

# Curation is an essential part of doing research

Egon Willighagen 

Published June 29, 2025

## Citation

Willighagen, E. (2025). Curation is an essential part of doing research. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/c4k5q-h8849>

## Keywords

Curation, Openscience, Nmrshiftdb

## Abstract

Depending on your exact definition of doing science, keeping track as precise as possible of your observations is an essential part of doing science. The precision should be high enough that mistakes are obvious. This pattern is, of course, not limited to doing science and we see this in open source development too. Unfortunately, in the modern way of doing science, this is not getting the attention it should get.

## Copyright

Copyright © Egon Willighagen 2025. 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.

Depending on your exact definition of doing science, keeping track as precise as possible of your observations is an essential part of doing science. The precision should be high enough that mistakes are obvious. This pattern is, of course, not limited to doing science and we see this in open source development too. Unfortunately, in the modern way of doing science, this is not getting the attention it should get. Worse, with narratives (stories) about the research, in the form of journal articles, are generally considered more important than a precise description of the observations.

Is that a big issue? Hell, yes. Where do you think the FAIR ideas came from? And why FAIR in ten years has not brought about the change it was hoping for?

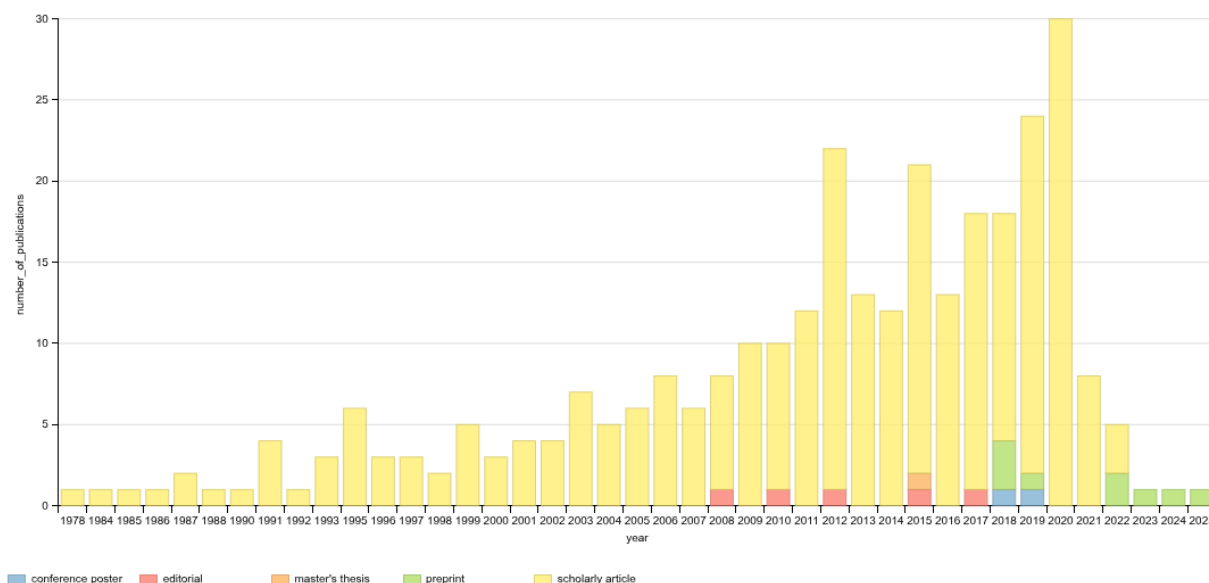
For me, my fascination for curation started as a student, around 1995, with the *Dictionary on Organic Chemistry*. At that time, my interest came from wanting to learn about chemistry and biology. During my M.Sc. and PhD, it was obvious how essential it was to deriving correct scientific conclusions from your experiment. Data, knowledge, and software alike, imo. And because curation is expensive, not having to repeat it, I prefer to do it as Open Science.

## Curation

Of course, curation has been part of doing science, but to a large extent is a separate step from doing science. It is done by database developers, librarians, and chemo- and bioinformaticians. For example, Chemical Abstracts Service (CAS) [started over 100 years ago](#) and started indexing chemical structures in 1965. The curation is an ongoing process, [also for old records](#).

[Biocuration](#) is getting [more and more attention](#):

### Publications per year



The recognition and rewarding by having the [International Society for Biocuration](#) (ISB, [Scholia page](#)) should not be underestimated (doi:[10.1038/455047A](#)). Their [Annual International](#)

[Biocuration Conferences](#) have been running since [2005](#). And with their awards, they give the biocuration work recognition and, literally, rewarding:

- [Biocuration Career Award](#) (2016-2021)
- [Excellence in Biocuration Early Career Award](#) (2022-)
- [Excellence in Biocuration Advanced Career Award](#) (2022-)
- [Exceptional Contribution to Biocuration Award](#) (2017-)

## My curation Curriculum Vitae

I don't have a good *curation CV*. For a large extend because the curation has been part of a study. The curation itself does not get recognized, and only the *journal article* does. With datasets slowly getting more recognition, so does data curation, but data curation is not really part of how we do FAIR at this moment, and via this route not getting the attention it gets.

But since I have been updating [my CV anyway](#), I dug up some curation I am proud of:

- the Dictionary on Organic Chemistry, which no longer exists, but it started my Open Science chemistry research
- the [Blue Obelisk Data Repository](#) (BODR), which has been part of various GNU/Linux distributions (see also doi:[10.1021/ci050400b](#)). A new version is [long overdue](#)
- I contributed hundreds of NMR spectra with uncommon nuclei to [NMRShiftDb](#)
- Wikidata, see [this preprint](#), but also many small projects, like adding CXSMILES for polymers, and [main subject annotation in Scholia](#)
- WikiPathways (see [these blog posts](#)), where I started [curating metabolites in 2012](#), set up [a computer-assistent curation platform using SPARQL](#), and were an early curator of [SARS-CoV-2 biological processes](#)
- citation intent annotation with the Citation Typing Ontology, see this [Scholia overview](#)
- nanosafety ontology and data: the [eNanoMapper Ontology](#) (ENMO), [NanoWiki](#), [JRC nanomaterial index](#) and [the ERM indentifier database](#)
- made RDF for supplementary information (e.g. [this NanoE-Tox spreadsheet](#), full databases, like [ChEMBL](#) and [NMRShiftDb](#)
- organized [an online ChemCuration event](#) (inspired by the ISB annual meetings!)

I am also curation my blog, which was [originally in blogger.com but being ported to Markdown with extra annotation](#). That includes [updating URLs](#) and annotation of blog posts [with chemicals, grants](#), and [intention-typed citations](#).

## Long tail

Of course, I have my Wikipedia edits, and contributed to projects like [Bioregistry.io](#), [FAIRsharing](#), regularly submit [missed mentions to Altmetric.com](#), etc. There is a long tail in curation. And there is a lot of curation hidden in [my literature list](#).

And that long tail matters to me. I want every researcher to pick up the challenge to curate their own research output. Put your experimental data in databases, add important provenance, get

## chem-bla-ics

the details rights. This is essential to reduce the cost of doing research, and that is more important than ever.

BTW, I must note that our bioinformatics team colleagues too have done a tremendous amount of biocuration, in WikiPathways ([Denise](#), [Freddie](#), [Susan](#)), in nanosafety ([Jeaphianne](#), [Ammar](#)), and in toxicology ([Marvin](#)), just to name a few. Often together with B.Sc. and M.Sc. students (which [can work really well](#)).

## Award nomination

And I hope this makes it clear why I am delighted to was [nominated last week](#) for an ISB *Excellence in Biocuration Advanced Career Award*. The list of past awardees is impressive, as are the other nominations: [Laurel Cooper](#), Oregon State University/USA, [Steven Marygold](#), University of Cambridge/UK, [Saurabh Raghuvanshi](#), University of Delhi/India, and [Kimberly Van Auken](#), California Institute of Technology/USA.

It's an honor to be listed along these other nominees and being nominated is a great recognition! With a *thank you* to the person who proposed my nomination.