Fail2Fair: Recovering Discarded Macromolecular Crystallographic Data

OSCARS
Open Science Clusters' Action for Research & Society

The Fail2Fair project seeks to recover value from discarded crystallographic datasets by developing a pipeline to annotate, classify, and integrate them into SciCat metadata catalogue, an Open Research Data infrastructure established at PSI. By enriching metadata with detailed descriptions of crystallographic failure modes, the project will make such datasets FAIR, while supporting the advancement in the development of new tools, including those based on AI.

PaNOSC
Photon and Neutron Science

Challenge

Current repositories such as the Protein Data Bank (PDB) only include successful and processed experiments, excluding vast amounts of potentially informative discarded or unsolved crystallographic data. Moreover, SciCat lacks the capacity to document why certain datasets failed or were discarded.

Solution

Extension of the SciCat metadata framework with a new schema for recording crystallographic failure modes. Using an AI-powered pipeline, Fail2Fair will allow automatically annotating and classifying crystallographic datasets that have been discarded for various reasons. These include datasets with pathologies, low-quality datasets and other crystallographic pitfalls.

Scientific Impact

By making previously discarded crystallographic data FAIR, Fail2Fair will improve data sharing and reuse. The reanalysis of such data deepens our understanding of common crystallographic problems, helps us identify systematic experimental errors, and supports troubleshooting in future experiments, ultimately reducing experimental efforts.

Partners

Institute of Molecular Biology of Barcelona - Spanish National Research Council - IBMB-CSIC, University of Konstanz, Paul Scherrer Institute

https://oscars-project.eu/projects/fail2fair-recovering-discarded-macromolecular-crystallographic-data