

DRAFT FOR DISCUSSION

The principal work undertaken by COMCIFS this year has been the development of the 'methods' Dictionary Definition Language (DDLm), a language for writing CIF dictionaries that is more rigorous and flexible than those currently in use (DDL1 and DDL2). It builds on the experience we have had with CIF over the past seventeen years. It will be easier to maintain, but more particularly it will contain machine executable mathematical expressions that will allow CIF dictionaries to instruct the software how to calculate derivative items, such as the crystal density, from other items in the CIF. DDLm will complement rather than replace the existing DDLs. Software that works with dictionaries written in DDLm will be able to read the full existing archive of CIFs written using the DDL1 and DDL2 standards, but will bring considerable added value to the archive. In the immediate future we are likely to see little change in the CIFs themselves, but we can expect to see powerful software that exploits the new standard.

COMCIFS agreed to develop DDLm at the Florence Congress and this year has seen the proof-of-principle prototype DDLm developed by Syd Hall and Nick Spadaccini expanded to a version suitable for adoption by COMCIFS. This work has been completed by a group consisting of Syd Hall, Nick Spadaccini, John Westbrook and others, and the draft DDLm has been submitted to COMCIFS for approval. COMCIFS has commissioned James Hester to provide a detailed review and testing to ensure that the draft is complete and consistent. When Hester's report is received DDLm will be presented to COMCIFS for adoption. At the same time, David Brown has been adapting the coreCIF dictionary to the DDLm standard and the IUCr has funded Herbert Bernstein's group at Dowling College to update the CIF software used in the journal publication so that it can validate CIFs against the new CIFm dictionaries. Because of the tighter standards and greater flexibility of DDLm compared to DDL1 and DDL2 in current use, the updated software will allow more subtle discrepancies and inconsistencies to be caught than has been the case in the past. The report from Hester, the draft of the CIFm core dictionary and the software from Bernstein is expected to be ready for review by COMCIFS at the IUCr Congress in Osaka.

During 2007 work continued to facilitate the use of mmCIFs in biological articles in *Acta Crystallographica*. A set of recommended data items has been drawn up by Howard Einspahr for describing information on: the sample and its treatment (including crystallization); data collection and structure solution; and structure refinement details. These can be harvested automatically from an mmCIF and will generate a table for publication in the article. A tool has also been developed for creating enhanced figures in all IUCr journals; these figures are three-dimensional visualizations of molecular structure with associated animations, specified views and schematic representations, and use the CIF (or mmCIF) directly for atomic coordinates and crystallographic symmetry information. Work on pubCIF as an authoring and editorial tool has continued.

The development of the traditional CIF dictionaries has also continued. A major revision of the core dictionary was completed and should receive final COMCIFS approval early in 2008. A further addition covering the description of the restraints and constraints used in refinement is

being developed. In addition work has resumed in the development of dictionaries for reflectivity and small angle scattering.

2007 was a year of rapid progress on imgCIF. Thanks in part to workshop funding from DOE, NSF and NIH, issues that had been holding up adoption of imgCIF at beamlines were addressed, new code was written, and we are pleased to report that imgCIF was adopted as the output format for the new Dectris Pilatus 6M detector at the Swiss Light Source in Villigen Switzerland and as of early January 2008, ADSC had prepared software to produce imgCIF from all its detectors for use at the Diamond Light Source in Chilton, England. Work and workshops will continue in 2008 and the results of this effort will be discussed at the IUCr Congress in Osaka.

A microsposium at the Osaka Congress has been arranged jointly with the Commission on Crystallographic Computing to introduce the DDLm dictionaries and software as well as discuss the interexchange of CIFs with other standards.

The Protein Data Bank continues to extend the content of the wwPDB Exchange Data Dictionary (PDBX). PDBx a superset of mmCIF, which in addition to macromolecular X-ray methods, includes structure and experimental representations of NMR, 3D electron microscopy, homology modeling, and experimental details of protein production (<http://mmcif.pdb.org>) A translated version of this dictionary is maintained as an XML schema (PDBML) (<http://pdbml.pdb.org>). In the past year significant extensions have been added to represent large molecular assemblies and more detailed chemical description of both polymer and non-polymer molecular components in macromolecular structures. The wwPDB has released a remediated version of the PDB archive which takes advantage of these latter extensions. This work is described in "Remediation of the protein data bank" by K. Henrick; et al. Nucleic Acids Research 2008 36(Database issue):D426-D433; doi:10.1093/nar/gkm937

Now that CIF is firmly established in the crystallographic community and the nature of the work of COMCIFS is moving from dictionary production to coordinating the production of software, it is time to introduce younger members with different expertise into COMCIFS. Half of our membership, including myself, will be retiring at the Osaka Congress and I would like to take this opportunity of thanking Helen Berman, Syd Hall and Gotzon Madariaga for the great work they have done in guiding COMCIFS through its first critical years.

David Brown  
Chair of COMCIFS