

Opinion

Fair enough: A sideways look

Jeremy G. Frey^{1,2,*}

¹School of Chemistry, University of Southampton, Southampton, SO17 1BJ, UK

²Twitter: @profechem

*Correspondence: j.g.frey@soton.ac.uk

<https://doi.org/10.1016/j.patter.2021.100331>

Even when we all agree to make data usefully available with our publications, the problem remains of how to do this in a reasonable and sensible manner. Just what do we need to do? I suggest a change of perspective: what would you wish other authors had done to make their data easily useful to us?

When asked by my colleagues, “What exactly is needed to comply with data management policies and to make my data FAIR [findable, accessible, interoperable, and reusable]?”¹ my short answer is in line with the Golden Rule, “Do unto others how you would wish them to do unto you.”

My longer reply is along these lines: if you read about an interesting experiment in a paper in your research area, and you want to use the result and incorporate the experiment in your own work, what information, what details would you need the authors to have provided? Go ahead and provide that level of detail.

It might be better even to suggest that the colleague reflects on when they need information from a paper not in their immediate area, where they are less acquainted with all the fine details of the research, what information would they need (hope) to find?

The next level of this spiral argument would be to think about what details would a new graduate student working in the area need to be able to locate the paper in the first place and then extract the necessary information to be able to (correctly) use the outputs of your research (and of course acknowledge the source).

But at what point should this stop? Does the paper need a whole undergraduate module to be followed first before it can be understood? Probably, and perhaps we should provide a link to potential courses as part of the background? But surely this is taking things to an extreme, unless the research is in a novel area, when it might be necessary and indeed worthwhile to go to this amount of work to ensure your research does

have the opportunity to create impact in the wider(est) community.

How to decide when is the appropriate point to halt this spiral where the contextual information could easily overwhelm the new results? The Royal Society Report on Science and an Open Enterprise² highlighted the concept of “intelligently open” to reflect the necessary interplay between the author and the reader, to reflect open communication. Without this concept or something similar, one could either ignore the need to provide the necessary context, or in a revision of the famous English oxymoron “fair unfairly kept them falsely true,” by too great an attention to the detail of the letter of the data management plans, one could miss the whole aim of the open science ideal.

At conferences over the last few years, I have noticed that industry interest in data sharing (and thus uptake of FAIR) has been much greater than in the academic world. Initially this might seem surprising—commercial interests are usually viewed and benefit from closely guarding access to data and information, while academia is about contributing to the scientific literature. However, industry is also often about a team effort to achieve and aim—industries need to work together, and any barrier to integrating the team efforts has a direct commercial cost. This makes very clear the benefit of easy exchange of data and the costs of failing to do so. In contrast, the academic environment can drive (though not necessarily) a very personal perspective. Not only should this provide an example to universities but highlight the need to train the next generation of researchers to take their place in these industrial teams.

Having worked out what data to present to one’s colleagues to ensure your work is understandable and repeatable, then one of the most difficult things to establish is whether the format (file format) we use for our data are sufficiently universally understood so that they can in fact make use of the data. The FAIRSharing project³ is a useful resource in this regard.

The ubiquitous CSV file provides a simple data management example that was brought home to me when teaching Python (I am usually an R programmer). It is great that the file format has the ability (not always used) to give sensible names to each column of data, and this clearly helps the use and re-use of the data. But when teaching, I realized that I was falling into the trap of “do as I say not what I do”—all data files should have at least a title within them so that even when the file name is changed, it will still be clear what the data in the file is about. What is the catch? Well, all the obvious routines for reading CSV files don’t consider the possibility of a title and comments as well as column headers. This is easy to fix in a program but changes the standard.

What would I like to see happen? What we need to achieve is an enhancement to the research culture, at the start of the UK e-Science program we suggested in Publication @ Source^{4–6} that the mindset of those undertaking research should, right from the start, have the view that others will use their work and so plan, collect, and record with this in mind. I was inspired by the 2014 BBC Reith Lectures given by Dr. Atul Gawande⁷ in discussing complex medical systems and how to ensure high-quality outcomes. Providing and explaining incentives (encourage



and enforcement (Medieval punishment) to push adoption only go so far; proper design is critical in imbedding a culture change. In his book *The Checklist Manifesto*,⁸ he demonstrates the importance of having a process for whatever you are doing to ensure these quality outcomes.

In many ways my own areas of research in physical sciences are simple. Once a decision to publish is made, there is little reason not to make all the data available, and indeed, doing so will help with assigning credit where credit is due, something that has often been overlooked. However, in my wider work with the digital economy, with the need to obtain ethical approval, I am well aware that in many research areas, for example those involving people, even if only expressing simply personal views, exposing data, even with some degree of anonymization, can be very problematic for the communities involved.

As part of the international standards work, IUPAC is working on recommendations for FAIR Chemistry (see <https://www.fairsfair.eu/> and specifically <https://www.fairsfair.eu/recommendations-practice-support-fair-data-principles> and <https://www.go-fair.org/fair-principles/>), because while high-level guidelines do exist (for examples see https://iupac.org/wp-content/uploads/2020/10/CPCDS_Shots_JulyAug2020.pdf and https://iupac.org/projects/project-details/?project_nr=2019-031-1-024), useful applicable guidelines at a disciplinary level are lacking but are essential to the digital transformation of IUPAC and chemistry as a whole.^{9,10} We need to start from professional understanding of the necessity to collaborate and communicate, and following Gawande, we need clear processes. Reflecting on what we would

want others to do for us, is enormously helpful in thinking what these processes should be.

Have I always followed these ideals? No, but I am trying!

ACKNOWLEDGMENTS

Title inspired by the BBC Radio 4 series “Side-ways” by Matthew Syed (<https://www.bbc.co.uk/programmes/m000s2kt/episodes/player>). I would like to thank all my research group and my academic and industrial colleagues for informative discussions and the inspiration that led to these thoughts and UKRI for the funding to support our work, particularly the e-Science and Digital Economy grants (EP/S000356/, EP/K003569/, EP/C008863/, GR/R67729/)

DECLARATION OF INTERESTS

J.G.F. is a member of the *Patterns* advisory board, and he has interacted with *Patterns* in the organization of some meetings. He is a member of the editorial boards of the *Journal of Chemical Informatics* and *Scientific Data*. His research has been funded largely by UK Research Councils with some industrial collaboration and funding, but he does not believe this would be cause for any conflicts of interest.

WEB RESOURCES

Committee on Publications and Cheminformatics Data Standards August 2 Update, https://iupac.org/wp-content/uploads/2020/10/CPCDS_Shots_JulyAug2020.pdf

FAIR Principles, <https://www.go-fair.org/fair-principles/>

FAIRsFAIR, <https://www.fairsfair.eu/>

International Union of Pure and Applied Chemistry, https://iupac.org/projects/project-details/?project_nr=2019-031-1-024

Recommendations on practice to support FAIR data principles, <https://www.fairsfair.eu/recommendations-practice-support-fair-data-principles>

REFERENCES

1. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data manage-

ment and stewardship. *Sci. Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

2. The Royal Society (2012). Final Report – Science as an open enterprise. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>.
3. FAIRsharing. <https://fairsharing.org/>.
4. Frey, J.G., De Roure, D., and Carr, L. (2002). Position Paper: Publication at Source: Scientific Communication from a Publication Web to a Data Grid (EuroWeb).
5. Frey, J. (2008). The curation of laboratory experimental data as part of the overall data lifecycle. *Int. J. Digit. Curation* 3. Published online December 2, 2008. <https://doi.org/10.2218/ijdc.v3i1.41>.
6. Frey, J., Coles, S.J., Bird, C., and Willoughby, C. (2015). Collection, Curation, Citation at Source: Publication@Source 10 Years On. *Int. J. Digit. Curation* 10. Published online June 16, 2015. <https://doi.org/10.2218/ijdc.v10i2.377>.
7. BBC (2021). Dr Atul Gawande - 2014 Reith Lectures. <https://www.bbc.co.uk/programmes/articles/6F2X8TpsxrJpnsq82hggHW/dr-atul-gawande-2014-reith-lectures>.
8. Gawande, A. (2011). *The Checklist Manifesto: How to Get Things Right* (Picador).
9. Frey, J.G. (2014). Digital IUPAC: A Vision and a Necessity for the 21st Century. *Chem. Int.* Published online January 25, 2014. <https://doi.org/10.1515/ci.2014.36.1.14>.
10. McEwen, L.R. (2020). Towards a Digital IUPAC. *Chem. Int.* Published online April 30, 2020. <https://doi.org/10.1515/ci-2020-0203>.

About the author

Jeremy G. Frey obtained his PhD in the University of Oxford with Prof. Brian Howard followed by a NATO/SERC fellowship at the Lawrence Berkeley Laboratory with Prof. Yuan T. Lee. In 1984, Jeremy took up a lectureship at the University of Southampton, where he is now professor of physical chemistry. Jeremy is an enthusiastic supporter of interdisciplinary research. His experimental research probes molecular organization laser spectroscopy and imaging from the IR to soft X-rays. In parallel he investigates how e-Science infrastructure make intelligent laboratories. This blends into a computational and theoretical focus on chemical informatics and the application of novel mathematical approaches to chemical modeling. Jeremy is the PI of the EPSRC Network+ on Artificial Intelligence and Automated Scientific Discovery, encouraging the collaborations at the cutting edge of AI and chemical sciences.