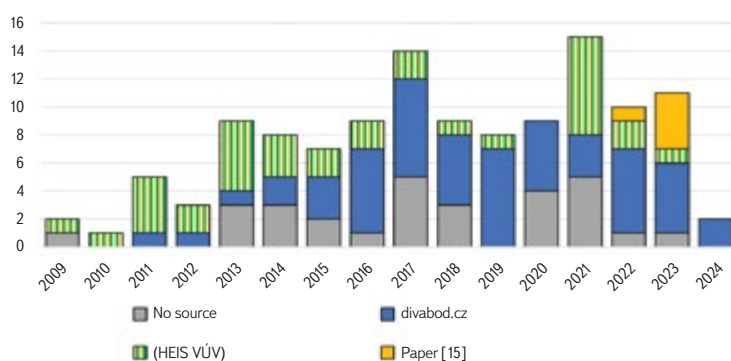
Fig. 2. Development of DIBAVOD citations over time

Fig. 3. Distribution of citation types over time

CONCLUSIONS

The analysis showed that citing the DIBAVOD dataset cannot be considered optimal. Of the 122 papers analysed, 58 only used the reference in the text and, of these, only 29 mentioned the use of DIBAVOD without more detailed information about the dataset or its authors or originator. A total of 54 papers provided a link to the dibavod.cz website and 34 papers cited DIBAVOD in the form

of a reference to the originator, i.e. TGM WRI. Only five papers used the recommended citation via the article *GIS and Cartography in TGM WRI*; however, this may be mainly due to the fact that this recommended article is quite new. The study thus demonstrated that citing water management datasets, such as DIBAVOD in particular, is not widespread in the Czech Republic, and there is no established form and method of citing these datasets. The importance of citing geographic data should therefore be emphasized both within university study programmes and through public events and professional committees. Similarly, the study demonstrated high heterogeneity in the form of citations of the DIBAVOD dataset. Much more awareness-raising is needed in this regard as well.

All data used in the study can be obtained from Dimension.AI, Scopus, and Web of Science databases using the procedures described in this study. The source file in MS Excel format in which all analyses were performed is available upon request from the corresponding author.

Acknowledgements

Conflict of interest declaration

The corresponding author is part of the TGM WRI management, which publishes the VTEI journal, and the chairman of the VTEI journal Editorial Board. However, these facts had no influence on the results of the presented study. TGM WRI did not provide any funds for the preparation of this study.

References

[1] COSTELLO, M. J. Motivating Online Publication of Data. *BioScience* [on-line]. 2009, 59(5), pp. 418–427. ISSN 0006-3568. Available at: https://doi.org/10.1525/bio.2009.59.5.9

[2] KHAN, N., THELWALL, M., KOUSHA, K. Data Sharing and Reuse Practices: Disciplinary Differences and Improvements Needed. *On-line Information Review* [on-line]. 2023, 47(6), pp. 1 036–1 064. ISSN 1468-4527. Available at: https://doi.org/10.1108/OIR-08-2021-0423

[3] ZHAO, M., YAN, E., LI, K. Data Set Mentions and Citations: A Content Analysis of Full-Text Publications. *Journal of the Association for Information Science and Technology* [on-line]. 2018, 69(1), pp. 32–46. ISSN 2330-1643. Available at: https://doi.org/10.1002/asi.23919

[4] ANSORGE, L., STEJSKALOVÁ, L., SOLDÁN, P. Nové znečišťující látky v odpadních vodách – výsledky Společného průzkumu Dunaje 4 pohledem šedé vodní stopy. *Vodohospodářské technicko- -ekonomické informace* [on-line]. 2024, 66(1), pp. 38–45. ISSN 0322–8916, 1805-6555. Available at: https://doi.org/10.46555/VTEI.2023.11.002

[5] ANSORGE, L., STEJSKALOVÁ, L. Srovnání šedé vodní stopy způsobené běžným znečištěním a mikropolutanty: případová studie ČOV Bandung v Indonésii. *Vodohospodářské technicko-ekonomické informace* [on-line]. 2024, 66(4), pp. 20–27. ISSN 0322–8916, 1805-6555. Available at: https://doi.org/10.46555/VTEI.2023.11.00210.46555/VTEI.2024.05.001

[6] BORGMAN, C. L. The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology* [on-line]. 2012, 63(6), pp. 1 059–1 078. ISSN 1532-2890. Available at: https://doi.org/10.1002/asi.22634

[7] GERASIMOV, I., SAVTCHENKO, A., ALFRED, J., ACKER, J., WEI, J., BINITA, K. C. Bridging the Gap: Enhancing Prominence and Provenance of NASA Datasets in Research Publications. *Data Science Journal* [on-line]. 2024, 23(1). ISSN 1683-1470. Available at: https://doi.org/10.5334/dsj-2024-001

[8] WILKINSON, M. D., DUMONTIER, M., AALBERSBERG, Ij. J., APPLETON, G., AXTON, M., BAAK, A., BLOMBERG, N., BOITEN, J.-W., DA SILVA SANTOS, L. B., BOURNE, P. E., BOUWMAN, J., BROOKES, A. J., CLARK, T., CROSAS, M., DILLO, I., DUMON, O., EDMUNDS, S., EVELO, C. T., FINKERS, R., GONZALEZ- -BELTRAN, A., GRAY, A. J. G., GROTH, P., GOBLE, C., GRETHE, J. S., HERINGA, J., 'T HOEN, P. A. C., HOOFT, R., KUHN, T., KOK, R., KOK, J., LUSHER, S. J., MARTONE, M. E., MONS, A., PACKER, A. L., PERSSON, B., ROCCA- -SERRA, P., ROOS, M., VAN SCHAIK, R., SANSONE, S.-A., SCHULTES, E., SENGSTAG, T., SLATER, T., STRAWN, G., SWERTZ, M. A., THOMPSON, M., VAN DER LEI, J., VAN MULLIGEN, E., VELTEROP, J., WAAGMEESTER, A., WITTENBURG, P., WOLSTENCROFT, K., ZHAO, J., MONS, B. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* [on-line]. 2016, 3(1), 160018. ISSN 2052-4463. Available at: https://doi.org/10.1038/sdata.2016.18

[9] PARSONS, M. A., DUERR, R. E., JONES, M. B. The History and Future of Data Citation in Practice. *Data Science Journal* [on-line]. 2019, 18(1), pp. 52. ISSN 1683-1470. Available at: https://doi.org/10.5334/dsj-2019-052

[10] DATA CITATION SYNTHESIS GROUP. *Joint Declaration of Data Citation Principles* [on-line]. San Diego CA: FORCE11, 2014. Available at: https://doi.org/10.25490/A97F-EGYK

[11] COUSIJN, H., FEENEY, P., LOWENBERG, D., PRESANI, E., SIMONS, N. Bringing Citations and Usage Metrics Together to Make Data Count. *Data Science Journal* [on-line]. 2019, 18(1), 9. ISSN 1683-1470. Available at: https://doi.org/10.5334/dsj-2019-009

[12] SILVELLO, G. Theory and Practice of Data Citation. *Journal of the Association for Information Science and Technology* [on-line]. 2018, 69(1), pp. 6–20. ISSN 2330-1643. Available at: https://doi.org/10.1002/asi.23917

[13] GREGORY, K., NINKOV, A., RIPP, C., ROBLIN, E., PETERS, I., HAUSTEIN, S. Tracing Data: A Survey Investigating Disciplinary Differences in Data Citation. *Quantitative Science Studies* [on-line]. 2023, 4(3), pp. 622–649. ISSN 2641-3337. Available at: https://doi.org/10.1162/qss_a_00264

[14] SMITH, J. A., RAJA, N. B., CLEMENTS, T., DIMITRIJEVIĆ, D., DOWDING, E. M., DUNNE, E. M., GEE, B. M., GODOY, P. L., LOMBARDI, E. M., MULVEY, L. P. A., NÄTSCHER, P. S., REDDIN, C. J., SHIRLEY, B., WARNOCK, R. C. M., KOCSIS, Á. T. Increasing the Equitability of Data Citation in Paleontology: Capacity Building for the Big Data Future. *Paleobiology* [on-line]. 2024, 50(2), pp. 165–176. ISSN 0094-8373, 1938-5331. Available at: https://doi.org/10.1017/pab.2023.33

[15] FOJTÍK, T., JAŠÍKOVÁ, L., KURFIŘTOVÁ, J., MAKOVCOVÁ, M., MAŤAŠOVSKÁ, V., MAYER, P., NOVÁKOVÁ, H., ZAVŘELOVÁ, J., ZBOŘIL, A. GIS a kartografie ve VÚV TGM. *Vodohospodářské technicko-ekonomické informace* [on-line]. 2022, 64(1), pp. 47–52. ISSN 0322–8916, 1805-6555. Available at: https://www.vtei.cz/2022/02/gis-a-kartografie-ve-vuv-tgm/

[16] SNYDER, H. Literature Review As a Research Methodology: An Overview and Guidelines. *Journal of Business Research* [on-line]. 2019, 104, pp. 333–339. ISSN 0148-2963. Available at: https://doi.org/10.1016/j.jbusres.2019.07.039

[17] HODGKINSON, G. P., FORD, J. K. Narrative, Meta-Analytic, and Systematic Reviews: What Are the Differences and Why Do They Matter? *Journal of Organizational Behavior* [on-line]. 2014, 35(S1), pp. S1–S5. ISSN 1099-1379. Available at: https://doi.org/10.1002/job.1918

[18] TRANFIELD, D., DENYER, D., SMART, P. Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management* [on-line]. 2003, 14(3), pp. 207–222. ISSN 1467-8551. Available at: https://doi.org/10.1111/1467-8551.00375

[19] PAGE, M. J., MCKENZIE, J. E., BOSSUYT, P. M., BOUTRON, I., HOFFMANN, T. C., MULROW, C. D., SHAMSEER, L., TETZLAFF, J. M., AKL, E. A., BRENNAN, S. E., CHOU, R., GLANVILLE, J., GRIMSHAW, J. M., HRÓBJARTSSON, A., LALU, M. M., LI, T., LODER, E. W., MAYO-WILSON, E., MCDONALD, S., MCGUINNESS, L. A., STEWART, L. A., THOMAS, J., TRICCO, A. C., WELCH, V. A., WHITING, P., MOHER, D. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* [on-line]. 2021, 372, n71. ISSN 1756-1833. Available at: https://doi.org/10.1136/bmj.n71

[20] THELWALL, M. Dimensions: A Competitor to Scopus and the Web of Science? *Journal of Informetrics* [on-line]. 2018, 12(2), pp. 430–435. ISSN 1751-1577. Available at: https://doi.org/10.1016/j.joi.2018.03.006

[21] BAAS, J., SCHOTTEN, M., PLUME, A., CÔTÉ, G., KARIMI, R. Scopus As a Curated, High-Quality Bibliometric Data Source for Academic Research in Quantitative Science Studies. *Quantitative Science Studies* [on-line]. 2020, 1(1), pp. 377–386. ISSN 2641-3337. Available at: https://doi.org/10.1162/qss_a_00019

[22] BIRKLE, C., PENDLEBURY, D. A., SCHNELL, J., ADAMS, J. Web of Science As a Data Source for Research on Scientific and Scholarly Activity. *Quantitative Science Studies* [on-line]. 2020, 1(1), pp. 363–376. ISSN 2641-3337. Available at: https://doi.org/10.1162/qss_a_00018

[23] PARK, H., YOU, S., WOLFRAM, D. Informal Data Citation for Data Sharing and Reuse is More Common than Formal Data Citation in Biomedical Fields. *Journal of the Association for Information Science and Technology* [on-line]. 2018, 69(11), pp. 1 346–1 354. ISSN 2330-1643. Available at: https://doi.org/10.1002/asi.24049

[24] ROGERS, G., SZOMSZOR, M., ADAMS, J. Sample Size in Bibliometric Analysis. *Scientometrics* [on-line]. 2020, 125(1), pp. 777–794. ISSN 1588-2861. Available at: https://doi.org/10.1007/s11192-020-03647-7

[25] FOLTÝNEK, T., ČERNIKOVSKÝ, P., FONTANA, J., GOJNÁ, Z., HENEK DLABOLOVÁ, D., HOLEČEK, T., HRADECKÝ, J., KOZMANOVÁ, I., MACH, J., PRUSEK, O., ŘÍMANOVÁ, R., TESAŘÍKOVÁ, K., VOREL, F., VORLOVÁ, H. *Jak předcházet plagiátorství ve studentských pracích: příručka pro akademické pracovníky*. Praha: Karolinum – Charles University Press, 2020. ISBN 978-80-246-4786-9.

[26] GROTH, P., COUSIJN, H., CLARK, T., GOBLE, C. FAIR Data Reuse – the Path through Data Citation. *Data Intelligence* [on-line]. 2020, 2(1–2), pp. 78–86. ISSN 2641-435X. Available at: https://doi.org/10.1162/dint_a_00030

[27] JIAO, H., QIU, Y., MA, X., YANG, B. Dissemination Effect of Data Papers on Scientific Datasets. *Journal of the Association for Information Science and Technology* [on-line]. 2024, 75(2), pp. 115–131. ISSN 2330-1635, 2330-1643. Available at: https://doi.org/10.1002/asi.24843

[28] NUMAJIRI, H., HAYASHI, T. Analysis on Open Data As a Foundation for Data-Driven Research. *Scientometrics* [on-line]. 2024, 129, pp. 6 315–6 332. ISSN 1588-2861. Available at: https://doi.org/10.1007/s11192-024-04956-x

Authors

Ing. Libor Ansorge, Ph.D.[1]
✉ libor.ansorge@vuv.cz
ORCID 0000-0003-3963-8290

Anna Ansorgeová[2]
✉ st102751@students.ujep.cz

[1]T. G. Masaryk Water Research Institute, Prague (Czech Republic)
[2]Faculty of Science, Jan Evangelista Purkyně University in Ústí nad Labem (Czech Republic)