

Impact On Validation and Quality of Structures Deposited To PDB

Wladek Minor¹, Marcin Cymborowski¹, David R Cooper¹

¹*University of Virginia*

wladek@iwonka.med.virginia.edu

Preservation and public accessibility of primary experimental data are cornerstones necessary for the reproducibility of empirical sciences. We present the results obtained with Integrated Resource for Reproducibility in Molecular Crystallography (IRRCM) resource. In its first five years, several hundred crystallographers have deposited over 9600 datasets representing more than 6,100 diffraction experiments performed at over 60 different synchrotron beamlines or home sources worldwide. 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. 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 Findable, Accessible, Interoperable, and Reusable data management principles guide the IRRCM. Data from IRRCM have already enabled several novel research projects.