

“What You’re Doing Is Rather Desperate”

Egon Willighagen 

Published March 22, 2015

Citation

Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/ef06k-1xh03>

Keywords

Citeulike, Cito, Publishing, Pubmed

Copyright

Copyright © Egon Willighagen 2015. 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.

One blog that I have been (occasionally but repeatedly) reading for a long time is the [What You're Doing Is Rather Desperate](#) blog by Self-portrait by [Gustave Courbet](#). HT to [WoW!ter](#) for pointing me to this nice post where Saunders shows [how to calculate the number of entries in PubMed marked as retracted](#) (in two separate ways). Source: Wikipedia. [CC-BY-SA](#) The spoiler (you should check his R scripts anyway!): about 3900 retractions.

I have been quite interested in this idea of retractions and recent discussions with Chris ([Christopher](#): mostly offline this time, and even some over a beer :) about if retractions are good or bad (BTW, check [his inaugural speech on YouTube](#)). He points out that retractions are not always for the right reasons, and, probably worse, have unintended consequences. An example he gave is a more senior researcher with two positions; in one lab someone misbehaved and this lab did not see any misconduct of this senior researcher; however, his other affiliation did not agree and fired him.

Five years ago I would have said any senior researcher on a paper should still understand things in detail, and if misconduct was found, the senior authors are to blame too. I still believe this is the case, that's what you're co-researcher for, but feeling the pressure of publishing enough and just not having enough time, I do realize I cannot personally reproduce all results my post-docs and collaborators do. But we all know *that* publishing has made a wrong turn and, yes, I am trying to make it return to a better default.

Why I am interested in retractions

But that is not why I wanted to blog and discuss Saunders post. Instead, I want to explain why I am interested in retractions and, another blog you should check out, [Retraction Watch](#). In fact, I am very much looking forward to their database! However, this is not because of the blame game. Not at all.

Instead, I am interested in noise in knowledge. Obviously, because this greatly affects my chemblaics research. Particularly, I like to reduce noise or at the very least take appropriate measures when doing statistics as we have plenty of means to deal with noise (like cancelling it out). Now, are retractions then a appropriate means to find incorrect, incomplete, or just outright false knowledge? No. But there is nothing better.

There are better approaches: I [have long and still am advocating](#) the [Citation Typing Ontology](#), though I have to admit I am not up to date with [David Shotton](#)'s work. The CiTO allows to annotate if two papers disagree or if it agrees and perhaps even uses the knowledge. It can also annotate the citation as merely being included because the cited paper has some authority (expect many of those to Nature and Science papers).

But we have a long way to go before using CiTO becomes a reality. If interested, please check out [the CiteULike support](#) and Shotton's [OpenCitations](#).

What does this mean for databases?

chem-bla-ics

Research depends on data, some you measure, some you get from literature and increasingly databases. The latter, for example, to compare your own results with other findings. It is indeed helpful that databases provide these two functions:

1. provide means to find similar experiments
2. provide a gold standard of true knowledge

These database will have two different approaches: the first will present the data as reported or even better as raw data (as it came from the machine, unprocessed, though increasingly the machines already do processing to the best of its knowledge); the second will filter out true facts, possibly normalizing data along the way, e.g. by correcting obvious typing and drawing errors.

Indeed, database can combine these features, just like [PubChem](#) and [ChemSpider](#) do for small compounds. PubChem has the explicit approach of providing both the raw input from sources (the substances with SIDs) and the normalized result (the compounds with CIDs).

But what if the original paper turned out to be wrong? There are (at least) two phases:

1. can we detect when a paper turns out wrong?
2. can we propagate this knowledge into databases?

The first clearly reflects my interest in CiTO and retractions. We must develop means to filter out all the reported facts that turn out to be incorrect. And, can we efficiently keep our thousands of databases clean (many valid approaches!)? Do retractions matter here? Yes, because research in so-called higher impact journals is also seeing more retractions (whatever the reason is for that correlation), see [this post by Bjoern Brembs](#).

Where we must be heading

What the community needs to develop in the next few years is approaches for propagation of knowledge about correct and incorrect knowledge. That is, high impact knowledge must enter databases quickly, e.g. the exercise myokine [irisin](#), but also the fact that it was recently shown it very likely doesn't exist, or at least that the original paper most likely measured something else (doi:[10.1038/srep08889](#)). Now, this is clearly a high-profile "retraction" of facts that few of us will have missed. Right? And that's where the problem is, the very long tail of disproven knowledge is very long, and we cannot rely on such facts to propagate quickly if we do not use tools to help us. This is one reason why my attention turned to semantic technologies, so that contradictions can be found more easily.

But I am getting rather desperate about all the badly annotated knowledge in databases, and I am also getting desperate about being able to make a change. The research I'm doing may turn out rather desperate.

References

Albrecht, E., Norheim, F., Thiede, B., Holen, T., Ohashi, T., Schering, L., Lee, S., Brenmoehl, J., Thomas, S., Drevon, C. A., Erickson, H. P., & Maak, S. (2015). Irisin – a myth rather than an exercise-inducible myokine. *Scientific Reports*, 5(1). <https://doi.org/10.1038/srep08889> [CrossRef](#) [Scholia](#)