Raw diffraction data reuse: the good, the bad and the challenging

The role of *IUCrData's* new *Raw Data Letters*

Loes Kroon-Batenburg

IUCr

IUCr 2023

Raw diffraction data reuse: the good, the bad and the challenging

Facility and raw data providers.

Raw data reusers
- **Macromolecular Crystallography**
- **Chemical Crystallography**
- **Powder Diffraction**
- Electron Crystallography?

IUCrData **raw data letters**

short descriptions of crystallographic raw data sets from X-ray, neutron or electron diffraction experiments, in the biological, chemical, materials science or physics fields

- persistent link to the location of the raw data
- describe interesting features in raw data sets
- methods and software developers for purposes such as reanalysis by newer methods
- follows FAIR principles
- Metadata capture in imgCIF

Raw Data Letters

In 2022, *IUCrData* launched a new section – Raw Data Letters, a collaborative innovation of IUCr Journals with the IUCr Committee on Data. This section publishes short descriptions of crystallographic raw data sets from X-ray, neutron or electron diffraction experiments, in the biological, chemical, materials science or physics fields, and provides a persistent link to the location of the raw data. Information for each dataset includes an imgCIF containing core metadata, a diffraction image, figures and a description of the data and their processing. See the infographic overleaf.

Editor

Loes Kroon-Batenburg

Co-editors

Miguel Aranda
Elena Boldyreva
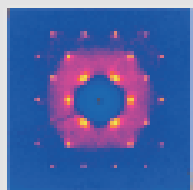Aaron Brewster
Simon Coles
John Helliwell

Read the **Editorial** at
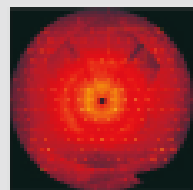iucrdata.iucr.org/x/services/rawdataletters.html

Recent Raw Data Letters

IUCrData (2023). **8**, x230114
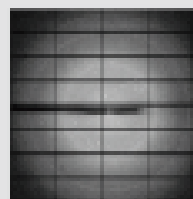**Cyclohexane plastic phase I: single-crystal diffraction images and new structural model**
S. Bernès and S. Camargo

IUCrData (2022). **7**, x221059
Accurate intensity integration in the twinned γ-form of *o*-nitroaniline
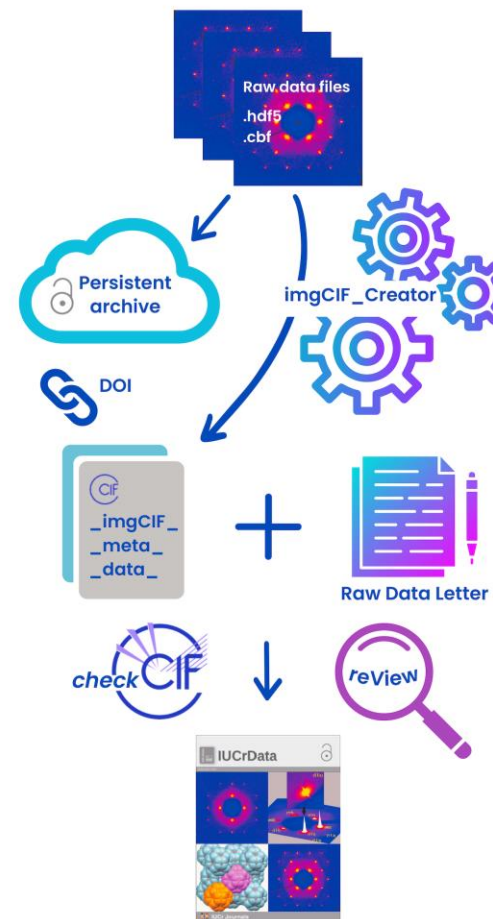M. Lutz and L. Kroon-Batenburg

IUCrData (2022). **7**, x220852
Crystal structure of the second extracellular domain of human tetraspanin CD9: twinning and diffuse scattering
V. Neviani, M. Lutz, W. Oosterheert, P. Gros and L. Kroon-Batenburg

Raw Data Letters — RAW DATA >>> ARCHIVE >>> PUBLISH

A Raw Data Letter allows researchers to highlight aspects of their raw data that could be of interest to a wide range of researchers working in the biological, chemical, materials science or physics fields, from methods and software developers to structural scientists interested in reanalysis, validation or review.

Raw data files
.hdf5
.cbf

Persistent archive

DOI

imgCIF_Creator

_imgCIF_meta_data_

+

Raw Data Letter

checkCIF

reView

IUCrData

Raw data files

Of the variety of image formats, HDF5, CBF, SMV (ADSC) are the most popular and standardized. Where possible, any other formats should be converted to CBF before archiving. CBF logically separates into two standards: one for describing the image encoding, and one for the metadata.

Archiving raw data

As part of ensuring that raw data are *Findable, Accessible, Interoperable and Re-usable* (FAIR), raw data files should be deposited in open-access persistent archives that assign a Digital Object Identifier (DOI) to a data set (for example Zenodo). For raw image data, the recommended archive formats are zip, tar+gzip, bzip2, xz, HDF5 (with standard compression filters).

Documenting metadata – imgCIF

Using tools that can read a collection of raw data files, a description of the associated metadata, including references to the locations of the raw data (binary) files, can be captured in plain text as an 'imgCIF'. An imgCIF file provides data in CIF format, which is both machine-readable and human-readable, with a comprehensive set of tags (defined in the imgCIF dictionary) for describing detector geometries and other experimental parameters, thus facilitating FAIR 're-usability'. The archived raw data, imgCIF metadata and ultimately the published Raw Data Letter are all linked via DOIs.

Publishing raw data

A Word template for writing a Raw Data Letter is available. Submission of the Raw Data Letter to *IUCrData* should be accompanied by the imgCIF describing the archived raw data. The submitted Raw Data Letter along with a *checkCIF* validation report will be subject to peer review before publication.

IUCr

IUCr 2023