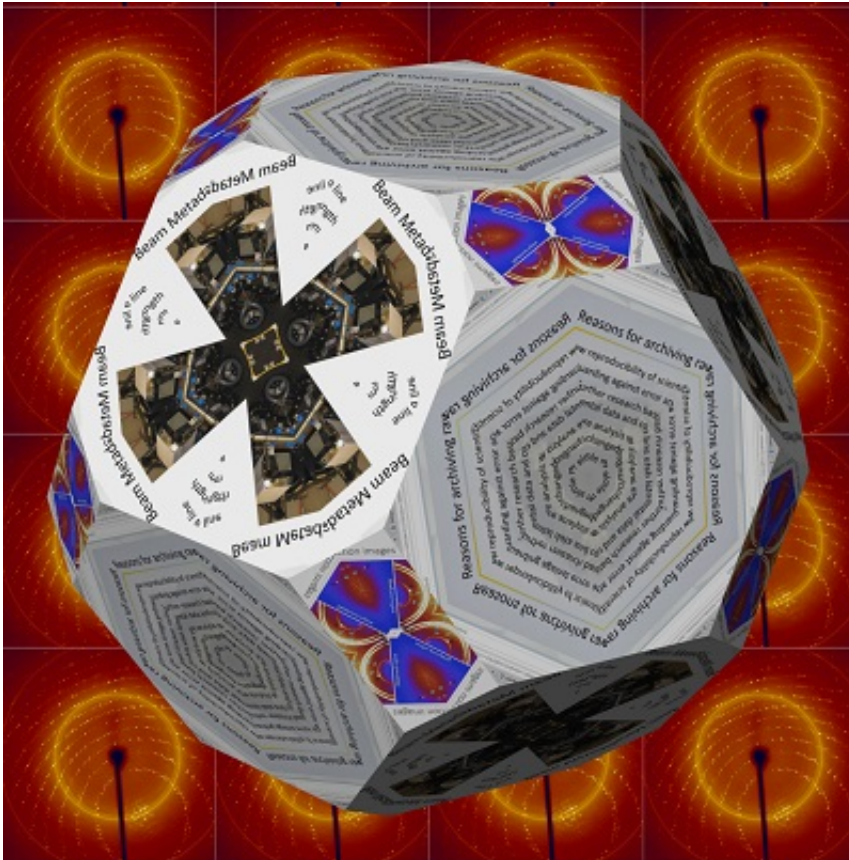


Why keep the raw data?

December 7 2016



Graphical image of re-using data. Credit: Kroon-Batenburg et al.

The increasingly popular subject of raw diffraction data deposition is examined in a Topical Review in *IUCrJ*. Building on the 2015 workshop organised by the IUCr Diffraction Data Deposition Working Group (DDDWG), the authors bring the story up to date with accounts of new subject-specific and institutional data repositories, and of growing policy

pressures on research data management such as the European Open Science initiative.

The article is, however, more than just a workshop report or a survey of evolving policy. It seeks to inform the cost-benefit arguments over diffraction data deposition with examples from real front-line research. For example, Kroon-Batenburg and Helliwell have collaborated on studies of protein binding of the chemotherapeutic agent cisplatin, and have made all their 34 raw data sets available through the University of Manchester Data Library. Some of these datasets have been reanalysed and resulted in fresh understanding of cisplatin-lysozyme models.

The prospect of extracting further information from archived primary data sets in this way (either by the insights of fresh pairs of eyes or through subsequent improvements in software analysis) has implications for structural databases, facilitating the idea of continuous improvement of studies, such as for macromolecular structure models (long championed by Terwilliger).

It is not only in the field of macromolecular structure determination that these considerations are important. One of the greatest challenges to reusing any raw data is the need for complete metadata associated with any [raw data](#) set, to allow its subsequent interpretation and full evaluation.

Various IUCr Commissions are actively publishing their summaries of the essential metadata that need to be captured alongside all experimental data sets. These initiatives and their relationship to the IUCr's standard for data characterization (CIF, the Crystallographic Information Framework) are reviewed within the article. Again, practical pointers are given to essential metadata that need to be captured alongside diffraction [data sets](#).

While there are encouraging signs that the scientific community is taking more informed interest in data management and its scientific potential, fresh challenges are being thrown up by the latest generation of instrumentation, capable of generating vast amounts of data at an incredible rate. It may not be possible to archive or even thoroughly analyse all the data that is being produced. However, this article will help to supply a deep understanding of the reasons why society should invest effort and resources into extracting the greatest value possible from the data deluge, in crystallography as in any science.

More information: Loes M. J. Kroon-Batenburg et al, Raw diffraction data preservation and reuse: overview, update on practicalities and metadata requirements, *IUCrJ* (2017). [DOI: 10.1107/S2052252516018315](https://doi.org/10.1107/S2052252516018315)

Provided by International Union of Crystallography

Citation: Why keep the raw data? (2016, December 7) retrieved 18 April 2024 from <https://phys.org/news/2016-12-raw.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--