



**Report from
Dowling College, Oakdale, NY, USA
to the International Union of Crystallography, Chester, England
concerning provision of new and upgraded CIF software to facilitate publication in
IUCr journals**

1 August 2005

The Executive Secretary of the IUCr
Mr M. H. Dacombe
International Union of Crystallography
2 Abbey Square
CHESTER CH1 2HU
England

This is the 6th bimonthly progress report on the IUCr funded project at Dowling College to support the evolving needs of the community for new and upgraded CIF software to facilitate publication in IUCr journals.

I. Project Summary

Dowling College is providing to the IUCr the services of Professor Herbert J. Bernstein as project director (PI/PD) and certain of his students to modify existing software and to create new software as detailed in section I below in order to support the evolving needs of the community for new and upgraded CIF software to facilitate publication in IUCr journals. This is a major set of inter-related projects, expected to take more than two years to complete. However, a phased release to the Chester office of the IUCr of partial preliminary versions of all of these packages is expected before the end of the first year, and feedback from the Chester office of the IUCr will be used to guide completion of the packages.

As versions of these packages mature they will be released to the community as open source software without charge to encourage wide use. The software will be released using the GNU GPL or a similar license. "CIF Applications" articles will be submitted to help make the community aware of these new and upgraded tools, and the IUCr will be given first refusal in publication of such articles produced from the work of this project.

II. Description of goods and services to be delivered to the IUCr

These include:

- (1) CIFTTEST – Creation of a new and extended test file suite and test protocol for validation of CIF parsers.
- (2) vcif2 – an extended version of vcif (a program used to validate the syntax of CIF files). vcif2 will verify compliance with CIF 1.0, CIF 1.1 or mmCIF file formats. vcif2 will also accept an arbitrary list of layered DDL1 and DDL2 dictionaries against which units, enumerations, parent–child relationships and category integrity will be checked.
- (3) CIFFOLD – a new utility to fold and unfold long-line CIFs to allow existing CIF 1.0 applications to work with CIF 1.1 files.
- (4) CIFtbx3 – a new release of CIFtbx to provide support for CIF 1.1 (as well as CIF 1.0 and mmCIF) for Fortran applications and to provide extended integrity checking comparable to that in vcif2.

III. Timetable

The agreement started 1 August 2004. The Agreement will terminate when the work under Item I is complete and this will be no later than 31 July 2006.

IV. 1 August 2005 Status

Overview: The project started on time on 1 August 2004, and has continued to date. This is a report for the work done from 29 May 2005 through 1 August 2005. The project is fully staffed and operational and in this period the major effort has been on preparations to present the status of the project at the IUCr Congress in Florence in August, with a primary focus on CIFFOLD. The progress-status web site has been updated.

Staffing: The PI/PI is Professor Herbert J. Bernstein. The students working on the project are Kostadin Mitev and Georgi Todorov. Isaac Awuah Asiamah, Ricky Chachra, Clarice Chigbo and Stavros Louris have assisted in testing. For this reporting period, K. Mitev has been away, and Clarice Chigbo completed her work in the lab. Most of the work reported for this period has been done by H. J. Bernstein and G. Todorov, with testing by I. Awuah Asiamah, and some remote contributions by K. Mitev.

Funding and Administration: Cash flows and burn rates are appropriate to the needs of the project. Some of the funds will be used for part of the costs of travel to the Congress in Florence.

Project Activities: In this project period the focus has been on getting presentations and web sites ready for the IUCr Congress in Florence with testing of packages and resolution of problems discovered. Releases of CIFFOLD and CIFtbx3 were accepted by B. McMahon for inclusion in the ITVG CDROM. The internal project web site has been promoted to become an external web site at:

<http://arcib.dowling.edu/cifiucr>

and the foils for the talk have been posted at

http://arcib.dowling.edu/cifiucr/IUCr2005_Presentation_27Jul05/

The two released packages have been placed at

<http://www.bernstein-plus-sons.com/software/ciftbx>

and

<http://www.bernstein-plus-sons.com/software/ciffold>

CIFtbx: As reported in the prior reporting period, fully operational folding and unfolding with acceptable performance on long lines was integrated with CIFtbx and released as CIFtbx3. During this reporting period, CIFtbx3 was made the default release of CIFtbx (see the link above), and, with S. R. Hall's approval, released under the GPL. The package was given extensive testing as the base for a new version of the program cif2pdb being used in another project, and worked correctly with excellent performance.

CIFFOLD: During this reporting period, the program CIFFOLD was tested and documented and progressed from release 0.4.3 to release 0.5.2 as problems were discovered and corrected and the quality of the output was improved (e.g. to fold between blank separated words in the style of CIFtbx3, rather than precisely on column 80 as in earlier releases of CIFFOLD). The code has been working well, and a link has been placed on the public web page for CIFFOLD to allow retrieval of folded versions of the long-line mmCIF datasets currently being released by the RCSB PDB. As of this writing the 0.5.1 release is the default release. The 0.5.2 release, currently available for testing at

http://www.bernstein-plus-sons.com/software/CIFFOLD_0.5.2

changes the handling of folded long single or double quoted strings to include a terminal backslash in the resulting text field, so that an extra newline will not be added on reconstruction, and also deals with additional cases of embedded ";" sequences in text fields that might end up in column 1. If this release does well in testing, it should be the default release during the Congress.

A logo was created for CIFFOLD:



Work by K. Mitev, G. Todorov and H. J. Bernstein, with testing by I. Awuah Asiamah.

CIFTEST: Continuing from the prior reporting period the intensive work on CIFFOLD has resulted in interesting new test cases for long lines and for cases that might break parsers. I. Awuah Asiamah is cleaning up and organizing these test cases. Some of the test cases have become part of the “make tests” section of CIFFOLD already. Work by I. Awuah Asiamah, K. Mitev and H. J. Bernstein.

vcif: Since I. Awuah Asiamah has become the primary tester for the packages in this project, he is taking over responsibility for vcif. If the visa issues can be resolved in a timely manner, I. Awuah Asiamah will attend the IUCr Congress in Florence (as will K Mitev, G. Todorov and H. J. Bernstein) which will be helpful in discussions relative both to CIFTEST and vcif.

Summary: The project continues on track.

Respectfully Submitted,

Herbert J. Bernstein
Professor of Computer Science

Cc: Brian McMahon
Dawn Pierpoint