

CiTO updates #4: annotations in datasets

Egon Willighagen 

Published April 2, 2023

Citation

Willighagen, E. (2023, April 2). CiTO updates #4: annotations in datasets. *Chem-bla-ics*. <https://doi.org/10.59350/jakew-pe809>

Keywords

Cito, Data, Scholia

Copyright

Copyright © Egon Willighagen 2023. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

Okay, [the Pilot is over](#) ending with 17 papers, 16 of which have CiTO annotations (and so far 4 J.Cheminform. [papers after the pilot](#)), but my interest in the [Citation Typing Ontology](#) continues and we just need [more adoption](#).

Datasets as source of annotations

So, here's a quick [Wikidata](#) update. I have been using Wikidata as infrastructure to collect and share CiTO annotations (see also the below "Scholia patch" posts). Some time ago I recovered my CiteULike CiTO annotations and made this [available on Zenodo](#) (doi:10.5281/zenodo.7368209).

And while thinking about datasets with CiTO annotations, I found two other datasets. One was from an article in Portuguese and one from an [article by Peroni et al.](#) with [this data file](#). That data file is actually a zip, but inside the zip file is a CSV file with three interesting columns: `cited_doi`, `citing_doi`, and `intext_citation.intent`. There are many more columns and I can highly recommend browsing them. But these are the three I need to add data to Wikidata. The third column is free text, but using the CiTO for labels, making it relatively easy to convert to [citation intentions from Wikidata](#) (PS, thanks to [Fvtvr3r](#) for adding more!).

So, I had a cleaned file and started writing a Groovy Bioclipse script using [Bacting](#). It basically does a few things: extract all DOIs, check which ones are in Wikidata, analyze the `intext_citation.intent` column content, and then generate QuickStatements (see [this gist](#)). Out of the 600 lines from the input, it creates some 200 new CiTO-annotated citations in Wikidata between [some 150 article pairs](#):

Supports the following statement(s)

Reload

Statements in Wikidata supported by references to this work. Only a maximum of around 2000 statements are shown.

Show entries

Search:

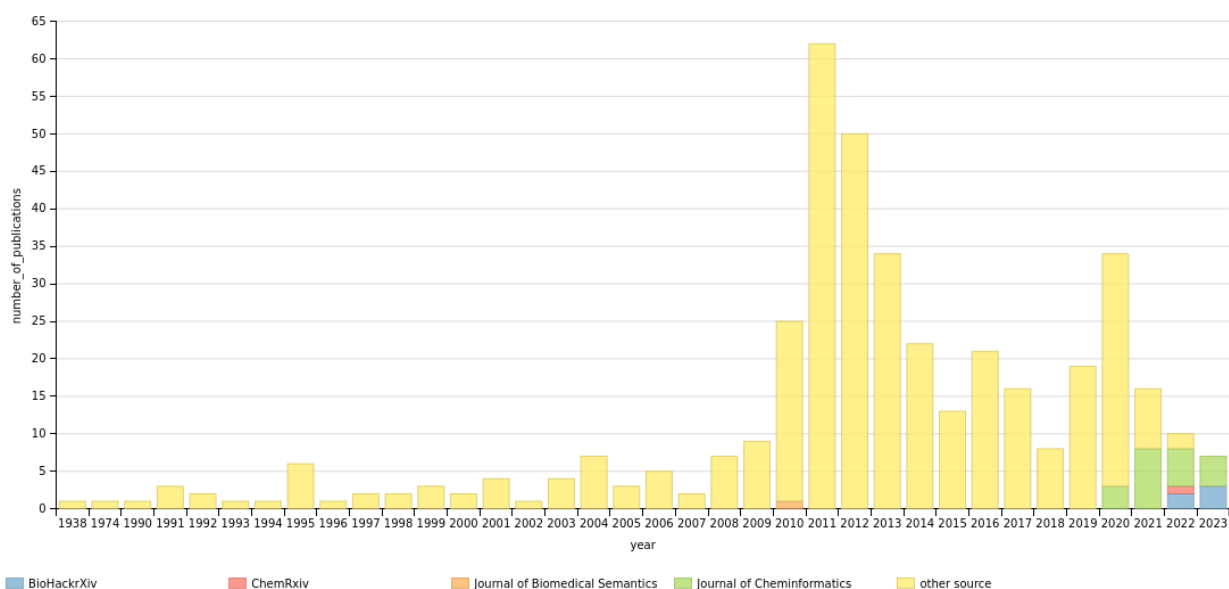
Item	Property	Value	Unit
'Fraternal Socialism': The International Reconstruction of North Korea, 1953–62	instance of	retracted paper	
Why do people spend money to help vulnerable people?	cites work	The Negative Association between Religiousness and Children's Altruism across the World.	
When Words Hurt: Affective Word Use in Daily News Coverage Impacts Mental Health	instance of	retracted paper	
What is the association between religious affiliation and children's altruism?	cites work	The Negative Association between Religiousness and Children's Altruism across the World.	
Using personality variables to predict cancer and heart disease	cites work	Personality, stress and disease: description and validation of a new inventory.	
Use of Modern Contraceptive Methods in Burkina Faso: What are the Obstacles to Male Involvement in Improving Indicators in the Centre-East and Centre-North Regions?	cites work	Religion, culture and male involvement in the use of the Standard Days Method: evidence from Enugu and Katsina states of Nigeria.	
Unpacking religious affiliation: Exploring associations between Christian children's religious cultural context, God image, and self-esteem across	cites work	The Negative Association between Religiousness and Children's Altruism across the World	

chem-bla-ics

The ability to include CiTO annotations from datasets is another welcome boost for the CiTO statistics in Wikidata. [This SPARQL query](#) shows an overview of sources that support the CiTO intention annotation, but note that a claim with a CiTO intention may also have CrossRef, PubMed, and COCI as reference. In those cases, they are primarily for the citations and not the intention.

There are [now](#) (the [latest stats are here](#)) **1202 citation intention** annotations in Wikidata for 992 citations from **405 articles in 199 venues**. Of these 27 articles have explicit annotations in the article itself and are found in 4 venues, two journals and two preprint servers). These annotated citations are to 510 articles in 190 different venues. [This Scholia patch](#) will add a new statistics, the number of datasets providing citation intentions, of which there are (as discussed) [currently two](#) in Wikidata. The latter two provide intentions for the majority of articles and are depicted in yellow in the below overview.

Number of articles with CiTO-annotated citations by year



With an annotation in [an 1938 article by Alan Turing!](#) I ran into this article in November 2011 noting an apparent duplicate title in his article list. I turned out an earlier article had a correction with the same name. I added [this clarification](#):

different from by Egon Willighagen	On Computable Numbers, with an Application to the Entscheidungsproblem 0 references
cites work by Egon Willighagen	On Computable Numbers, with an Application to the Entscheidungsproblem 0 references has goal updates the cited work

This is very trivial citation intention data that publishers could provide as open data.

chem-bla-ics

Okay, that will do for today. There are actually some really interesting things in the pipeline, but I will have to write about that later. I have some deadlines I should start looking at. Below is some extra reading. Some more history

- 2021: [BioHackathon Europe 2021 #1: CiTO annotations in BioHackrXiv](#)
- 2021: [Markdown template for the Journal of Cheminformatics with CiTO support](#)
- 2020: [CiTO updates #3: third paper in the collection and updated Scholia patch](#)
- 2020: [CiTO updates #2: annotation migration to Wikidata and first Scholia patch](#)
- 2020: [CiTO updates #1: first research paper in the Journal of Cheminformatics with CiTO annotation published](#)
- July 2020: [New Editorial: "Adoption of the Citation Typing Ontology by the Journal of Cheminformatics"](#)
- 2015: ["What You're Doing Is Rather Desperate"](#)
- 2012: [CiTO / CiteULike: publishing innovation](#)
- 2010: [CiteULike CiTO Use Case #1: Wordles](#)
- September 2010: [A list of things I miss in CiteULike](#)