

Minutes

COMCIFS Meeting

Montréal, Palais des Congrès, Room 516e, 8 August 2014, 16:45-17:30

1. Report on CIF-NeXuS interaction (H. J. Bernstein)

Herbert Bernstein reported on continuing work in progress to allow interoperability between imgCIF and HDF-5 files supporting the NeXuS data standard. This is necessary because HDF-5 files will be written by next-generation image detectors. Cooperation between COMCIFS and NIAC (the NeXuS International Advisory Committee) meant that *CBFlib* was now interoperable between CBF/imgCIF and HDF-5/NeXuS formats for existing single large frame detectors. Work was in hand to anticipate the requirements of the next generation Dectris detectors, but sample files were needed for free-electron laser and Dectris Eiger detectors. It was still expected that adoption of DDLm/dREL would be needed to provide a full suite of methods to map between imgCIF and HDF-5 representations.

2. Discussion and possible approval of new data citation data items (B. McMahon)

A small set of candidate data items was presented by Brian McMahon for adding to the core dictionary some information that would be helpful in document and dataset citation. The data names, agreed between IUCr and CCDC, were

```
_audit_block_doi  
_citation_doi  
_citation_id  
_citation_publisher  
_database_dataset_doi  
_publ_author_id_orcid  
_publ_contact_author_id_orcid
```

After brief discussion, these were approved by the meeting.

3. Dictionaries under construction: magCIF (J. Hester), propCIF

(a) The magnetic CIF dictionary magCIF was now under construction. The project was being coordinated by Branton Campbell (Chair of the IUCr Commission on Magnetic Structures), and already included a coalition of major software packages and databases including the Bilbao group, *JANA* and *FullProf*. The main list of concepts had been identified and agreed upon. Definitions and data names were now being developed, and would constitute additions to the coreCIF, msCIF and symCIF dictionaries. There was some discussion about the desirability of writing the dictionary in DDLm formalism to facilitate its use across both DDL1 and DDL2 applications (see also agenda item 7). [Subsequent discussion at the Open Meeting of the Magnetic Structures Commission confirmed that the magCIF dictionary would be prepared in DDLm form. Elsewhere at the Congress, Bob Hanson demonstrated that Jmol was already able to visualize magnetic structures using an early draft set of magCIF data names.]

(b) The materials properties dictionary, originally developed for the Materials Properties Open Database by Daniel Chateigner, was now being maintained in Mexico by Luis Fuentes and co-workers, in association with the Crystallography Open Database personnel. It was agreed that, for the time being, this was best developed within its own community, but advice and support would be available from COMCIFS. It was agreed that working drafts would be assessed for syntactic correctness and possible name clashes with CIF dictionaries, and their availability could be advertised through the IUCr CIF dictionary register. For the present, they would still be hosted on MPOD servers, but COMCIFS could provide mirroring from the IUCr servers also if required. Saulius Gražulis encouraged the development of some common revision control system that could be applied across all public CIF dictionaries.

4. Plan for next triennium

James Hester reported that COMCIFS should work in the next triennium to:

(a) Manage the formal introduction of CIF2/DDLM with associated documentation and relevant software support such as the CIFAPI (item 6a) and small standalone utilities that could assist in transitioning small-scale workflows. After discussion under item 7, it was accepted that efforts would be made to harmonise the published DDLm specification with existing large-scale publishing and database workflows. It was also suggested that COMCIFS should emphasise the commitment of publishing and database activities to long-term secure and stable information and data curation. One of the founding principles of CIF was as an archiving mechanism, and ongoing development to add future functionality to CIF would have no significant impact on the continued readability of all valid existing CIFs.

(b) A need had been identified for a modest fund that could help engage the software development community in upgrading CIF readers/writers to adapt to the new specifications. It might also be helpful for the journals publishing office to contract out or collaborate with external groups in developing specific projects on an accelerated timescale. A request had gone forward to the IUCr Executive Committee to approve such a plan.

(c) The IUCr Diffraction Data Deposition Working Group was an active body with great interest in developing metadata for describing a wider range of experimental raw data from different crystallographic and structural science techniques. COMCIFS was already well represented in the DDDWG membership, and affirmed its willingness to provide experience in data definition to IUCr Commissions and other groups developing their own experimental data ontologies.

(d) The possibility was raised of a further workshop along the lines of the successful event at ECM28 in Warwick in August 2013.

(e) It was anticipated that the preparation of a new edition of *International Tables* Volume G would begin during the triennium, to assist in the documentation of CIF2 and DDLm. It was hoped that the canonical dictionaries reported in the new edition would be presented in DDLm (as per item 7).

5. Report on meeting with IUCr Executive Committee

Following an initial request, James Hester had asked the Executive Committee (EC) to be prepared to make available modest funding in the next triennium if needed as seed corn for specific projects in the Chester office and amongst the wider software community. Following submission of a more detailed plan of work requested by the EC, the proposal was under consideration.

6a. Project update: CIFAPI

John Bollinger reported that a C-based CIFAPI offering fully-compliant CIF2 functionality but able to support all CIF1 functions was in an advanced state of preparation. It should be available as a beta release around late September 2014, and would most likely be released under LGPL licence. The core API and the initial release were in C; it was anticipated that Python, Java, C++ and Fortran bindings would become available in future. The meeting thanked John for this important work to date.

6b. Project update: wwPDB

John Westbrook reported the continuing growth of the wwPDB and the leadership transition at Rutgers, where Stephen Burley was welcomed as the new Director. He described the successful transition to mmCIF as a working format amongst macromolecular crystallographic software developers. The developer community had expressed a preference for moving directly to mmCIF instead of adopting a further hybrid format needed to handle the increasingly complex ribosomal-scale structures being submitted to the PDB. Ontology development continued vigorously with small-angle scattering definitions joining the existing extension dictionaries for NMR, protein production etc. These were all facilitated by the PDBX/mmCIF definition framework, and represented a community exercise in defining its data attributes that was unmatched. mmCIF submissions were now coming in routinely from packages such as *REFMAC* and *Phenix*. The meeting expressed its admiration and appreciation for the wwPDB success in its CIF-based data management strategies and practice.

7. Discussion, hopefully leading to decision: "that all new dictionaries accepted by COMCIFS use DDLm, and that COMCIFS updates existing dictionaries to DDLm as soon as possible."

Concern was expressed that the current draft DDLm versions of the existing suite of CIF dictionaries omitted significant semantic content (*e.g.* from the mmCIF dictionary). These drafts represent work in progress, and a working group (initially comprising John Westbrook, James Hester and Herbert Bernstein) would identify any additional attributes required for DDLm to ensure that no semantic loss would ensue from such translations. Attention would also be paid to the harmonisation of attribute names.

It was agreed that any new dictionary submitted for COMCIFS approval should have a canonical DDLm version. This would not preclude development of a dictionary using DDL1, but the reference version upon acceptance would be DDLm.

8. Any other business

Jim Kaduk inquired if any data items needed to be defined for hypothetical structures. Saulius Gražulis reported that some work had already been started on a dictionary in association with the Predicted Crystallography Open Database. It was intended that this would be referred to COMCIFS for review when suitably mature. John Westbrook mentioned a group set up to archive protein homology models that would work with the PDB to deposit structures of all types, including theoretical.

Brian McMahon
Coordinating Secretary
26 August 2014

