

A Survey About Raw Data Archival and Reuse in Chemical Crystallography

It is now common to deposit structure factors when publishing, which means that the small molecule crystallography community caters very well for routine structures. However this is generally only the case if everything in a raw image is fully and/or properly accounted for and the model is correct or appropriate. So for example, in some cases raw data may no longer be required, while in others it may be necessary to validate or 'do better' in the future. Moreover there are increasing pressures from bodies e.g. funders to make the data relating to research outputs Findable, Accessible, Interoperable and Reusable (FAIR). In acknowledgement of this situation and in order to begin addressing it, IUCr Journals now facilitate access to and citation of large raw diffraction datasets in its articles. Therefore it is important for our community understand and define how we manage our Raw Data* in this respect.

As Members of the IUCr Committee on Data we see the need to conduct this survey about exploring raw data archival practice and gathering opinions as to if/how raw data could/should be used if it were to be made more widely available.

* For the purposes of this survey work, we define Raw Data as a collection of single crystal diffraction images, along with the associated files and metadata necessary to interpret them.

The IUCr Committee on Data



1. If you don't archive raw data, what are the main reasons (please tick all that apply)? If you do archive your raw data, please proceed straight to the next question.

Check all that apply.

- No infrastructure
- Not technically adept
- Not allowed by institution
- It's not necessary
- No budget for this

Other: _____

2. Where do you keep your raw data (please tick all that apply)?

Check all that apply.

- The computer connected to the diffractometer
- Removable media e.g. DVD, tape, USB hard drive
- A computer / filestore in my laboratory
- A computer / filestore in my office
- A university/organization datastore
- Online (cloud) datastore specifically for Higher Education
- Commercial online (cloud) datastore e.g. Amazon

Other: _____

3. Do you manage your archive (please tick the closest option or give an alternative approach in the 'other' section)?

Mark only one oval.

- No – it simply accumulates
- I use solely a 'live' archive (where data can be readily accessed)
- I use solely a 'dark' archive (where data is securely stored but requires time or effort to retrieve it)
- I use both a 'live' archive for recent data and a 'dark' archive for old data which doesn't need to be readily accessed
- I periodically group raw data together e.g. by year, by person
- I actively review whether historic data is required any more (and delete?)
- I assess whether it is possible to actually do anything with old data
- Other: _____

4. Is your raw data searchable (please tick the closest option and/or provide more details)?

Mark only one oval.

- By browsing only
- By filename/sample identifier only
- By high-level terms eg examiner name, date
- By a range of terms, including some relating directly to the data eg unit cell parameters
- Other: _____

5. Is your raw data externally available?

Mark only one oval.

- Yes (e.g. open access; collaborators; other researchers requesting access)
- No

6. If it were to become Policy to make funded research raw diffraction data accessible after 3 years since measurement, would you endeavour to comply?

Mark only one oval.

- Yes, I would do this for all my raw data
- Yes, but I would only do so in cases where I had to
- No

7. Do any of the following organisations who might have a controlling hand in your research have a policy that mandates you manage and/or share your raw data?

Check all that apply.

- The research funder
- My Higher Education Institution e.g. as implemented by the Library
- My employer (not a HEI)
- My department e.g. Grad School for theses
- A Publisher that I submit to
- I am not aware of there being a raw data policy/mandate for any aspect of what I have to do

Other: _____

8. Would you anticipate being able to pay for external archive/repository services and facilities e.g. through research grants or institutional funding

Mark only one oval.

- Yes
- No
- Maybe

9. How likely is it that you will need to revisit raw data in your own work?

Mark only

Very likely 1 2 3 4 5 Highly unlikely

one
oval.

_____      _____

10. How likely is it that you will need to examine raw data when reviewing someone else's work?

Mark
only
one
oval.

Very likely _____ 1  2  3  4  5  _____ Highly unlikely

11. If you had access to a repository containing raw data, what would you want to use it for (please tick all that apply)?

Check all that apply.

- Training datasets
- Software/algorithm/methods development
- To validate a result that you are reviewing
- To validate a result that you wish to use / incorporate in your own work
- To better understand a complicated result
- For a new/extended study on a complicated result

Other: _____

12. There are some cases where publishing raw data may be necessary. Please tick all the cases that you agree would be worthwhile.

Check all that apply.

- Validation: a result provides a contribution to chemical knowledge, but is poor quality
- Validation: to support a 'grand' claim
- To back up modelling of disorder, twinning, incommensurate, modulated structures
- To back up modelling of diffuse scattering
- To make available disorder, twinning, incommensurate, modulated, diffuse scattering datasets so others can attempt to resolve them
- To support 'Advanced Experiments' e.g. charge density, high pressure, phase transition, gas environment, excited states
- When it is clear that future improvement may be possible through developments in software & modelling
- Training sets/benchmarking for software/methods developers

Other: _____

13. The description that best fits my role is:

Mark only one oval.

- Staff/service crystallographer at an academic research institution
- Principal Investigator Faculty member at a research institution
- Research faculty (non-tenure track) at an academic research institution
- Emeritus faculty member at an academic research institution
- Graduate student or post-doc at an academic research institution
- Staff/service crystallographer at an industrial/commercial organisation
- Researcher at an industrial institution
- Employee at a crystallographic software/equipment vendor company
- Other: _____

14. I come from

Mark only one oval.

- North America
 - South and Central America
 - Europe (including Russia)
 - Middle East
 - Africa
 - Asia
 - Australia, New Zealand and South Pacific territories
-

