

# Citing the Chemistry Development Kit

Egon Willighagen 

Published February 18, 2010

## Citation

Willighagen, E. (2010). Citing the Chemistry Development Kit. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/y34jk-2ds64>

## Keywords

Cdk, Cito, Bioclipse, Jchempaint, Jmol

## Copyright

Copyright © Egon Willighagen 2010. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## chem-bla-ics

Two weeks ago, a paper by Peter Ertl was published about [Molecular structure input on the web](#) (doi:10.1186/1758-2946-2-1). In this paper, he discusses the state of things and describes his contribution to this field, the [JME Molecule Editor](#). The article also cites the CDK, but only the website and not one of the two papers (doi:10.1021/ci025584y, or doi:10.2174/138161206777585274). This is not an isolated case, but a common pattern. In principle, the proper work is cited, and nothing is wrong. Practically it means, that a citation to the [CDK website](#) does not show up in the citation network. This is **not** a problem caused by these papers, but merely by the nature current citation databases work: they only count citations between journal articles, and only sometimes extend to books or conference abstracts.

Now, addressing the limitations of the current citation databases is technically simple, and purely blocked by social and commercial aspects. The Citation Typing Ontology by [David Shotton](#) defines the framework to define citation types, independent from any existing database. The semantic web technologies will take it from there, and allow aggregation etc.

There are some things to think about on how to use such citation networks, though. If we calculate the impact of the CDK project, we should combine citation counts to the website(s), papers, etc, after removal of duplicates, etc. The [cito:cites](#) does link to resources, and the CDK paper resources is not the same as the CDK website resource. But, we could define a Project Class, where both are `foo:partOf`. Then, we could define that *the triple chain the:citingWork cito:cites the:CDKArticle foo:partOf the:CDKProject* would imply *the triple the:citingWork cito:cites the:CDKProject*.

## Typed Citations

Now, while writing up this blog, I realize that my fork of this morning, [A BIBO Citation Typing Ontology](#), might actually be counter-productive in the long run, as I was only working out a solution to a simpler, but different problem, which the CiTO also addresses: a citation is not typed. When a paper does cite the CDK paper, we still do not know if it *uses* the CDK, or merely mentioned it as *related-but-unused*, or even *refuted* work.

Now, as I am leaning towards the Bibliography Ontology as RDF-based system for my references, and been using this already in the [RDF store hosting the ChEMBL data](#), I forked the CiTO to define `rdfs:domain` and `rdfs:range` on [bibo:Document](#). The CiTO 1.5 actually defines a large set of document types too, and I rather see BIBO reused.

This indeed has the downside that the `bibo:cites` cannot be used for the above chaining, and this might bite me seriously later. Well, nothing wrong with a failing experiment, right? For now, it will serve my purpose: setting up a citation database for the CDK project papers.

## The CDK citation database

So, here goes (it's [RDFa-enabled](#); check this [RDF pulled out](#)):

## chem-bla-ics

@prefix bibo: <http://purl.org/ontology/bibo/>.

@prefix bibocto: <http://github.com/egonw/bibo-cto/>.

```
<urn:doi:10.1186/1758-2946-2-1> a bibo:Article ;  
  bibocto:cites <urn:doi:10.1021/ci025584y> .
```

I am not entirely happy about the error-prone XHTML+RDFa of the above example, and [filed a question of better solution](#) on [SemanticOverflow](#).

While the above example merely defines the citation of Peter Ertl's article to the CDK (whether that is valid or not... would he have cited the other paper perhaps?), the citation typing allows me to state how the CDK paper is cited. Now, Peter states:

*It is also gratifying to see the advent of open source movement in cheminformatics on the Internet, as advocated for example by the Blue Obelisk Group (40) and witnessed by collaborative projects like Chemistry Development Kit CDK (41), Jmol (42), Bioclipse (43) and several others.*

So, I think it is fair to state that:

```
<urn:doi:10.1186/1758-2946-2-1> bibocto:credits <urn:doi:10.1021/ci025584y> .
```

which is very much appreciated!