

# FAIR data $\rightleftharpoons$ Raw data.

Henry Rzepa 

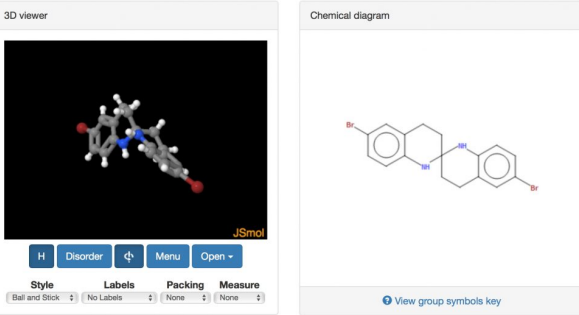
Published December 7, 2017

## Citation

Rzepa, H. (2017). FAIR data  $\rightleftharpoons$  Raw data. *Henry Rzepa's Blog*. <https://doi.org/10.59350/whf3t-b6v86>

## Keywords

Chemical IT, Crystal\_structure\_mining, Computing, Context, Data




3D viewer

Chemical diagram

Additional details

Deposition Number	1528662
Data Citation	Joshua Almond-Thynne, Andrew J. P. White, Anastasios Polyzos, Henry S. Rzepa, Philip J. Parsons, Anthony G. M. Barrett CCDC 1528662: Experimental Crystal Structure Determination, 2017, DOI: 10.5517/ccdc.csd.cc1n9ppb
Deposited on	20/01/2017
Raw data DOI(s)	<a href="https://doi.org/10.14469/hpcr/2300">10.14469/hpcr/2300</a>

Associated publications

 Joshua Almond-Thynne, Andrew J. P. White, Anastasios Polyzos, Henry S. Rzepa, Philip J. Parsons, Anthony G. M. Barrett, *ACS Omega*, 2017, 2, 3241, DOI: 10.1021/acsomega.7b00482

## Copyright

Copyright © Henry Rzepa 2017. 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.

## Henry Rzepa's Blog

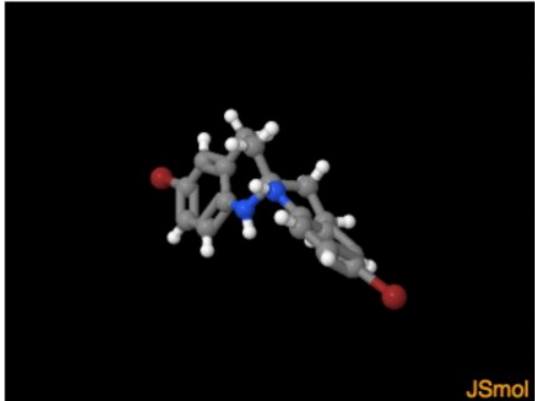
**FAIR data** is increasingly accepted as a description of what research data should aspire to; **F**indable, **A**ccessible, **I**nter-operable and **R**e-usable, with **C**ontext added by rich metadata (and also that it should be **O**pen). But there are two sides to data, one of which is the raw data emerging from say an instrument or software simulations and the other in which some kind of model is applied to produce semi- or even fully processed/interpreted data. Here I illustrate a new example of how both kinds of data can be made to co-exist.

I will start with a recent publication[cite]10.1002/anie.201407165[/cite] with the title \*Crystallographic Snapshot of an Arrested Intermediate in the Biomimetic Activation of CO<sub>2</sub>. \*The nature of this intermediate caught the eye of another research group, who responded with their own critique[cite]10.1002/anie.201411654[/cite]‡ along with the comment “\*However, since we have no access to the original crystallographic data ...” \*They might have been referring to the semi-processed data (containing the so-called *hkl* structure factors) but they may also have been alluding to the **raw image data** captured directly from the diffractometer cameras. That traditionally has not been available via the CSD (Cambridge structural database), but would be required for a complete re-analysis of the crystal structure. Now the first example of how both FAIR (processed) data and raw data can co-exist has appeared.

The latest version of the CSD database shows an entry resulting from the following publication[cite]10.1021/acsomega.7b00482[/cite] and the deposited data has its own DOI there ([10.5517/ccdc.csd.cc1n9ppb](https://doi.org/10.5517/ccdc.csd.cc1n9ppb)). That entry in turn has a DOI pointer to the **Raw data** ([10.14469/hpc/2300](https://doi.org/10.14469/hpc/2300)) held in a different location and the pointer is reciprocated (⇌) with the latter pointing back to the former. Both datasets point to the original article, thus completing a holy triangle.†

## Henry Rzepa's Blog

### 3D viewer



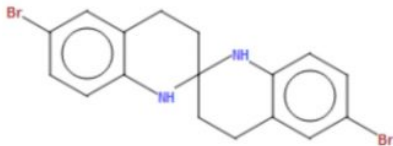
JSmol

H Disorder ↕ Menu Open ▾

Style Labels Packing Measure

Ball and Stick ↕ No Labels ↕ None ↕ None ↕

### Chemical diagram



[View group symbols key](#)


### Additional details

<b>Deposition Number</b>	1528662
<b>Data Citation</b>	Joshua Almond-Thynne, Andrew J. P. White, Anastasios Polyzos, Henry S. Rzepa, Philip J. Parsons, Anthony G. M. Barrett CCDC 1528662: Experimental Crystal Structure Determination, 2017, DOI: <a href="https://doi.org/10.5517/ccdc.csd.cc1n9ppb">10.5517/ccdc.csd.cc1n9ppb</a>
<b>Deposited on</b>	20/01/2017

### Raw data DOI(s)

<b>DOI</b>	<a href="https://doi.org/10.14469/hpc/2300">10.14469/hpc/2300</a>
------------	---

### Associated publications

 Joshua Almond-Thynne, Andrew J. P. White, Anastasios Polyzos, Henry S. Rzepa, Philip J. Parsons, Anthony G. M. Barrett, *ACS Omega*, **2017**, **2**, **3241**, DOI: [10.1021/acsomega.7b00482](https://doi.org/10.1021/acsomega.7b00482)

There is more. The Raw dataset ([10.14469/hpc/2300](https://doi.org/10.14469/hpc/2300)) declares it is a member of a superset, called *Crystal structure data for Synthesis and Reactions of Benzannulated Spiroaminals; Tetrahydrospirobiquinolines* ([10.14469/hpc/2297](https://doi.org/10.14469/hpc/2297)\*) where you can find information about six other related structures. That collection is in turn a member of a superset called *Synthesis and Reactions of Benzannulated Spiroaminals; Tetrahydrospirobiquinolines* ([10.14469/hpc/2099](https://doi.org/10.14469/hpc/2099)) \*where DOIs to other types of data associated with this project can be found, such as *Computational data* ([10.14469/hpc/2098](https://doi.org/10.14469/hpc/2098)) and *NMR data* ([10.14469/hpc/2294](https://doi.org/10.14469/hpc/2294)). Although a human can with some determination follow these associations up, down and across, the system is designed to also be followed by automated algorithms that could traverse this web quickly and efficiently.

So you can now see that a crystal structure held in the CSD could be the starting point for a journey of FAIR data discovery, in manner that has not hitherto been possible. How quickly the CSD will become populated by links to **Raw** (and other) **data** remains to be seen. I have not yet

## Henry Rzepa's Blog

discovered any mechanism for specifying a CSD query which stipulates that Raw data must be available, but no doubt this will come.

To end, back to the *Biomimetic Activation of CO<sub>2</sub>* referred to at the start. With no access to the original data, recourse was made to computational modelling.[cite]10.1002/anie.201411654[/cite] Which where I came in, since I wanted access to the original (computational) data. Sadly it did not appear to be available with the article,[cite]10.1002/anie.201411654[/cite] in much the same manner as the original complaint. Perhaps, when FAIR data becomes fully accepted as part of how science is done nowadays, such complaints will become ever rarer!

---

‡In fact the original authors did respond[cite]10.1002/anie.201504197[/cite] with an acknowledgement that their original conclusions were not correct.

†Almost. The article [cite]10.1021/acsomega.7b00482[/cite] cites DOI: [10.14469/hpc/2099](https://doi.org/10.14469/hpc/2099) (Ref 28), but it does not cite DOI: [10.5517/ccdc.csd.cc1n9ppb](https://doi.org/10.5517/ccdc.csd.cc1n9ppb) because the latter had not been minted yet at the time the final proofs were corrected, and there is no mechanism to add it at a later stage.

<https://orcid.org/0000-0002-8635-8390>