CIF 1 and CIF 2

MSWK.CF.01 WHAT IS A CIF, AND HOW IS IT USED? By B. McMahon, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

The Crystallographic Information File, or CIF, is the data exchange format defined and adopted by the IUCr for the transfer of crystallographic information between laboratories, databases, journals and applications programs. It has proven successful in all these areas: for example, Acta Crystallographica Section C is typeset entirely from electronic submissions in CIF format, and the data in the resultant papers are transferred by ftp direct to the Cambridge Crystallographic Data Centre. The highly structured file format allows a large amount of automated checking and validation of the included data, and the Union's goal of achieving a free flow of uncorrupted data from refinement program through to publication and archive has been largely achieved.

As the CIF project has developed, it has become clear that there are more profound benefits to be reaped. Precise definitions have been drawn up of the scientific terms that should be universally recognised within the discipline. The dictionary that holds these definitions is an electronic file. This file can also store machine-readable information about the ranges of permitted values for the quantities described, and about their interdependence. This gives rise to the possibility of writing sophisticated software able to manipulate novel concepts in the science as they are developed.

A technical committee of the IUCr (known as COMCIFS) maintains and develops dictionaries of terms that may be used in universal CIF applications.

As equipment manufacturers continue to add CIF capability to their software, as increasingly powerful applications are developed, as the IUCr makes available its archive of structures in CIF format, so will this novel data format continue to evolve into the universal language of crystallography.

MSWK.CF.02 HOW IS A CIF PREPARED AND EDITED? George Ferguson, Department of Chemistry and Biochemistry, University of Guelph, Guelph, Ontario, Canada N1G 2W1 (george@xray.chembio.uoguelph.ca)

This presentation will review the types of CIF typically produced by several well known crystallographic program systems and detail what MUST be added to each of them with any text editor to produce a "working" CIF which will pass the tests in the IUCr program CHECKCIF. The resultant CIF will then be acceptable to the IUCr program PRINTCIF which returns to the submitting author a postcript file of the CIF. Once the CIF is formally submitted to the Managing Editor of Acta Crystallographica, a similar printout is prepared in Chester and sent to Coeditors and referees. Specimen "raw" CIFs will be available both in print and electronic form, together with a list of the data items which should be added in each case to yield a "working" CIF. Specimen CHECKCIF and PRINTCIF outputs for correct CIFs and those with errors will be shown and discussed. Additional CIF features (which are not necessary for the formal acceptance of a CIF by Acta Crystallographica) which make the CIF a useful archival document will also be presented.

MSWK.CF.03 WHAT IS A CIF DICTIONARY? Sydney R Hall, Crystallography Centre, University of Western Australia, Nedlands 6907, Australia. (syd@crystal.uwa.edu.au)

In a CIF each data item to be identified by a tag (ie. a unique data name), and specified as a text string (ie. a data value) following the tag. To access a CIF data item one must know about the data name and value in advance - a common requirement for accessing data in any format! The important difference with CIF data is that this information resides in an electronic dictionary which is human and machine parsable. This talk will introduce the CIF dictionary concepts.

Data will be described as "data objects". A data item is not just "number" or "text string" but an entity with properties or "attributes" which collectively define its uniqueness, and its relationship to other items. For example, a common data item in crystallography is the "calculated structure factor Fc". How does one go about defining Fc? It turns out to be a non-trivial task! The most obvious approach might be to use the Fourier transform expression and link its definition to the atom types and atomic parameters. This introduces quite complex relationships with other data items, and these in turn are dependent on data such as the diffraction measurements. In fact, strictly speaking, Fc is related to almost every crystallographic measurement!

A common reaction to such a detailed definition is that "its over the top"! But is it? If one wanted to compare calculated and measured structure factors but only have access to the structural data then must know this information. If a data dictionary contained the "complete" description of Fc is understood by an accessing tool, then missing Fc values could be generated automatically. Fc is a "dependent" data item directly related to other crystallographic information. Other data items are not as dependent. For example, the crystal cell dimensions are often referred to as "primitive" data because they cannot be easily derived from other data (other than diffraction angles and indices, that is!). These aspects will be explained in the talk.

The second part of this talk will describe the format and syntax of the electronic dictionary. This is referred to as the "dictionary definition language: (DDL). The DDL1 specifications have been published[1], a newer version DDL2[2] is already in use for the mmCIF dictionary. A DDL dictionary file contains definition data which conforms to the STAR File syntax[3]. Each definition is composed of a sequence of DDL data items referred to as DDL attributes. The attributes are the vocabulary of the dictionary language, and, individually and collectively, they provide the semantic tools of the dictionary. The description and rationale of the DDL will be described.

1 Hall, S.R. & Cook, A.P.F. STAR Dictionary Definition Language: Initial Specifications. J. Chem. Inf. Comput. Sci. 1995, 35, 819-825.
2 Westbrook, J.D. STAR Dictionary Definition Language: Extended Syntax. To be presented in CIF-II.
3 Hall, S.R.; Spadaccini, N. The STAR File: Detailed Specifications. J. Chem. Inf. Comput. Sci. 1994, 34, 505-508.

MSWK.CF.04 INTRODUCTION TO THE CIF CORE DEFINITIONS. I.D.Brown, Brockhouse Institute for Materials Research, McMaster University, Hamilton, Ontario, Canada L8S 4M1

The original cif core dictionary (Hall, Allen and Brown, 1991) defined cif names for the basic crystallographic quantities, so that cifs could be used to report small molecule crystal structures to Acta Crystallographica C. The dictionary is now in routine use for the submission of structure reports to Acta Cryst. B and C and is being widely incorporated into crystallographic software.

The cif core definitions have proved to be quite robust despite the remarkable changes that have occurred in the philosophy of the crystallographic information file in the intervening years. Some weaknesses in the definitions have been identified and these will be addressed in the extension to the core cif dictionary which is in an advanced stage of preparation.

As of the beginning of 1996 the core cif dictionary is the only one that has been approved, but the core extension is expected to receive approval during 1996.

Hall, S.R., Allen, F.H. & Brown, I.D. (1991) Acta Cryst. A47, 655- 685.