INTEGRATING DATA WITH PUBLICATIONS: GREATER INTERACTIVITY AND CHALLENGES FOR LONG-TERM PRESERVATION OF THE SCIENTIFIC RECORD

Brian McMahon

International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, UK,
e-mail: bm@iucr.org

The journals of the International Union of Crystallography (IUCr) publish many reports of crystal or molecular structures which discuss associated data sets. These data sets (atomic positional coordinates and anisotropic displacements, molecular geometry, protein secondary structure) are stored by the journal or by public data banks. Recently, IUCr journals have begun to incorporate interactive three-dimensional visualizations of such data as enhanced figures in the published article. Such figures bind publication and data more closely together than ever. There is also, within the crystallographic community, increasing pressure to retain and publish primary experimental data as well as these derivative data sets. The primary data could then be linked to the publication for purposes of validation and reproducibility. Such a development would greatly extend the nature of the research report, integrating the discussion of new scientific results with access to all of the associated data. Proper management of such publications will require: formal descriptions of compound documents, with components possibly distributed at multiple sites; the ability to cite and extract specific slices of large data collections; persistent identifiers for data sets and the associated publications; suitable linking protocols; and appropriate allocation of responsibility and resources for archiving the separate components in a consistent and compatible framework.