

A convincing example of the need for data repositories. FAIR Data.

Henry Rzepa¹

¹Affiliation not available

April 17, 2023



A convincing example of the need for data repositories. FAIR Data.

HENRY RZEPA

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:

DATE RECEIVED:
June 11, 2015


DOI:
10.15200/winn.142313.30279

ARCHIVED:
February 05, 2015

CITATION:
Henry Rzepa, A convincing example of the need for data repositories. FAIR Data., *The Winnower* 2:e142313.30279, 2015, DOI: 10.15200/winn.142313.30279

© Rzepa This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



 Derek Lowe in his [In the Pipeline](#) blog is famed for spotting unusual claims in the literature and subjecting them to analysis. This one is entitled [Odd Structures, Subjected to Powerful Computations](#). He looks at this image below, and finds the structures represented there might be a mistake, based on his considerable experience of these kinds of molecules. I expect he had a gut feeling within seconds of seeing the diagram.

Indeed, so, you will now find that the [authors have apparently acknowledged](#) a mistake^[1]. My interest piqued, I went to the article, and immediately tracked down the [supplementary information](#). Surely, if these molecules had been subjected to *powerful computation*, this supporting information should contain coordinates of some kind that would allow a correlation with the 2D structural representation shown above. I have just returned from [FORCE2015](#), a three-day event in Oxford. From the detailed agenda, you can see that a lot of the conference centered around what is called [FAIR Data](#). FAIR stands for:

1. [Findable](#)
2. [Accessible](#)
3. [Interoperable](#)
4. [Re-usable](#)

So I then set out to find if the [supplementary information](#) WAS FAIR. Well, check for yourself (unlike the narrative article, the data should be accessible outside of the paywall, *i.e.* you should not need a subscription to access it). It is certainly big, running out to 45 pages, in the form of a paginated PDF file (the norm). The table of contents does not refer to data as such, but it does quote 25 figures, from which you might just be able to extract some data. But no molecules as such! So:

1. No data is findable, although the PDF which might contain it is reasonably so.
2. The data is not easily accessible,
3. let alone interoperable (thus many of the charts were probably created using spreadsheet software, but the source files for these are not available),
4. and not-reusable (certainly not without loss and possible error in any attempt at capture).

I think it fair to say that the data for these powerful computations are not FAIR. Had we had at least some coordinates (the computations involved molecular mechanics based dynamics simulations, which certainly involve manipulating atom coordinates in some form) then the structures shown in the figure above could be checked, and perhaps even the apparent error would have been quickly spotted.

Derek does not make the point about FAIR data (to be fair, he was not at FORCE2015) and so I will make the case. If you are reporting a computational model or simulation, there is no excuse for not supplying FAIR data to accompany it. If the data is FAIR it will be inter-operable and re-usable. And this will instantly allow anyone to check *e.g.* the structures above. You would not need to have Derek's vast experience and instinct (although having it is also helps). And of course we might presume that there were 2-3 referees that also looked at the article, and presumably none of them requested FAIR data.

Oh, if you are interested in my take on FAIR data, I gave [a talk about](#) that at FORCE2015, which you are welcome to view; I hope it constitutes a FAIR talk!

REFERENCES

1. K.J. Kohlhoff, D. Shukla, M. Lawrenz, G.R. Bowman, D.E. Konerding, D. Belov, R.B. Altman, and V.S. Pande, "Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways", *Nature Chem*, vol. 6, pp. 15-21, 2013. <http://dx.doi.org/10.1038/nchem.1821>
<!-- kcite-section 13248 -->