

**OCM04 COMMISSION ON BIOLOGICAL MACROMOLECULES***Coordinator:* M. Guss**OCM04.26.1***Acta Cryst.* (2005). **A61**, C129**What is the Issue with Autodeposit?**Zbyszek Otwinowski, *UTSouthwestern Medical Center at Dallas*. E-mail: [zbyszek@work.swmed.edu](mailto:zbyszek@work.swmed.edu)

In macromolecular crystallographic structure solution the flow of information between stages is the key component of success. The final model is deposited in the PDB together with the information collected during the process. The dispersion of information between different formats of different programs makes the data organization and the automation of deposition a hard to achieve task. The main issue is still deciding which information is the most valuable and should be deposited. The answer depends on the view of the PDB role. The Protein Data Bank is not only the structural models repository, but it also contains a limited description of facts related to structure determination. The question is: should we enhance this information and can we achieve it using automatic tools?

The important facts missing from the PDB are: the increase in temperature factor, comments about twinning, diffuse scattering, use of multiple crystals for a single data set. Log files from numerical calculations are mostly of interest to the process of structure determination; however, it is not clear how this information would enhance the PDB deposits.

Overall, we should concentrate on how to improve the process of structure determination, which is challenging due to diversity of problems and approaches. The issue of autodeposit is subordinated to an overall improvement in crystallographic computational and database methods.

**Keywords:** data harvesting, data deposition, missing data**OCM04.26.2***Acta Cryst.* (2005). **A61**, C129**Harvesting should be a Complete and Repeatable Record of Experiment**Thomas Terwilliger, *Bioscience Division, Los Alamos National Laboratory, Mail Stop M888, Los Alamos, NM 87545 USA*. E-mail: [terwilliger@lanl.gov](mailto:terwilliger@lanl.gov)

The procedural and crystallographic data that are harvested and stored as a record of a crystallographic experiment should be sufficient to (1) repeat the analysis and (2) to estimate uncertainties, preferably in a completely automated fashion. It is highly likely that automated methods will be constantly developed over the next years and that there will be repeatedly be utility to a re-analysis of all structures currently in the PDB. Such analyses will be possible only if the primary data and any associated knowledge about the experiment necessary to fully and automatically repeat the structure determination and analysis process are available. Further, even in the absence of full redeterminations of structure, re-estimates of errors and uncertainties (as in structure validation) will be done repeatedly on the entire PDB. The final structure and structure factors are not sufficient to repeat either the structure determination or the the error analysis (for example a free-R analysis is impossible if a test set is not specified; model bias can never be fully removed if an atom is placed and later removed). The data harvested from an experiment should therefore be a complete record of the process used to determine the structure, in a form that can be automatically read, repeated, and modified.

**Keywords:** PDB, LIMS, harvesting**OCM04.26.3***Acta Cryst.* (2005). **A61**, C129**Harvesting - from Data Collection to Deposition**Harold R. Powell<sup>a</sup>, Andrew G.W. Leslie<sup>a</sup>, T. Geoffrey Battye<sup>a</sup>, Kim Henrick<sup>b</sup>, <sup>a</sup>*MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK*. <sup>b</sup>*EMBL Outstation Hinxton, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*. E-mail: [harry@mrc-lmb.cam.ac.uk](mailto:harry@mrc-lmb.cam.ac.uk)

The best approach to data harvesting is undoubtedly to have all relevant information regarding the experiment recorded automatically as it is determined. Provided that this is done reliably (i.e. no data are missing and all items present are also correct), this can be used in an automated pipeline requiring minimal user intervention. By far the majority of errors in the harvesting process arise from manually entered user input.

While a standard format has been developed by the IUCr for area detector image data (imgCIF and its binary equivalent, the Crystallographic Binary File or CBF), it has not been adopted widely. In the absence of a commonly used format for the image data, it is important that files written using existing commercial formats contain the essential experimental information along with the intensity data. Further, the data processing programs should record these data in a format which is readily amenable to straightforward deposition in the appropriate database.

**Keywords:** data collection, data processing, data harvesting**OCM04.26.4***Acta Cryst.* (2005). **A61**, C129**Streamlining Data Bank Deposition and Manuscript Submission**Howard Einspahr, *Editor, Acta Crystallographica Section F, Lawrenceville, New Jersey, USA*. E-mail: [hmeinspahr@yahoo.com](mailto:hmeinspahr@yahoo.com)

One of the reasons an all-electronic *Acta Crystallographica Section F* was created was to meet the increasing need for rapid publication of structural biology reports. Few have failed to notice the explosive acceleration in depositions to the Protein Data Bank (PDB) in recent years, but unrealized by many is the growing potential for depositions that have not been accompanied by a publication documenting the experiment and its results. Failure to publish means results are inaccessible to abstracting services and thus to the wider biological audience. A strategy to speed publication has emerged from collaborations between the PDB and IUCr Journals: eliminate repetition and streamline submissions by facilitating transfer of data deposited in the PDB to the journal at submission. Transfer of information for publication is through data files with content described in the PDB Exchange Data Dictionary. Deposition tools are available at each wwPDB site to assist in the preparation of these data files. Additional tools such as PDB EXTRACT are also available to help automate the extraction of data items from popular structure determination software applications. It is hoped that a system combining publication and PDB deposition will reduce the overall effort required to perform both activities. Details and progress of this effort will be described.

**Keywords:** *Acta Cryst. F*, PDB, streamlining**OCM05 COMMISSION ON INORGANIC AND MINERAL STRUCTURES***Coordinator:* G. Ferraris**OCM05.27.1***Acta Cryst.* (2005). **A61**, C129-C130**Modular Aspects of Inorganic and Mineral Structures**Emil Makovicky, *Geological Institute, University of Copenhagen*. E-mail: [emilm@geol.ku.dk](mailto:emilm@geol.ku.dk)

With the increasing complexity of known inorganic and mineral structures, the importance of *moduli*, composed of a number of coordination polyhedra, became obvious for the understanding of these structures. The moduli are fragments of simpler structure types, so called *archetypes*, which have been recombined into a recombination structure according to a set of new structural principles, unit cell (chemical) twinning, crystallographic shear, non-commensurate interfaces, and coherent intergrowths. Structures built on the same principles, but with an incrementally increasing fragment size, form *accretional homologous series/polysomatic series*. The homologous approach allows distinction into *extensive* and *combinatorial* categories. *Variable-fit structures* have two (sub)lattices present in one structure. Combination of the accretional and variable-fit principle leads to new complex structures, including 'sliding series' and 'box-work structures'. *Configurational homology*