Two years of CiTO annotations

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, 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 [3].
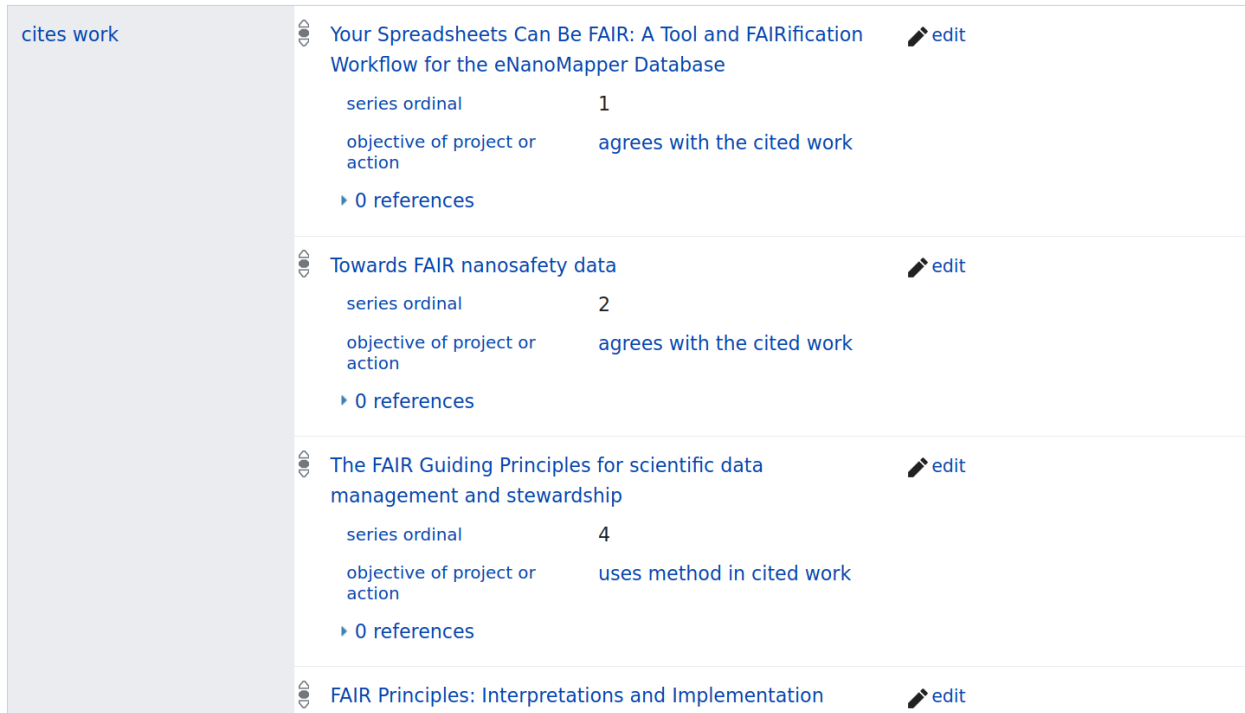
The Citation Typing Ontology Pilot

The pilot consists of a couple of components and the editorial explains some of them [3]. The Citation Typing Ontology was selected to express the intention [1], the intention is expressed a compact identifier wrapped in square brackets [4], and 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 [5]: [**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 [6] and includes citations from OpenCitations [7] 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).

cites work	Your Spreadsheets Can Be FAIR: A Tool and FAIRification Workflow for the eNanoMapper Database	edit
	series ordinal	1
	objective of project or action	agrees with the cited work
	▸ 0 references	
	Towards FAIR nanosafety data	edit
	series ordinal	2
	objective of project or action	agrees with the cited work
	▸ 0 references	
	The FAIR Guiding Principles for scientific data management and stewardship	edit
	series ordinal	4
	objective of project or action	uses method in cited work
	▸ 0 references	
	FAIR Principles: Interpretations and Implementation	edit

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 [8]. 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) [9].

Citation Typing Ontology Intentions

Page with general info about CiTO annotation in Wikidata.

Table of Contents

Statistics

Search:

Count	Description
603	Total number of annotations
494	Total number of annotated citations
387	Total number of articles receiving annotated citations
141	Number of cited venues with annotated citations
98	Total number of articles providing annotated citations
48	Number of citing venues with annotated citations
10	Number of article with explicit CiTO annotation
1	Number of journals with explicit CiTO annotation

Wikidata Query Service   cito-index: statistics.sparql

Figure 2: XXXX.

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.

Adoption by *Journal of Cheminformatics* authors

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* [3]	cito:citesAsAuthority, cito:citesForInformation, cito:usesMethodIn	42%
Methodology	Predicting target profiles with confidence as a service using docking scores [10]	cito:agreesWith, cito:extends, cito:citesAsAuthority, cito:citesAsDataSource, cito:usesMethodIn	47%

Type	Article	Intentions	%CiTO
Research article	Too sweet: cheminformatics for deglycosylation in natural products [11]	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 [12]	cito:agreesWith, cito:citesAsAuthority, cito:citesAsDataSource, cito:discusses, cito:usesDataFrom, cito:usesMethodIn	98%
Database	COCONUT online: Collection of Open Natural Products database [13]	cito:citesForInformation, cito:usesDataFrom, cito:usesMethodIn	41%
Editorial	What is the role of cheminformatics in a pandemic? [14]	cito:agreesWith, cito:citesAsAuthority, cito:citesAsDataSource, cito:citesForInformation, cito:citesAsPotentialSolution	91%
Research article	Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag [15]	cito:citesAsAuthority, cito:citesAsDataSource, cito:citesAsMetadataDocument, cito:discusses, cito:extends, cito:usesDataFrom, cito:usesMethodIn	100%
Software	IDSM ChemWebRDF: SPARQLing small-molecule datasets [16]	cito:citesAsAuthority, cito:citesAsRelated, cito:usesDataFrom, cito:usesMethodIn	97%
Letter to the Editor	FAIR chemical structures in the Journal of Cheminformatics [17]	cito:citesAsAuthority, cito:documents	100%
Letter Response	Reply to "FAIR chemical structure in the Journal of Cheminformatics"	cito:agreesWith, cito:citesAsAuthority, cito:citesAsDataSource, cito:obtainsBackgroundFrom, cito:repliesTo	100%

Technological innovation

Markdown template. . .

To make life easier for authors, and following a Twitter discussion in Spring 2021, we developed a Markdown template with native CiTO support. Here, the author merely indicates the CiTO type when they cite the article.

The manuscript can then easily be saved as a Word file to enable submission to the journal. The *Journal of Cheminformatics* template is available from our GitHub repository, and authors and editors should feel free to adapt it to their own journal needs.

Discussion

Formatting issues

Citation level intention annotation: only Markdown

What is next?

With X 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 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.

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.   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:citesForInformation]**

4.   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]**

5.   (2020) Citation Typing Ontology (CiTO) Pilot

6.   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]**

7.   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

8.   E. W (2021) Typed Citations in the Journal of Cheminformatics. On Physical Sciences **[cito:citesAsDataSource]**

9.   (2022) Citation Typing Ontology Intentions - Scholia **[cito:citesAsDataSource]**

10.   Predicting target profiles with confidence as a service using docking scores. 12: https://doi.org/10.1186/s13321-020-00464-1 **[cito:citesForInformation]**

11.   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]**

12. 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**]

13. 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**]

14. 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**]

15. 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**]

16. 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**]

17. 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**]