

IRRCM ([https:// proteindiffraction.org](https://proteindiffraction.org)): Impact on quality of structures in PDB**Marek Grabowski, Marcin Cymborowski, David Cooper, Wladek Minor***University of Virginia, Charlottesville, United States of America;**marek@iwonka.med.virginia.edu*

Preservation and public accessibility of primary experimental data are cornerstones necessary for the reproducibility of empirical sciences. Many crystallography journals recommend that authors of manuscripts presenting a crystal structure deposit their primary experimental data (X-ray diffraction images) to one of the dedicated resources created in recent years. We present the Integrated Resource for Reproducibility in Molecular Crystallography (IRRCM). In its first five years, several hundred crystallographers have deposited over 9000 datasets representing more than 5,700 diffraction experiments performed at over 60 different synchrotron beamlines or home sources all over the world. We describe several examples of the crucial role that diffraction data can play in improving previously determined protein structures. In addition to improving the resource and annotating and curating submitted data, we have been building a pipeline to extract or generate the metadata necessary for seamless, automated processing. Preliminary analysis shows that about 95% of the data received by our resource can be automatically reprocessed. A high rate of reprocessing success shows the feasibility of automated metadata extraction and automated processing as a validation step that ensures the correctness of raw diffraction images. The IRRCM is guided by the Findable, Accessible, Interoperable, and Reusable data management principles. Data from IRRCM have already enabled several novel research projects.

Keywords: protein crystallography, X-ray diffraction, reproducibility, big data