OSCARS
Open Science Clusters' Action for Research & Society

Funded Project

Metadata Collection and validation for Re-use of raw Diffraction Data (MC-ReDD)

Presenter: Thomas Kluyver, European XFEL, 0000-0003-4020-6364

Implemented by

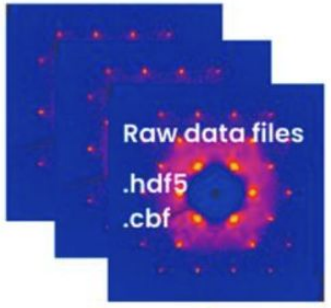
European XFEL

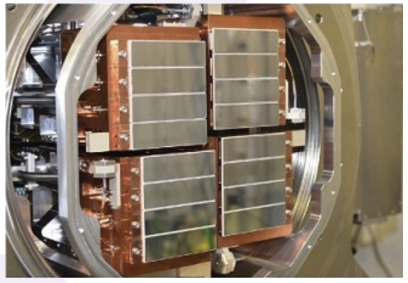
IUCr International Union of Crystallography

Funded by the European Union

CHALLENGE ADDRESSED

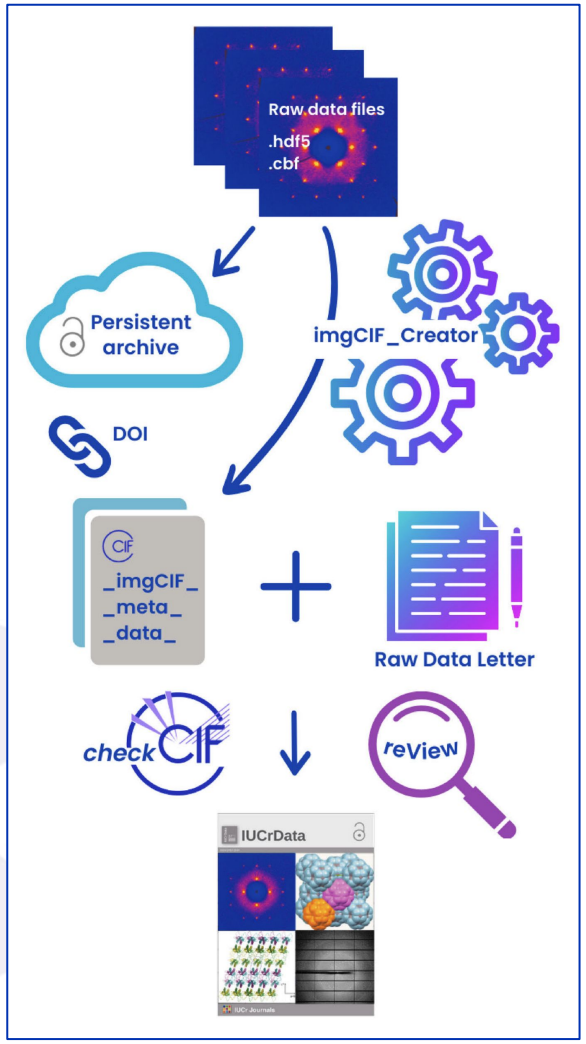
- Lack of standardisation in crystallographic diffraction image data: various file and data formats, inconsistent and incomplete metadata
- Resulting lack of interoperability, impeding the re-use of raw diffraction data
- Experiments at XFELs and synchrotrons need a partly different metadata description





SOLUTION

- Facilitate the creation of imgCIF-format files:
  - complete and standardised
  - human- and machine-readable
  - metadata with lightweight links to the raw detector image data
- Offer a user-friendly service supporting raw data publications
  - data ingestion from various input sources
  - conversion to imgCIF
  - validation

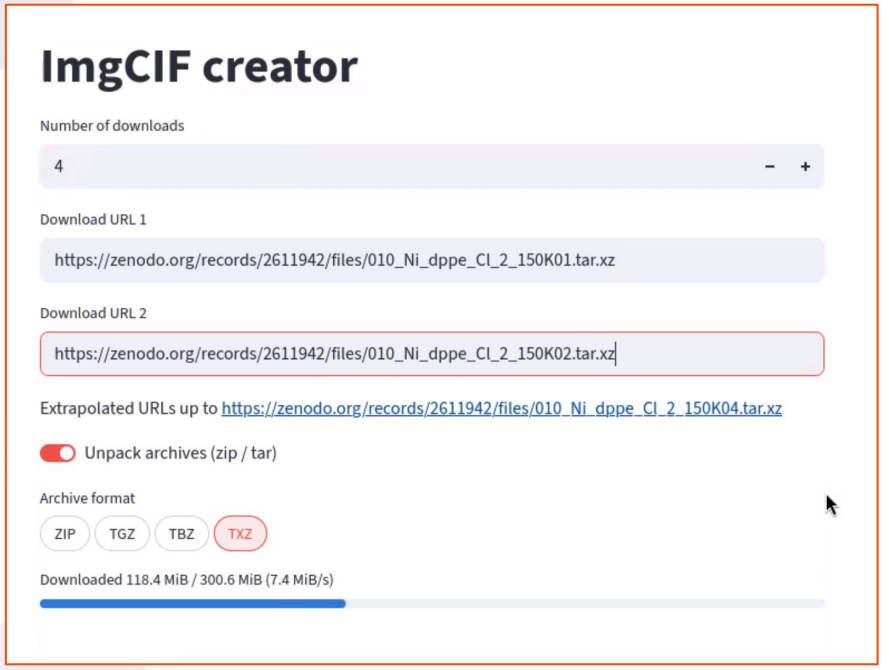


Main achievement

- imgCIF_creator, an open-source Python tool has been developed:
  - fetching input from remote repositories or local storage
  - filling incomplete metadata
  - converting to imgCIF output

Release to the public

- standalone, command-line-driven program
- web application with Streamlit frontend



Sustainability

- The web-based version is close to completion; to be first deployed as IUCr service.
- Mid-term, it shall be linked from or hosted by the PaNOSC-EOSC node.
- PaN facilities can host the service and will add imgCIF files to open data repositories, in the long run.

Result exploitation

- A streamlined process supporting sound data publications will raise their number, and the opportunities for re-use.
- Reproducible crystallographic results will increase the trust in this field of science and foster further insight through re-use.
- Structural information from crystallographic methods benefits society through application in life sciences, material science and many more domains.

TEAM

OSCARS

- Loes Kroon-Batenburg, IUCr raw data letters editor: Project PI
  0000-0002-5321-1392

- James Hester, ANSTO: Technical advisor
  0000-0002-2004-8672

- Thomas Kluyver, European XFEL: Software developer
  0000-0003-4020-6364

- Fabio Dall'Antonia, European XFEL: Coordinator/spokesperson
  0000-0003-0799-2244