Two years of CiTO annotations

Type: Correspondance

Egon Willighagen [1,*,orcid:0000-0001-7542-0286]

- Correspondance: egon.willighagen@maastrichtuniversity.nl

Abstract

Citations are an essential aspect of research communication and have become the basis of many metrics in the academic publishing world. Some see citation counts as a mark of scientific impact or even quality, but in reality the reasons for citing other work are manifold. Two years ago the *Journal of Cheminformatics* proposed a pilot for the adoption of annotating citations with their annotations. Basically, when you cite a source like a journal article or datasets, you also explain why specifically you cite that article. Particularly, the agreement and disagreement and reuse of methods and data is interesting. This article explores what happened after the launch of the pilot, gives examples of how people used the idea, and what needs to happen next.

Main text

Communiting new research findings is primarily done by written texts in the form of scholarly articles, books, and book chapters. To not having to repeat past research by themselves or others, authors cite research relevant [1]. However, the reasons why authors cite literature vary, which complicates how we use citations [2]. Typing citations is therefore of interest: it allows us to navigate ltierature more easily: it points us to essential research methods, data, and can warn us of research that cannot be reproduced, or others disagree with [2].

With the use of citations increasingly being picked up to help researchers with tools like scite.ai [2] en Connected Papers [3], having typed citations will help us explore literature. Therefore, the *Journal of Cheminformatics* started a pilot for the adoption of annotating citations with their annotations [4].

The Citation Typing Ontology Pilot

The pilot consists of a couple of components and the editorial explains some of them [4]. The Citation Typing Ontology was selected to express the intention [1], the intention is expressed a compact identifier wrapped in square brackets, also called a safe CURIE, standard proposed by the W3C in 2010 [6]. The *cito* prefix is registered in BioRegistry [7]. The *bibnotes* concept of the Springer Nature publishing platform was used as carrier. Authors are guided by a landing page consisting of a BioMedCentral Collection at https://www.biomedcentral.com/collections/cito and author guidelines explaining to authors how they can add the annotations with their favoriate editor at https://jcheminform.github.io/jcheminform-author-guidelines/cito.

Because the CiTO ontology has many terms for many different citation intentions, we made a selection of CiTO terms authors could use [8]: [**cito:citesAsDatasource**] to indicate to a source the provides data to back up a claim, [**cito:usesDataFrom**] to indicate that the authors reused data, [**cito:usesMethodIn**] when a method or protocol explain in that source is used, and a few more general intentions like [**cito:discusses**], [**cito:extends**], [**cito:agreesWith**], and [**cito:disagreesWith**]. The journal itself would adopt the following CiTO annotations: [**cito:retracts**], [**cito:repliesTo**], and [**updates**]. Fortunately, it has not been used yet, but the first would be used if an article with be retracted from the journal. The second would be used when

an Letter to the Editor replies to an earlier published article, and [**cito:updates**] when an Correction was published.

Wikidata and Scholia

To track the uptake but also to demonstrate the impact, we extended Scholia to visualize citation intention data. Scholia is a graphical interface around the data stored in Wikidata [9] and includes citations from OpenCitations [10] and PubMed. Wikidata allows adding qualifier to statements which allowed us to define a data model for citations annoted with CiTO intention; the Wikidata property P3712 has been used for this, labeled *objective of project or action* (see Figure 1).

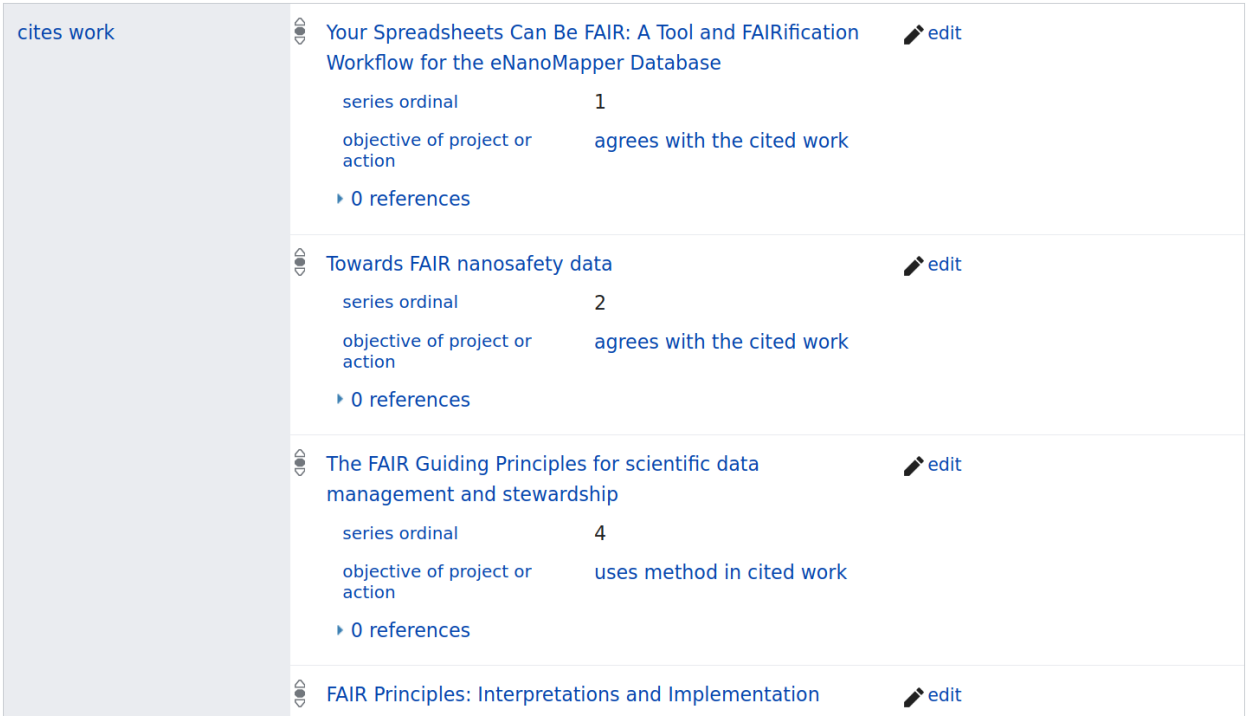

Figure 1: Screenshot of the citation statements for an article where the *objective of project or action* qualifier is used to annotate the citation with their CiTO intentions.

This data in Wikidata can then be accessed via Scholia and Scholia tells use some overall statisticss of the number of annotations, which we reported on about a year ago too [11]. Since last year and written down on August 25, the number of annotations and the number of annotated citations have almost doubled (from 377 to 603 and from 304 to 494, respectively). The first number is higher because one citation can have more than one citation intention. To continue, the current number of citations are citing 387 articles in 141 different scholarly journals, and they are found in 98 articles in 48 different journals (see Figure 2) [12].

It must be noted that the *Journal of Cheminformatics* is only one possible source, but is still the only journal that uses CiTO annotation explicitly in the articles itself. And with 335 annotated citations in 32 articles it also is the major source of CiTO annotations in Wikidata at the time of writing. However, CiTO intention annotations in Wikidata can come from other sources too and be added both manually and automatically using the tools around Wikidata. When all annotation is combined, Scholia shows us that [**cito:citesAsAuthority**] is the most used intention, with 226 annotated citations (out of 603) in 38 articles. [**cito:usesMethodIn**] follows with 102 annotated citations.

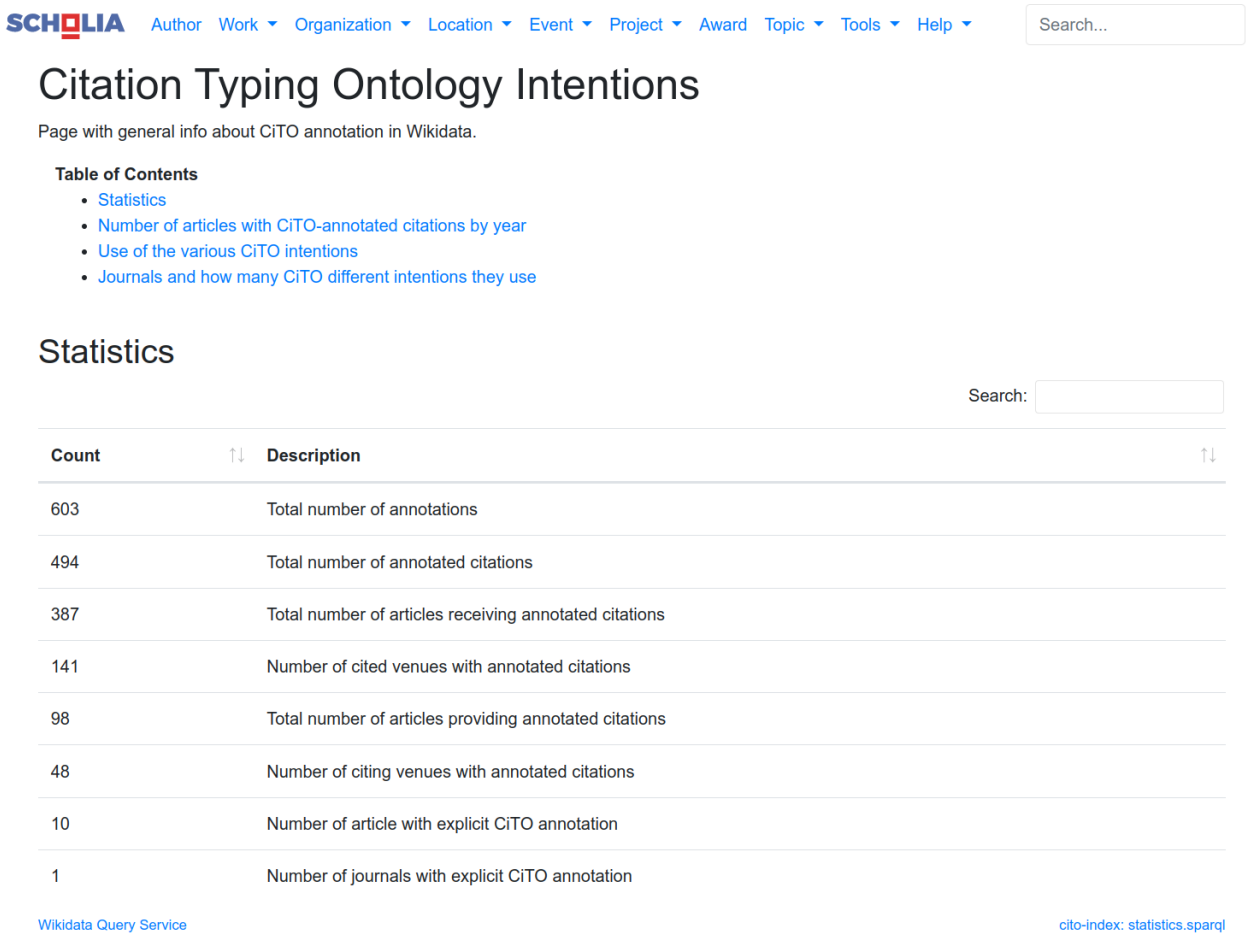

Figure 2: Screenshot of the Scholia Citation Typing Ontology page showing the daily statistics.

Adoption by *Journal of Cheminformatics* authors

In the first two years after the start of the Pilot, including the seminal editorial [4], the *Journal of Cheminformatics* has published fifteen articles with explicit CiTO annotation: three Editorials, four Research articles, two Database, and one of each of Data Note, Software, Letter to the Editor, Letter Response, Educational, and Methodology. Ten were published in the first year (Table 1) and five in the second year (Table 2). Each article annotated one or more citations with CiTO intentions, and several articles annotated every citation, far exceeding the original anticipation.

Also exceeding expectations is the diversity of the chosen CiTO intention types. The original guidance focused on [**cito:citesAsDataSource**], [**cito:usesDataFrom**] (the first is used to cite an article with data, the second when you reused data and citate the article where it comes from), [**cito:usesMethodIn**], [**cito:citesAsAuthority**], [**cito:discusses**], [**cito:extends**], [**cito:agreesWith**], and [**cito:disagreesWith**]. Not only have all of these been used, authors also used [**cito:citesForInformation**], [**cito:citesAsPotentialSolution**], [**cito:citesAsRelated**], [**cito:documents**], and [**cito:obtainsBackgroundFrom**].

Table 1: *Journal of Cheminformatics* articles in the first year of the pilot.

Type	**Article**	**Intentions**	**%CiTO**
Editorial	Adoption of the Citation Typing Ontology by the *Journal of Cheminformatics* [4]	cito:citesAsAuthority, cito:citesForInformation, cito:usesMethodIn	42%
Methodology	Predicting target profiles with confidence as a service using docking scores [13]	cito:agreesWith, cito:extends, cito:citesAsAuthority, cito:citesAsDataSource, cito:usesMethodIn	47%
Research article	Too sweet: cheminformatics for deglycosylation in natural products [14]	cito:citeAsAuthority, cito:citesForInformation, cito:usesDataFrom, cito:usesMethodIn	
Educational	A ligand-based computational drug repurposing pipeline using KNIME and Programmatic Data Access: case studies for rare diseases and COVID-19 [15]	cito:agreesWith, cito:citesAsAuthority, cito:citesAsDataSource, cito:discusses, cito:usesDataFrom, cito:usesMethodIn	98%
Database	COCONUT online: Collection of Open Natural Products database [16]	cito:citesForInformation, cito:usesDataFrom, cito:usesMethodIn	41%
Editorial	What is the role of cheminformatics in a pandemic? [17]	cito:agreesWith, cito:citesAsAuthority, cito:citesAsDataSource, cito:citesForInformation, cito:citesAsPotentialSolution	91%
Research article	Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag [18]	cito:citesAsAuthority, cito:citesAsDataSource, cito:citesAsMetadataDocument, cito:discusses, cito:extends, cito:usesDataFrom, cito:usesMethodIn	100%
Software	IDSM ChemWebRDF: SPARQLing small-molecule datasets [19]	cito:citesAsAuthority, cito:citesAsRelated, cito:usesDataFrom, cito:usesMethodIn	97%

Type	Article	Intentions	%CiTO
Letter to the Editor	FAIR chemical structures in the Journal of Cheminformatics [20]	cito:citesAsAuthority, cito:documents	100%
Letter Response	Reply to "FAIR chemical structure in the Journal of Cheminformatics" [21]	cito:agreesWith, cito:citesAsAuthority, cito:citesAsDataSource, cito:obtainsBackgroundFrom, cito:repliesTo	100%

Table 2: *Journal of Cheminformatics* articles in the second year of the pilot. The * indicates that the percentage is based on CiTO intends different from cito:cites.

Type	Article	Intentions	%CiTO
Research article	DECIMER 1.0: deep learning for chemical image recognition using transformers [22]	cito:agreesWith, cito:citesAsAuthority, cito:citesAsDataSource, cito:extends, cito:usesMethodIn	66% (*)
Research article	*DrugEx* v2: de novo design of drug molecules by Pareto-based multi-objective reinforcement learning in polypharmacology[23]	cito:extends, cito:citesAsAuthority, cito:usesMethodIn	100%
Database	PSnpBind: a database of mutated binding site protein–ligand complexes constructed using a multithreaded virtual screening workflow [24]	cito:citesAsDataSource, cito:usesDataFrom, cito:usesMethodIn	23%
Editorial	Diversifying cheminformatics [25]	cito:citesAsAuthority, cito:citesAsEvidence, cito:citesForInformation, cito:citesAsRecommendedReading, cito:citesAsSourceDocument, cito:containsAssertionFrom	100%
Data Note	DECIMER - hand-drawn molecule images dataset [26]	cito:agreesWith, cito:cites, cito:citesAsAuthority, cito:extends, cito:usesDataFrom, cito:usesMethodIn	41% (*)

Technological innovation

To make life easier for authors, and following a Twitter discussion in Spring 2021, a Markdown template was developed with native CiTO support: https://jcheminform.github.io/jcheminform-author-guidelines/cito-guidelines/markdown.html. Here, the author merely indicates the CiTO type when they cite the article. This is using a method introduced by Krewinkel et al. [27]. The manuscript can then be converted to a Microsoft Word file with Pandoc (https://pandoc.org/) for submission to the journal.

The *Journal of Cheminformatics* template is available from the journal's GitHub organization, and authors and editors should feel free to adapt it to the needs of other journals. The BioHackrXiv (https://biohackrxiv.org/) preprint server also support CiTO annotations [28] and this template can be found at https://github.com/biohackrxiv/publication-template.

Discussion

Because many publishing platforms currently do not support display of citation level intention annotation, the default model only provides the annotation in the bibliography. The choice was made to use the `bibnotes` idea which allows giving additional notes to a reference in the bibliography. The CiTO Pilot uses the *safe CURIE* standard, but the typesetting in the *Journal of Cheminformatics* has shown to be inconsistent, even within single articles, suggesting additional training or guidance may be needed.

This links to a limitation to the current use of CiTO annotation: the citation-level annotation may be supported by some authoring systems (Markdown), even then the depiction may not be supported. If we convert a Markdown file to a Microsoft Word file, the annotation is kept in the bibliography. However, when writing a manuscript in Microsoft Word, LibreOffice, or Google Docs, combined with a reference manager like Zotero, the CiTO annotation cannot be stored as part of the manuscript easily. The problem here is that CiTO annotation in the reference manager are no longer linked to when they are cited. The workaround is to add the CiTO annotation after completion of the Word document, directly to the bibliography.

LaTeX users may find them in a situation between that of Markdown users and reference manager users: only if a manuscript-level Bib(La)TeX file is used the CiTO annotation can be added as notes to the BiBTeX file. This is explained in this guidance document: https://jcheminform.github.io/jcheminform-author-guidelines/cito-guidelines/latex.html.

What is next?

With fourteen articles published in the CiTO Collection, the adoption is not as high as one would have hoped. Nevertheless, the pilot is triggering interest with both authors and publishing platforms. The *Journal of Cheminformatics* has already published a few more articles with CiTO annotation and a search on the preprint servers ResearchSquare and ChemrXiv show a few more manuscripts. The support by BioHackrXiv is a nice example of adoption beyond BioMedCentral.

The Scholia use case shows how the information can be used downstream. Further uptake of this idea of typed citations depends on the combined willingness of journal editors, authors, publishers and indexing services alike. The rise of services like scite.ai shows that the research community needs this kind of information.

Availability of data and materials

CiTO annotation in the *Journal of Cheminformatics* are available from the journal's articles. CiTO annotation data in Scholia is available from Wikidata. Markdown templates that support CiTO are available from https://github.com/jcheminform/markdown-jcheminf and https://github.com/biohackrxiv/publication-template.

Competing interests

No competing interests.

Funding

Part of this work was supported by ELIXIR, the research infrastructure for life-science data.

Authors' contributions

Not applicable.

Acknowledgments

This work would not be possible without the support from Springer Nature and Matthew Smyllie in particular and the editors of the *Journal of Cheminformatics*, Rajarshi Guha, Nina Jealizakova, and Barbara Zdrazil. Carlin MacKenzie is thanked for integration the sections on CiTO use into the main Scholia journal aspects [29]. Huge thanks goes to Albert Krewinkel for developing the Markdown/Pandoc integration. Finally, thanks to all authors for including CiTO annotation in their articles.

References

1. Shotton D (2010) CiTO, the citation typing ontology. Journal of Biomedical Semantics 1:S6+. https://doi.org/10.1186/2041-1480-1-s1-s6 [**cito:agreesWith**] [**cito:citesAsAuthority**] [**cito:usesMethodIn**]
2. Nicholson JM (2021) Smart(er) citations. Matter 4:756–758. https://doi.org/10.1016/j.matt.2021.02.007 [**cito:agreesWith**] [**cito:citesAsAuthority**]
3. Tarnavsky-Eitan A, Smolyansky E, Knaan-Harpaz I, Perets S (2022) Connected Papers. https://connectedpapers.com. Accessed 28 Aug 2022 [**cito:citesAsAuthority**]
4. Willighagen E (2020) Adoption of the citation typing ontology by the journal of cheminformatics. Journal of Cheminformatics 12:47, s13321-020-00448-1. https://doi.org/10.1186/s13321-020-00448-1 [**cito:discusses**] [**cito:extends**] [**cito:citesForInformation**]
5. McCarron S, Birbeck M (2010) CURIE syntax 1.0. https://www.w3.org/TR/2010/NOTE-curie-20101216/ [**cito:usesMethodIn**]
6. Wimalaratne SM, Juty N, Kunze J, et al (2018) Uniform resolution of compact identifiers for biomedical data. Scientific Data 5:180029. https://doi.org/10.1038/sdata.2018.29 [**cito:usesMethodIn**]
7. Hoyt CT, Balk M, Callahan TJ, et al (2022) Unifying the identification of biomedical entities with the bioregistry. BioRxiv. https://doi.org/10.1101/2022.07.08.499378 [**cito:usesDataFrom**]
8. (2020) Citation Typing Ontology (CiTO) Pilot. https://www.biomedcentral.com/collections/cito
9. Nielsen FÅ, Mietchen D, Willighagen E (2017) Scholia, scientometrics and wikidata. In: Blomqvist E, Hose K, Paulheim H, et al (eds) The semantic web: ESWC 2017 satellite events. Springer International Publishing, Cham, pp 237–259 [**cito:usesMethodIn**]
10. Peroni S, Shotton D (2020) OpenCitations, an infrastructure organization for open scholarship. Quantitative Science Studies 1:428–444. https://doi.org/10.1162/qss_a_00023
11. Willighagen E (2021) Typed Citations in the Journal of Cheminformatics. In: On Physical Sciences. https://blogs.biomedcentral.com/on-physicalsciences/2021/07/20/typed-citations-journal-of-cheminformatics/ [**cito:citesAsDataSource**]
12. (2022) Citation Typing Ontology Intentions - Scholia. https://scholia.toolforge.org/cito/. Accessed 26 Aug 2022 [**cito:citesAsDataSource**]
13. Ahmed L, Alogheli H, McShane SA, et al (2020) Predicting target profiles with confidence as a service using docking scores. Journal of Cheminformatics 12:62. https://doi.org/10.1186/s13321-020-00464-1 [**cito:citesForInformation**]
14. Schaub J, Zielesny A, Steinbeck C, Sorokina M (2020) Too sweet: Cheminformatics for deglycosylation in natural products. Journal of Cheminformatics 12:67. https://doi.org/10.1186/s13321-020-00467-y [**cito:citesForInformation**]
15. Tuerkova A, Zdrazil B (2020) A ligand-based computational drug repurposing pipeline using KNIME and programmatic data access: Case studies for rare diseases and COVID-19. Journal of Cheminformatics 12:71. https://doi.org/10.1186/s13321-020-00474-z [**cito:citesForInformation**]
16. Sorokina M, Merseburger P, Rajan K, et al (2021) COCONUT online: Collection of open natural products database. Journal of Cheminformatics 13:2. https://doi.org/10.1186/s13321-020-00478-9 [**cito:citesForInformation**]
17. Guha R, Willighagen E, Zdrazil B, Jeliazkova N (2021) What is the role of cheminformatics in a pandemic? Journal of Cheminformatics 13:16, s13321-021-00491-6. https://doi.org/10.1186/s13321-021-00491-6 [**cito:citesForInformation**]
18. Schymanski EL, Kondić T, Neumann S, et al (2021) Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. Journal of Cheminformatics 13:19. https://doi.org/10.1186/s13321-021-00489-0 [**cito:citesForInformation**]
19. Galgonek J, Vondrášek J (2021) IDSM ChemWebRDF: SPARQLing small-molecule datasets. Journal of Cheminformatics 13:38. https://doi.org/10.1186/s13321-021-00515-1 [**cito:citesForInformation**]

20. Schymanski EL, Bolton EE (2021) FAIR chemical structures in the journal of cheminformatics. Journal of Cheminformatics 13:50. https://doi.org/10.1186/s13321-021-00520-4 [**cito:citesForInformation**]

21. Guha R, Jeliazkova N, Willighagen E, Zdrazil B (2021) Reply to "FAIR chemical structure in the journal of cheminformatics." Journal of Cheminformatics 13:49. https://doi.org/10.1186/s13321-021-00521-3 [**cito:citesForInformation**]

22. Rajan K, Zielesny A, Steinbeck C (2021) DECIMER 1.0: Deep learning for chemical image recognition using transformers. Journal of Cheminformatics 13:61. https://doi.org/10.1186/s13321-021-00538-8 [**cito:citesForInformation**]

23. Liu X, Ye K, Vlijmen HWT van, et al (2021) DrugEx v2: De novo design of drug molecules by pareto-based multi-objective reinforcement learning in polypharmacology. Journal of Cheminformatics 13:85. https://doi.org/10.1186/s13321-021-00561-9 [**cito:citesForInformation**]

24. Ammar A, Cavill R, Evelo C, Willighagen E (2022) PSnpBind: A database of mutated binding site protein-ligand complexes constructed using a multithreaded virtual screening workflow. Journal of Cheminformatics 14:8. https://doi.org/10.1186/s13321-021-00573-5 [**cito:citesForInformation**]

25. Zdrazil B, Guha R (2022) Diversifying cheminformatics. Journal of Cheminformatics 14:25, s13321-022-00597-5. https://doi.org/10.1186/s13321-022-00597-5 [**cito:citesForInformation**]

26. Brinkhaus HO, Zielesny A, Steinbeck C, Rajan K (2022) DECIMER—hand-drawn molecule images dataset. Journal of Cheminformatics 14:36. https://doi.org/10.1186/s13321-022-00620-9 [**cito:citesForInformation**]

27. Krewinkel A, Winkler R (2017) Formatting open science: Agilely creating multiple document formats for academic manuscripts with pandoc scholar. PeerJ Computer Science 3:e112. https://doi.org/10.7717/peerj-cs.112 [**cito:usesMethodIn**]

28. Prins P, Ohta T, Willighagen E (2021) Metadata for BioHackrXiv markdown publications. https://github.com/biohackrxiv/bhxiv-metadata/blob/main/doc/elixir_biohackathon2021/paper.md [**cito:citesAsDataSource**]

29. MacKenzie C (2022) Add cito panel to work aspect with ask query. In: GitHub. https://github.com/WDscholia/scholia/commit/0af60cf7732bc8664d828cbe51f233aa63201df9. Accessed 26 Aug 2022 [**cito:citesAsEvidence**]