

Citování geografických datových sad: pilotní studie DIBAVOD – systematický přehled literatury

LIBOR ANSORGE, ANNA ANSORGEOVÁ

Klíčová slova: DIBAVOD – citovanost dat – GIS – geografická databáze

ABSTRAKT

Společnost, prostřednictvím zejména státních a veřejných institucí, vynakládá značné finanční prostředky na pořizování, správu a sdílení dat financovaných z veřejných prostředků, a to i výzkumných dat. Hodnocení úspěšnosti politiky otevřených dat je velmi problematické. Jednou z možností je využít citovanost těchto datových sad pro sledování využívání otevřených dat. Citace datových sad je relativně novou záležitostí a stále se potýká s řadou metodických a technických problémů, včetně malého povědomí vědecké komunity o pozitivních aspektech citování datových sad. Problematická je také nízká úroveň dovedností v citování datových sad, což vede obecně k různým formám a způsobům zápisu citací. V této studii byla provedena analýza citovanosti geografické databáze DIBAVOD, kterou spravuje Výzkumný ústav vodohospodářský T. G. Masaryka (VÚV TGM). Do studie bylo zahrnuto 122 citujících dokumentů. Studie ukázala, že formy a způsoby citování se velmi liší a nenaznačují v průběhu času nějaké patrné trendy. Pouze počet citací vykazuje určitý mírně rostoucí trend. Skoro čtvrtina prací pak uvedla použití DIBAVOD, aniž by udala zdroj těchto dat nebo je citovala jinou formou.

ÚVOD

Problematika citování dat

V dnešní digitální éře hrají data – včetně geografických dat – klíčovou roli v oblasti vědy. Státní i vědecké instituce investují nemalé prostředky do tvorby, správy a zpřístupnění datových sad. S rozvojem digitálního prostředí a s růstem velikosti datových sad v uplynulých desetiletích neustále rostou náklady na tato data. Sledování, jak byly datové sady využívány v rámci výzkumné komunity, umožňuje ověřit smysluplnost prostředků vložených do tvorby datových sad, jejich správy i zpřístupnění veřejnosti a dalším uživatelům.

Současné veřejné politiky podporují či přímo vyžadují, aby datové sady vznikající s veřejnou finanční podporou byly sdíleny a využívány pro další typy analýz. Costello [1] zmapoval řadu pozitiv při sdílení výzkumných dat i obav a argumentů vědců, proč se tomu brání. Ochota sdílet výzkumná data se liší podle jednotlivých oborů a kromě datových úložišť jsou dosud hojně využívány osobní webové stránky jednotlivých vědců [2]. Zhao et al. [3] analyzovali 600 dokumentů publikovaných v časopise PLoS One a došli k závěru, že vědci stále raději vytvářejí vlastní sady, než aby analyzovali již pořízená data. V časopise VTEI lze nalézt také jen velmi málo článků, které jsou postaveny na znovuvyužití již publikovaných dat [4, 5].

Požadavek na zpřístupnění výzkumných dat je v České republice (ČR) zakotven v zákoně č. 130/2002 Sb., o podpoře výzkumu a vývoje z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje), který v § 12a zavedl povinnost bezplatně poskytnout na základě žádosti výzkumná data včetně výzkumných dat, jež jsou přílohou k vědeckým publikacím nebo jejich součástí, a to nejpozději jeden rok po skončení financování projektu z veřejných prostředků. Informace o pořizovaných výzkumných datech jsou evidovány prostřednictvím Informačního systému výzkumu, vývoje a inovací (IS VaVaI). Cílem je, aby nemalé prostředky poskytované na pořízení a správu

výzkumných dat byly prokazatelně vynakládány „pro veřejné blaho“, tedy aby ostatní vědci mohli řešit jiné vědecké úlohy s využitím již pořízených datových sad, ať už na základě jednotlivých datových sad nebo kombinací více datových zdrojů [6]. Samozřejmě zákon č. 130/2002 Sb., zavádí i určité výjimky, které upravují, kdy výzkumná data nemusejí být poskytnuta.

Citace jsou ve vědecké komunitě způsobem ocenění práce citovaných autorů a v současném světě se používají jako nástroj pro evaluaci vědy, který často slouží ke kariernímu postupu i jako podklad pro přidělování finančních prostředků na vědu. Citace dat však nemá nahradit citace relevantní literatury, nýbrž zajistit ověřitelnou a opakovaně použitelnou informaci o dostupnosti výzkumných dat, jež podporují publikované závěry a tvrzení. Nedostatek řádných citací datových sad činí recenzovanou publikaci méně transparentní, ohrožuje reprodukovatelnost a brání otevřené vědě [7].

Citování použité datové sady je též nezbytné pro dodržení zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). Datové sady spadají pod autorská díla. Použití autorského díla pro vědecké účely je podle § 31 autorského zákona povoleno pouze tehdy, pokud je uvedeno, „je-li to možné, jméno autora, nejde-li o dílo anonymní, nebo jméno osoby, pod jejímž jménem se dílo uvádí na veřejnost, a dále název díla a pramen“.

Citace dat je proto důležitým nástrojem pro uznání práce původců a kurátorů dat a umožňuje jim sledovat, jak jsou jejich data používána. Citování dat umožňuje vědcům a dalším uživatelům snadno najít data, která byla použita v konkrétním výzkumu, případně tento výzkum replikovat a ověřit jeho výsledky. Bez správného citování dat by bylo obtížné dosáhnout cílů otevřené vědy, jež usiluje o sdílení dat a vědeckých poznatků s cílem urychlit vědecký pokrok. V neposlední řadě přispívá citování dat k tomu, aby jejich sdílení bylo spravedlivější a i původci a kurátoři datových sad získali patřičné uznání za svou práci [8].

Citování datových sad se potýká s řadou problémů [9], jako jsou např. jednoznačnost a ověřitelnost citace, tj. jak citovat datové sady, aby citace umožnila přesně identifikovat použitou datovou sadu a aby bylo možno ověřit, že byla použita právě citovaná data. Dalším problémem je, jak citovat dynamické datové sady, které se v průběhu času mění, či zda citovat datovou sadu nebo článek, jenž popisuje danou datovou sadu. Citace článku, který popisuje danou datovou sadu, přispívá k H-indexu autora, což může být výhodné pro jeho vědeckou kariéru. Naopak, citace datové sady, i když je relevantní pro výzkum, nemá obvykle přímý vliv na H-index autora článku. Tento rozdíl v dopadech na akademické metriky může ovlivnit rozhodování o doporučované formě citace. V neposlední řadě řeší vědecká komunita otázku, jak můžeme sledovat a vyhodnocovat používání datových sad.

K řešení těchto problémů vznikla řada standardů a osvědčených postupů pro citování dat. Mezi nejznámější patří Data Citation Principles, což je soubor principů pro citování dat, vyvinutý FORCE11 [10]. Dodržování navržených standardů přispívá ke zvýšení dopadu citovaného i citujícího díla [11]. Základním úkolem systému citování dat je však zaručení trvalosti citovaných dat i samotných citací [12], tj. aby citovaná datová sada byla i v budoucnu dostupná v citované podobě.

V uplynulých letech vznikla řada studií, které sledují, jak jsou datové sady citovány. Gregory et al. [13] zkoumali postupy, preference a motivace citování dat; přitom rozlišují tři typy citace datových sad. Prvním typem je citace dat v seznamu literatury. To znamená, že datové sady jsou citovány stejně jako jakýkoli jiný bibliometrický zdroj, a to zkrácenou citací v textu článku a plnou citací v seznamu literatury. Taková forma citace umožňuje snadné sledování citovanosti pomocí nástrojů citační analýzy a specializovaných citačních služeb. Druhým typem citace dat je pouhá zmínka o použitých datech v textu publikace. Posledním typem je nepřímá citace, kdy je odkaz na data uveden formou citace jiné související publikace (např. článek analyzující data nebo datový dokument).

Smith et al. [14] poukazují na jiný problém spojený s citováním datových sad na příkladu *Paleobiology Database*. Spočívá v tom, že souborná díla, jako jsou rozsáhlé datové sady tvořené příspěvky řady

autorů (a články na nich založené), jsou citována častěji než původci zdrojových dat, kteří do těchto velkých datových sad přispívají.

Digitální báze vodohospodářských dat

Základní báze geografických dat ČR (ZABAGED) je primární datovou geografickou sadou v Česku. Správcem ZABAGED je Zeměměřický úřad, který jej spravuje a rozšiřuje ve veřejném zájmu podle zákona č. 200/1994 Sb. Financování správy ZABAGED je tak zajištěno ze státního rozpočtu ČR. Vedle ZABAGED existují i další sady geografických dat. V oblasti vodního hospodářství je to zejména *Digitální báze vodohospodářských dat (DIBAVOD)*. DIBAVOD je spravován veřejnou výzkumnou institucí VÚV TGM a jeho správa je zajišťována z vnitřních zdrojů této instituce.

DIBAVOD je referenční geografická databáze vytvořená primárně z odpovídajících vrstev ZABAGED. Slouží k tvorbě tematických kartografických výstupů z oblasti vodohospodářství a ochrany vod nad základní mapou ČR 1 : 10 000. DIBAVOD je např. využíván pro prostorové analýzy v prostředí geografických informačních systémů (GIS) a pro zpracování reportingových dat podle rámcové směrnice 2000/60/ES v oblasti vodní politiky.

DIBAVOD lze tedy charakterizovat jako dynamickou databázi obsahující 75 různých objektů, jež popisují vodohospodářské prvky pro tvorbu základních vodohospodářských map. Objekty jsou členěny do 10 účelových skupin:

- A. Základní jevy povrchových a podzemních vod
- B. Účelová klasifikace povrchových a podzemních vod
- C. Chráněná území
- D. Záplavová území
- E. Měřicí a kontrolní místa povrchových vod
- F. Měřicí a kontrolní místa podzemních vod
- G. Objekty subsystému užívání vody
- H. Místa odběrů a vypouštění
- I. Objekty na toku
- J. Objekty meteorologických pozorování

V rámci sdílení jednotlivých objektů s širokou veřejností či zájmovými subjekty vznikla webová mapová aplikace na webové adrese dibavod.cz. Tato aplikace je přehledným interaktivním nástrojem pro publikaci dat a služeb s on-line přístupem přes webový prohlížeč. Může obsahovat rastrové i vektorové datové sady a umožňuje využít analytické i publikační nástroje.

V současnosti není zajištěno systémově stabilní financování DIBAVOD v celém rozsahu [15], což vede k tomu, že některé objekty jsou nedostupné nebo delší dobu neaktualizované. Informace o užívání DIBAVOD je tak významným aspektem při rozhodování o dalším financování správy DIBAVOD. VÚV TGM nemá k dispozici detailní informace o užívání této databáze vědeckou komunitou, protože data DIBAVOD je možno stáhnout z portálu dibavod.cz zdarma. Cílem této studie je proto zmapovat citovanost DIBAVOD a analyzovat typy citací této datové sady. Sami autoři DIBAVOD doporučují na hlavní stránce aplikace dibavod.cz citovat datovou sadu DIBAVOD formou nepřímé citace článku *GIS a kartografie ve VÚV TGM* [15] publikovaném v roce 2022 v časopise VTEI.

DATA A METODY

Jako základní výzkumná metoda pro tuto studii byl zvolen systematický přehled literatury. Systematický přehled literatury je specifický typ přehledů (review), který je zaměřen na hledání odpovědi na předem formulovanou výzkumnou otázku pomocí analýzy důkazů shromážděných ve vyhledané literatuře [16]. Základním rozdílem oproti tzv. „narativním“ přehledům literatury je omezení subjektivity

prostřednictvím jasně definovaných pravidel pro výběr a zařazení literatury do přehledu [17, 18]. Systematické přehledy literatury využívají metodiku PRISMA [19]. Svoji náročností jsou systematické přehledy literatury vhodné pro případy, kdy je analyzováno několik desítek či nižší stovky příspěvků.

Pro analýzu citovanosti byly zvoleny bibliometrické databáze Dimensions.AI [20], Scopus [21] a Web of Science [22]. Pro sběr dat byl použit přístup prostřednictvím webového rozhraní všech tří databází. Sběr dat byl uskutečněn 7. března 2024 pomocí vyhledání řetězce DIBAVOD ve všech polích a následně zopakován 1. července 2024. V databázi Dimensions.AI bylo nalezeno celkem 216 vědeckých publikací a jejich metadata byla vyexportována ve formátu csv a nahrána do tabulkového procesoru. V databázi Scopus bylo nalezeno 47 vědeckých publikací, které byly opět vyexportovány ve formátu csv a nahrány do tabulkového procesoru. V databázi Web of Science – Core collection byly nalezeny tři články, při rozšíření dotazu na všechny databáze ve Web of Science bylo nalezeno pět odkazů na dvě datové sady odvozené z DIBAVOD. Záznamy o těchto datových sadách nebyly do analýzy zahrnuty.

V prvním kroku byly vyloučeny duplicity, čímž vznikl seznam 231 vědeckých publikací určených ke screeningu. V rámci screeningu byl každý nalezený dokument zkontrolován, zda skutečně obsahuje citaci DIBAVOD. Z další analýzy bylo takto vyřazeno 104 záznamů, jež necitovaly DIBAVOD, a pět záznamů, u nichž se nepodařilo ověřit, zda citují DIBAVOD (např. z důvodu nedostupnosti daného díla pro autory).

Do analýzy citovanosti tak vstupovalo 122 prací, citujících DIBAVOD. Pro sledování typů citací byla použita modifikovaná typologie popsána Gregorym et al. [13]. U každého díla citujícího DIBAVOD byla prověřena forma citace a způsob citace zdroje. Forma citace byla zařazena do jedné ze dvou kategorií – do „citace v textu“ nebo „citace v seznamu literatury“. V případě formy citace v seznamu literatury byly vytvořeny kategorie způsobu citace zdroje „není citován žádný zdroj“, „je citován doporučený článek“ (tj. je citován článek *GIS a kartografie ve VÚV TGM* [15]), „je citován VÚV TGM“, „jsou citovány stránky dibavod.cz“. Na základě analýzy citací byla dodatečně doplněna nová kategorie „je citován Hydroekologický informační systém VÚV TGM“ (alias HEIS VÚV TGM). HEIS VÚV TGM je jiný informační systém provozovaný VÚV TGM, který poskytuje atributová data o vodním hospodářství v ČR. V případě formy „citace v textu“ byly zvoleny stejné kategorie způsobu citace zdroje, logicky se však u této formy citace DIBAVOD nemůže objevit „je citován doporučený článek“. Analýzu citací prováděli oba autoři této studie; druhý autor měl na starosti prvotní analýzy, první autor kontroloval výsledky a rozhodoval v případě nejasných zařazení.

Následně byly tyto údaje statisticky zpracovány a obsahově analyzovány.

VÝSLEDKY A DISKUZE

Citace dle jednotlivých kategorií formy citace a způsobu citování zdroje uvádí *obr. 1*. Celkem bylo nalezeno 122 prací citujících DIBAVOD. V citacích DIBAVOD mírně převládá citování formou odkazů v seznamu literatury, kterých je 64 (tj. 52,5 %). To lze považovat za dobrý výsledek, jelikož v obecné rovině převládají citace v textu nad citacemi v seznamu literatury [23]. Předložené výsledky lze však jen těžko generalizovat, neboť počet citujících článků je nízký. Jak uvádějí Rogers et al. [24], vzorky 1 000 dokumentů poskytují dobré vodítko pro relativní (nikoli však absolutní) citační analýzy; a studie s méně než 200 dokumenty trpí vysokou variabilitou výsledků.

Citování formou odkazu v seznamu literatury je z pohledu kurátorů datových sad vhodnější, protože umožňuje snáze sledovat užívání datové sady pomocí specializovaných bibliometrických databází. Neznamena to však, že by citace datové sady v textu citujícího dokumentu měly být považovány za nevhodné. Citování dat je zatím stále v počátcích svého rozšíření a autoři se mu teprve učí. Je proto důležité, aby se citování použitých datových sad stalo součástí všeobecné „dovednosti citovat“ a dobrých publikačních zvyků.

Poněkud nepříjemné zjištění představuje 29 děl (23,8 %), které v textu pouze zmiňují DIBAVOD, aniž by poskytly jakékoli uznání autorům a kurátorům DIBAVOD, resp. odkaz na zdroj DIBAVOD. Ačkoli se může zdát, že 23,8 % není vysoký podíl, stále jde o projev neznalosti či nedodržování citačních pravidel. Nepřesné, neúplné či nedbalé citování, kdy není možné identifikovat citovaný zdroj, je považováno většinou příruček publikační etiky [např. 25] za porušení publikační etiky, resp. plagiátorství. Navíc jde de facto i o porušení českého autorského zákona, který kromě názvu použitého díla požaduje i uvedení autora a zdroje.

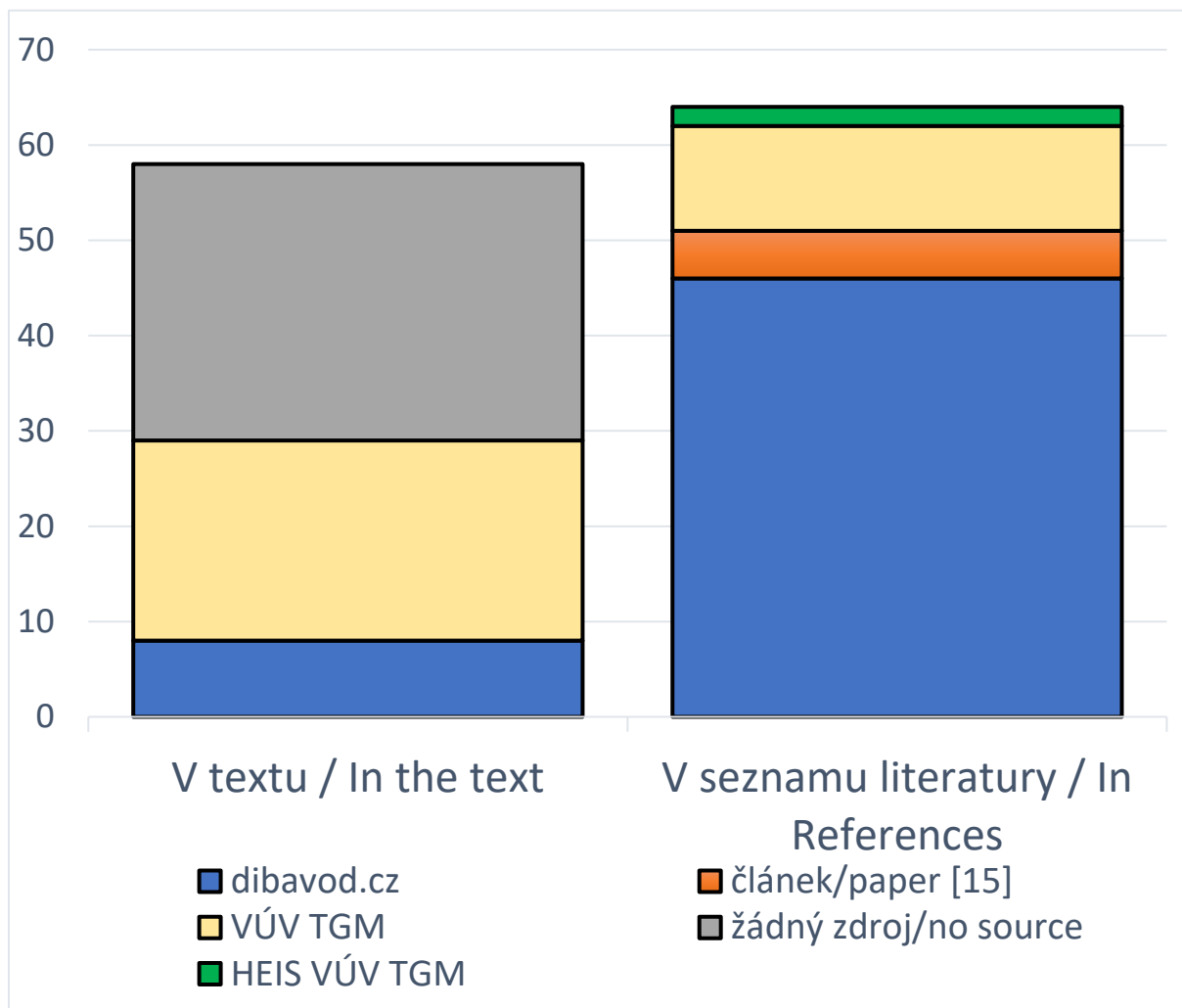
V této souvislosti se nabízí otázka, zda v současné době lze neuvedení přesného zdroje považovat za porušení publikační etiky, pokud z (ne)citujícího díla vyplývá, že byla použita konkrétní datová sada, kterou lze na internetu snadno dohledat. S ohledem na dodržení principů FAIR [8] by i citace datových sad měla přispět k nalezení použité datové sady (Findable), její dostupnosti (Accessible), interoperabilitě (Interoperable) a opětovné využitelnosti (Reusable). Citace dat hrají důležitou roli při zajišťování jejich dohledatelnosti a dostupnosti, zejména pokud jsou v citacích používány trvalé (perzistentní) identifikátory, jako např. DOI. Groth et al. [26] diskutují přínos citování datových sad pro jejich opětovné použití. Podle našeho názoru bude mít častější citování dat vliv i na jejich interoperabilitu, jelikož data s vyšší interoperabilitou by měla být více využívána i citována. Přiznání významu citování dat, např. zahrnutím citací dat do systémů hodnocení, vyvolá tlak na kurátory dat k zajištění vyšší interoperability jimi spravovaných datových sad.

Celkem 54 prací se odkazovalo na webové stránky dibavod.cz, což byl nejčastější způsob citování zdroje v případě datové sady DIBAVOD. Z toho 46 odkazů na dibavod.cz bylo formou odkazu v seznamu citované literatury a osm odkazů na dibavod.cz bylo uvedeno přímo v textu citujícího článku. Popisný článek [15], který doporučují citovat sami kurátoři databáze DIBAVOD na stránkách dibavod.cz, byl citován pouze pětkrát, což je velmi malé číslo. Jednak to může být tím, že článek vznikl poměrně nedávno (v roce 2022), a také tím, že datové články nejsou ještě příliš využívány pro citování datových sad, přesto ale dochází k trvalému nárůstu citací datových článků [27]. Celkové citování datových sad je však stále na velmi nízké úrovni, a to bez ohledu na datový repozitář, v němž jsou data nahrána [28].

Dva články citovaly DIBAVOD jako součást HEIS VÚV TGM. V obou případech šlo o poměrně staré citace, první byla z roku 2009, druhá sice z roku 2021, ale citovala zdroj z roku 1965. Jiné dva články citovaly HEIS VÚV TGM, protože však oba tyto články citovaly zároveň VÚV TGM nebo dibavod.cz, byly započítány do kategorií citujících tyto zdroje.

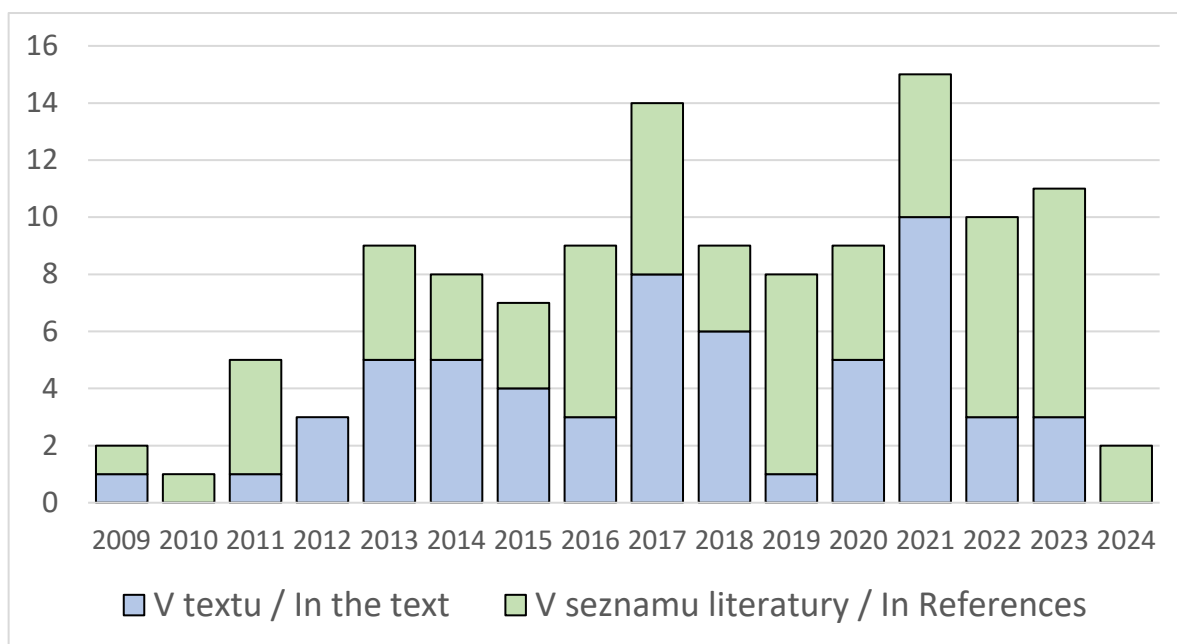
Poslední způsob citování zdroje představují citace, jež uvádějí, že DIBAVOD je spravován VÚV TGM. Těchto citací bylo celkem 32, z toho 21 formou citace v textu a 11 formou citace v seznamu literatury. Tento způsob citace nelze považovat za optimální, nicméně přiznává VÚV TGM alespoň zásluhy za správu datové sady DIBAVOD.

Obr. 2 ukazuje, že lze vysledovat určitý rostoucí trend v celkovém počtu citací, ale nikoli v tom, zda se mění podíl citací formou citací v textu a citací v seznamu literatury. Obdobně je na *obr. 3* patrné, že ani způsob citování zdroje nevykazuje žádný patrný trend a jednotlivé kategorie jsou náhodně zastoupeny v jednotlivých letech. Z toho lze usuzovat, že citování datových sad dosud není v české vědecké komunitě zažité. Přitom vzdělání v oblasti citování datových souborů je zásadní pro podporu akademické integrity, rozvoj kritických digitálních dovedností a pro zlepšení etického a efektivního využívání dat.



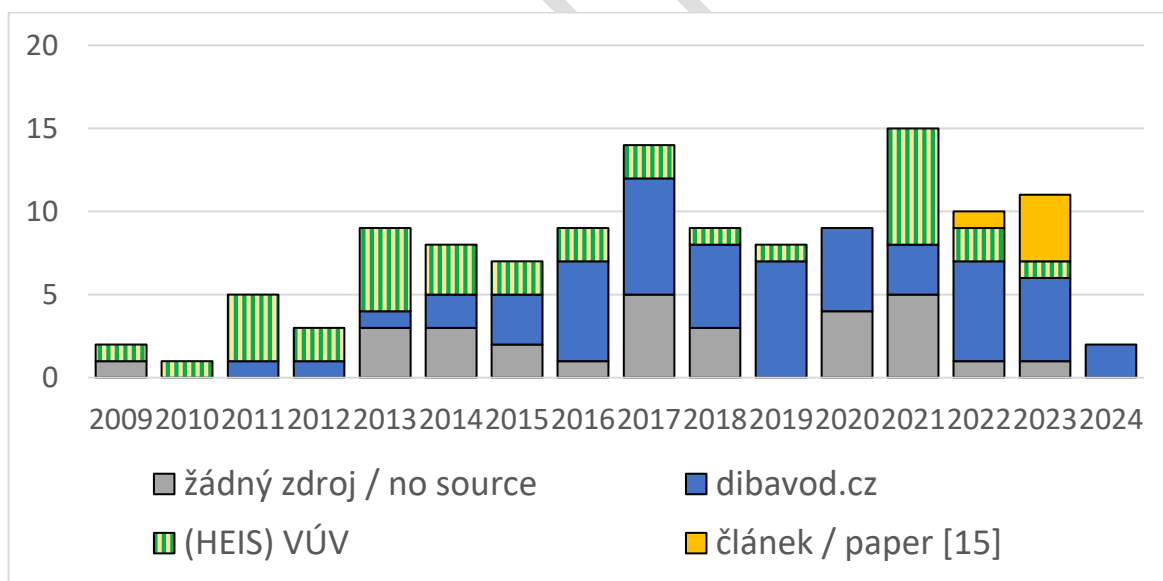
Obr. 1. Struktura typů citací DIBAVOD

Fig. 1. Structure of DIBAVOD citation types



Obr. 2. Vývoj citovanosti DIBAVOD v čase

Fig. 2. Development of DIBAVOD citations over time



Obr. 3. Rozdělení typů citací v čase

Fig. 3. Distribution of citation types over time

V rámci této studie jsme se soustředili na citování ve vědeckých časopisech, které jsou zařazeny do citačních databází, což umožnilo relativně jednoduché zpracování dat. Geografické datové sady, jako jsou DIBAVOD či ZABAGED, však vznikají zejména s cílem poskytnout podklady i pro laickou veřejnost. Jednou z výzev pro systémy hodnocení znovuvyužívání výzkumných dat je právě způsob použití těchto dat běžnou odbornou veřejností mimo akademickou sféru. To zahrnuje zejména citování

v tzv. šedé literatuře, ve strategických dokumentech a politikách, v rozhodnutích správních orgánů apod. Tyto citační analýzy jsou však samy o sobě vysoce náročné, protože na rozdíl od vědecké publikační produkce neexistují pro tyto typy dokumentů snadno použitelné zdroje informací. Nejsnadněji lze na tyto činnosti využít webové vyhledávače, jako je Google. Následná analýza výsledků vyhledávání se však velmi špatně automatizuje. Citování datových sad nemusí být jediným způsobem prokázání používání výzkumných dat. Dalšími možnostmi jsou např. počty stažení, počty smluv o využití apod. Největší komplikací je zde opět nedostatek snadno dostupných informací o těchto typech indikátorů.

ZÁVĚRY

Provedená analýza ukázala, že citování datové sady DIBAVOD nelze považovat za optimální. Ze 122 analyzovaných prací jich 58 použilo odkaz pouze v textu a z toho 29 jen prostou zmínku o využití DIBAVOD bez podrobnějších informací o datové sadě nebo jejích autorech či původci. Celkem 54 prací uvedlo odkaz na stránky dibavod.cz a 34 prací citovalo DIBAVOD formou odkazu na původce, tj. VÚV TGM. Pouhých pět prací použilo doporučenou citaci prostřednictvím článku *GIS a kartografie ve VÚV TGM*, to však může být hlavně důsledkem nízkého stáří tohoto doporučeného článku. Studie tak prokázala, že citování vodohospodářských datových sad, jako je zejména DIBAVOD, není v ČR široce rozšířené a neexistuje ustálená forma a způsob citování těchto sad. Důležitost citování geografických dat by proto měla být zdůrazňována jak v rámci vzdělávacích programů vysokých škol, tak prostřednictvím veřejných akcí i odborných grémií. Stejně tak studie prokázala vysokou heterogenitu ve formě citování datové sady DIBAVOD. I v tomto ohledu je třeba mnohem větší osvěty.

Veškerá data použitá ve studii lze získat z databází Dimension.AI, Scopus a Web of Science postupy popsanými v této studii. Na vyžádání u korespondenčního autora je možno získat zdrojový soubor ve formátu MS Excel, v němž byly všechny analýzy prováděny.

Poděkování

Autoři chtějí poděkovat oběma recenzentům za velmi podnětné připomínky a doporučení.

Prohlášení o střetu zájmů

Korespondenční autor je součástí managementu VÚV TGM, který časopis VTEI vydává, a předsedou Redakční rady časopisu VTEI. Tyto skutečnosti však neměly žádný vliv na výsledky prezentované studie. VÚV TGM neposkytl žádné prostředky na zpracování této studie.

Literatura

[1] COSTELLO, M. J. Motivating Online Publication of Data. *BioScience* [on-line]. 2009, 59(5), s. 418–427. ISSN 0006-3568. Dostupné z: <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), s. 1 036–1 064. ISSN 1468-4527. Dostupné z: <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), s. 32–46. ISSN 2330-1643. Dostupné z: <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), s. 38–45. ISSN 0322–8916, 1805-6555. Dostupné z: <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), s. 20–27. ISSN 0322–8916, 1805-6555. Dostupné z: <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), s. 1 059–1 078. ISSN 1532-2890. Dostupné z: <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. Dostupné z: <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. Dostupné z: <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), s. 52. ISSN 1683-1470. Dostupné z: <https://doi.org/10.5334/dsj-2019-052>
- [10] DATA CITATION SYNTHESIS GROUP. *Joint Declaration of Data Citation Principles* [online]. San Diego CA: FORCE11, 2014. Dostupné z: <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. Dostupné z: <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), s. 6–20. ISSN 2330-1643. Dostupné z: <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), s. 622–649. ISSN 2641-3337. Dostupné z: 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), s. 165–176. ISSN 0094-8373, 1938-5331. Dostupné z: <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), s. 47–52. ISSN 0322–8916, 1805-6555. Dostupné z: <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, s. 333–339. ISSN 0148-2963. Dostupné z: <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) s. S1–S5. ISSN 1099-1379. Dostupné z: <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), s. 207–222. ISSN 1467-8551. Dostupné z: <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. Dostupné z: <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), s. 430–435. ISSN 1751-1577. Dostupné z: <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), s. 377–386. ISSN 2641-3337. Dostupné z: 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), s. 363–376. ISSN 2641-3337. Dostupné z: 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), s. 1 346–1 354. ISSN 2330-1643. Dostupné z: <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), s. 777–794. ISSN 1588-2861. Dostupné z: <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), s. 78–86. ISSN 2641-435X. Dostupné z: 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), s. 115–131. ISSN 2330-1635, 2330-1643. Dostupné z: <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, s. 6 315–6 332. ISSN 1588-2861. Dostupné z: <https://doi.org/10.1007/s11192-024-04956-x>

Autoři

Ing. Libor Ansorge, Ph.D.¹

libor.ansorge@vuv.cz

ORCID 0000-0003-3963-8290

Anna Ansorgeová²

st102751@students.ujep.cz

¹Výzkumný ústav vodohospodářský T. G. Masaryka, Praha (Česká Republika)

²Přírodovědecká fakulta Univerzity J. E. Purkyně, Ústí nad Labem (Česká Republika)

Příspěvek prošel recenzním řízením.

DOI: 10.46555/VTEL.2024.11.001

ISSN 0322-8916/© 2025 Autoři. Tuto práci je kdokoli oprávněn šířit a využívat za podmínek licence CC BY-NC 4.0

GEOGRAPHIC DATA CITATIONS: CASE STUDY OF DIBAVOD – SYSTEMATIC REVIEW

ANSORGE, L.¹; ANSORGEOVÁ, A.²

¹T. G. Masaryk Water Research Institute, Prague (Czech Republic)

²Faculty of Science, J. E. Purkyně University in Ústí nad Labem (Czech Republic)

Keywords: DIBAVOD – data citations – GIS – geographic database

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 recording citations. In this study, an analysis of the citations of the geographic database DIBAVOD, which is managed by the T. G. Masaryk Water Research Institute, was performed. 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.

Accepted for print