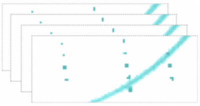


ImgCIF and the Interaction with NeXus

Herbert J. Bernstein, Dowling College, Oakdale, NY USA

Work supported in part by grants from the US Department of Energy, the US National Science Foundation, the US National Institutes of Health and the International Union of Crystallography

ImgCIF Workshops

Tuesday, 14 Aug 2007, 12:45 – 13:45 Barbirolli Room

12:45 – 13:00 Introduction and review of status of imgCIF – Herbert J. Bernstein
13:00 – 13:15 Experience with imgCIF for the Pilatus 6M detector at SLS – Clemens Schulze-Briese
13:15 – 13:30 "imgCIF: a developers perspective" -- Graeme Winter

Friday, 17 Aug 2007, 12:30 – 17:00, room 1.17 Diamond House, Diamond Light Source, Didcot, Chilton

12:30 – 12:45 imgCIF and the interaction with NeXus – Herbert J. Bernstein
12:45 – 13:00 Experience with imgCIF at Diamond – Alun Ashton
13:00 – 13:15 HDF – Mike Folk
13:15 – 13:30 Status of Data Formats in Cryo-electron Microscopy – Matt Dougherty
13:30 – 13:45 (Topic to be announced) – Chris Nielsen
13:45 – 14:00 Break
14:00 – 15:00 Discussion – interaction among imgCIF, NeXus, HDF and XML
15:00 – 15:30 Discussion – open issues in imgCIF
15:30 – 16:30 Formulation of recommendations
16:30 – 16:45 Future plans

Introduction

What is imgCIF/CBF?

1. A clearly defined set of terms to use in describing raw diffraction images (data) and the way in which they were collected (metadata); and
2. A workable and efficient format in which to record, archive and transmit data and metadata; and
3. Support software (e.g. CBFlib)

Status?

Little use, some support (fit2d, mosflm, adxv)

Community uses dozens of distinct formats

Use starting now at SLS and Diamond

Introduction (cont.)

Pending Issues?

Integration with other data frameworks

NeXus, XML, HDF

Integration with SAS

Axis definitions without a goniometer

MiniCBFs

Just the data with
the metadata gathered under just one tag

Jiffies

Cylindrical Detectors

New category for axis origins

Versioning

Why Does imgCIF/CBF exist?

Issue:

Managing massive amounts of data
in structural biology; e.g. in proteomics

Problem:

Many different data formats**.**

Practical Solution:

Provide a common approach to interchange

Crystallographic Information Framework (CIF)

Crystallographic Information Framework (CIF)

Started in 1990 by the International Union of Crystallography to reduce friction in managing structural data

[Hall, Allen, Brown 1991] [Hall, McMahon 2005]

Great success in handling of structural data for small molecules

Clarified the terms for macromolecules

mmCIF, PDBML

Effort to bring similar organization to synchrotron image data

imgCIF/CBF

Other Important Frameworks

CCP4 **[CCP4 1994]**

Collaborative Computational Project, No. 4, (CCP4) is a suite for protein crystallography. It "is a set of separate programs which communicate via standard data files". Many programs have been adapted to CCP4 file formats

XML **[Bray, Paoli, Sperberg-McQueen 98]**

The eXtensible Markup Language is a text-based approach to marking up documents that has become a standard tool for managing data about data (metadata), e.g. in PDBML and CML.

Other Important Frameworks

NeXus **[Klosowski et al. 1998].**

"NeXus is a data format for the exchange of neutron and synchrotron scattering data between facilities and user institutions. It has been developed by an international team of scientists and computer programmers from neutron and X-ray facilities around the world. The NeXus format uses the hierarchical data format (HDF) that is portable, binary, extensible and self-describing. The NeXus format defines the structure and contents of these HDF files in order to facilitate the visualization and analysis of neutron and X-ray data. In addition, an application program interface (API) [was] produced in order to simplify the reading and writing of NeXus files. The details of the format are available at http://www.neutron.anl.gov/NeXus/".

What Makes CIF Different

CIF clearly defines its terms in dictionaries

core CIF, powder CIF, modulated structures CIF, electron density CIF, macromolecular CIF, image CIF, symmetry CIF

The same terms can then be used in multiple formats with no loss of information

e.g. mmCIF vs PDBML

CIF is very database-friendly

CIF is documented (International Tables Volume G)

CIF supports arrays, binary and compression well

What Makes CIF Relevant?

"Diamond intends to provide their PX beam lines with imgCIF data, but to work to having a true [NeXus] infrastructure. Soleil intends to produce imgCIF for their PX beam lines and simply to encapsulate the imgCIF data sets as data (headers and all) within a [NeXus] infrastructure. ISIS is working purely in [NeXus]. At the moment, the critical need seems to be for both [NeXus] and imgCIF instrument definitions. [NeXus] seems to be well equipped with neutron instrument tokens and the X-ray tokens are a work soon to be in progress. …"

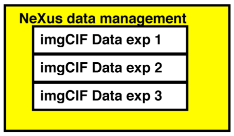
-- from email summary for the group by HJB subsequent to a meeting at Diamond, 10 January 2006

Where Should imgCIF, NeXus and mmCIF fit?

imgCIF dictionary extends mmCIF to manage data images

NeXus to provide a data framework within which to manage multiple sets of experimental data for multiple experiments

mmCIF provides the base within which to manage the entire crystallographic experiment



NeXus data management

imgCIF Data exp 1

imgCIF Data exp 2

imgCIF Data exp 3

Integration of imgCIF with NeXus

ImgCIF is tightly structured

- makes it easy to map from imgCIF to other frameworks with less structure
- makes it hard to map from less structured frameworks to imgCIF

Matches need to be found between

- **NeXus groups and data items versus**
- **imgCIF categories and tags**
  - not just for the items, but for their relationships

Propose to

align NeXus category NXdata
    with ARRAY_DATA as CBF_array_data

insert rest of CBF tree as CBF_… groups

Integration with NeXus (cont.)

Handling binary

- No problem with HDF version of NeXus
- Use binUTF to embed binaries in XML version of NeXus

Going from NeXus to imgCIF

- Need to flatten the NeXus hierarchy to 2 levels
- Map each NeXus class to a new CIF category with an “NX_” prefix
- Add explicit tags pointing to parent categories to link the hierarchy together

---

imgCIF Categories

ARRAY_DATA
  presents the actual numeric data
  (e.g. the numeric values of the pixels in an image)
ARRAY_INTENSITIES
  tells you what you need to do to recover
  intensities from ARRAY_DATA values
ARRAY_STRUCTURE
  how the bits and bytes are organized
 ARRAY_STRUCTURE_LIST
   how the array dimensions are organized
  ARRAY_STRUCTURE_LIST_AXIS
    how axis settings relate to array indices
AXIS
  the physical parameters of each axis

---

imgCIF Categories (cont.)

DIFFRN
  mmCIF category describing diffraction data
 DIFFRN_DATA_FRAME
   details about each frame of data
 DIFFRN_DETECTOR
   information about each detector
  DIFFRN_DETECTOR_AXIS
    information about each detector axis
  DIFFRN_DETECTOR_ELEMENT
    layout of detector elements
 DIFFRN_MEASUREMENT
   goniometer information
  DIFFRN_MEASUREMENT_AXIS
    information about each goniometer axis

---

imgCIF Categories (cont.)

DIFFRN_RADIATION
  incident radiation (crossfire, polarization, etc.)
 DIFFRN_REFLN
   reflection-by-reflection parameters for each frame
 DIFFRN_SCAN
   relationship of axis settings to scans
  DIFFRN_SCAN_FRAME
    relationship of particular frames to scans
   DIFFRN_SCAN_FRAME_AXIS
     relationship of axis settings to particular frames

---

imgCIF Categories (cont.)

Categories under development

AXIS_ROTATION_ORIGIN
  specify direction of axis at zero setting

MAP
  density maps and masks
 MAP_SEGMENT
   bricks, slices and other segments of maps

VARIANT
  identifiers, timestamps and relations for variants

NeXus Classes and Groups

A NeXus class is like a CIF category
A NeXus group is like a row in a CIF category
NXentry
  similar to a CIF data block
NXinstrument

- **NXsource**
- **NXmoderator**
- **NXcrystal**
- **NXdisk_chopper**
- **NXfermi_chopper**
- **NXvelocity_selector**
- **NXguide**
- **NXcollimator**
- **NXaperture**
- **NXfilter**
- **NXattenuator**
- **NXpolarizer**
- **NXflipper**
- **NXmirror**
- **NXdetector**
- **NXbeam_stop**

NXsample
NXmonitor
NXdata
  similar to imgCIF ARRAY_DATA
NXevent_data
NXuser
NXprocess
NXcharacterizations

Integration with SAS

- **1998 -- 2007 -- cif_sas.dic Marc Malfois, Dmitri Svergun, revisions by I. D. Brown**
- **Needs work to align with imgCIF, mmCIF**
- **Major issue: axis definitions**
  - If there is no goniometer the standard imgCIF definition of X does not work. We keep the origin in the sample, but use an X derived from the fast axis of the detector if there is no principal goniometer axis.

---

miniCBFs

- **Issues**
  - Not all information is available at time of data collection
  - Time constraints may limit complexity of what can be written
  - Processing programs may need a minimal header
- **Issues came to a head in setting up for SLS detector**
- **Proposed solution**
  - Write miniCBF at collection
  - Convert to full imgCIF with convert_minicbf program

---

miniCBFs (cont.)

```
###CBF: VERSION 1.5
# CBF file written by CBFlib v0.7.8

data_insulin_pilatus6m

_array_data.header_convention SLS_1.0
_array_data.header_contents
;
# Detector: PILATUS 6M SN: 60-0001
# 2007/Jun/17 15:12:36.928
# Pixel_size 172e-6 m x 172e-6 m
# Silicon sensor, thickness 0.000320 m
# Exposure_time 0.995000 s
# Exposure_period 1.000000 s
# Tau = 194.0e-09 s
# Count_cutoff 1048575 counts
# Threshold_setting 5000 eV
# Wavelength 1.2398 A
# Energy_range (0, 0) eV
# Detector_distance 0.15500 m
# Detector_Voffset -0.01003 m
# Beam_xy (1231.00, 1277.00) pixels
# Flux 22487563295 ph/s
# Filter_transmission 0.0008
# Start_angle 13.0000 deg.
# Angle_increment 1.0000 deg.
# Detector_2theta 0.0000 deg.
# Polarization 0.990
# Alpha 0.0000 deg.
# Kappa 0.0000 deg.
# Phi 0.0000 deg.
# Chi 0.0000 deg.
# Oscillation_axis  X, CW
# N_oscillations 1
;

_array_data.data
;
--CIF-BINARY-FORMAT-SECTION--
Content-Type: application/octet-stream;
    conversions="x-CBF_BYTE_OFFSET"
Content-Transfer-Encoding: BINARY
X-Binary-Size: 6247567
X-Binary-ID: 1
X-Binary-Element-Type: "signed 32-bit integer"
X-Binary-Element-Byte-Order: LITTLE_ENDIAN
Content-MD5: 8wO6i2+899If5iO8QPdgrw==
X-Binary-Number-of-Elements: 6224001
X-Binary-Size-Fastest-Dimension: 2463
X-Binary-Size-Second-Dimension: 2527
X-Binary-Size-Padding: 4095

...[binary data here]...
```

Jiffies

- Small programs to convert to/from imgCIF
- Recommended by Hawaii imgCIF workshop in 2006
- Conversion to imgCIF from MAR, ADSC and SLS miniCBFs working
- Work on Bruker-AXS starting
- 2 year project funded by DOE starting now

Cylindrical Detectors

- **Issue raised by James Hester**
  - Can be done using the current dictionary but requires implicit definition of the origin direction for rotation axes.
  - **Better approach**
    - new category, AXIS_ROTATION_ORIGIN
    - give a unit vector at right angles to the rotation axis representing the direction of the rotation axis at its zero setting.

Cylindrical Detectors (cont.)

```
loop_
_axis.id
_axis.type        #___
_axis.equipment   #___\___________
_axis.depends_on  #___|___________\____________
_axis.vector[1]   #___|___________|____________\______________
_axis.vector[2]   #___|___________|____________|______________\__
_axis.vector[3]   #___|___________|____________|______________|__\__
_axis.offset[1]   #___|___________|____________|______________|__|__\____
_axis.offset[2]   #___|___________|____________|______________|__|__|____\__
_axis.offset[3]   #___|___________|____________|______________|__|__|____|__\__
                  #   |           |            |              |  |  |    |  |  ¥
 DETECTOR_Z        translation detector        .              0  0 -1    0  0  0
 DETECTOR_Y        translation detector      DETECTOR_Z       0 -1  0    0  0  0
 DETECTOR_PITCH    rotation    detector      DETECTOR_Y       1  0  0    0  0  0
                  #   |           |            |              |  |  |    |  |  |
 ELEMENT_CAXIS     translation detector      DETECTOR_PITCH   1  0  0
                                                                     -99.95 0  0
 ELEMENT_ANGLE     rotation    detector      ELEMENT_CAXIS    1  0  0    0  0  0
 ELEMENT_RADIUS    translation detector      ELEMENT_ANGLE    0 0.341913983 -0.93973125
                                                                         0  0  0
######################|###########|############|##############|##|##|####|##|##|

loop_
_axis_rotation_origin.axis_id
_axis_rotation_origin.direction[0] #__
_axis_rotation_origin.direction[1] #__¥__________
_axis_rotation_origin.direction[2] #__|__________\____________
                                   #  |          |            \
DETECTOR_PITCH                        0          0             1
ELEMENT_ANGLE                         0          0.341913983 -0.93973125
```

Versioning

After (or even during) initial data collection we may have multiple versions of the same items, e.g. beam center

```
loop_
_variant.variant
_variant.role
_variant.timestamp
_variant.variant_of
_variant.details
.           "raw data"  2007-08-03T23:20:00 . .
indexed     "preferred" 2007-08-04T01:17:28 .
              "indexed cell and refined beam center"
```

Software status

CBFlib (http://www.bernstein-plus-sons.com/CBF) provides

- API (C function library, under GPL or LGPL, your choice)
  - more compressions, major speedup, support for maps
- Manual
- Sample files
- Utilities (under GPL only)
  - convert_image (works for Mar or ADSC)
  - convert_miniCBF (e.g. for SLS images)
  - cif2cbf
  - vcif2

mosflm (http://www.mrc-lmb.cam.ac.uk/harry/mosflm/) supports imgCIF

adxv (http://www.scripps.edu/~arvai/adxv.html) supports imgCIF

People Involved

Frances C. Bernstein

imgCIF Workshops:
- **Dowling:** HJB
- **BNL:** Robert M. Sweet

ARCiB Lab:
- **Dowling College:** HJB, Isaac Awuah Asiamah, Darina Boycheva, Georgi Darakev, Nikolay Darakev, John Jemilawon, Nan Jia, Greg McQuillan, Dan O'Brien, Georgi Todorov

SVEVSL Project:
- **Dowling College: ARCiB Lab group**
- **RIT:** Paul A. Craig, Jared Carter, Brett Hanson, Scott Mottarella, Charlie Westin

And many more over the years.

Additional Material

Additional material to help in understanding imgCIF is provided on the slides that follow

---

Status of CIF in PX

Core CIF (used for ligands):

An effective, working standard, heavily used

mmCIF (macromolecular structures):

Some use, community prefers old PDB format.

imgCIF (synchrotron data images, other images):

Little use, some support (fit2d, mosflm, adxv)

Community uses dozens of distinct formats

Use starting now at SLS and Diamond

Problem

Use of many different data formats causes delay and confusion; may lead to errors

Synchrotron image data formats:

Wladek Minor is "dealing with 197 (!!!!) frame formats" [email Wladek Minor to H. J. Bernstein, 15 May 2006].

2003 Denzo manual lists 107 available detector formats [Gewirth 2003] [Otwinowski, Minor 1997]

An Impractical Solution
(The Esperanto Solution)

Mandate one perfect format for everyone to use for everything (internals and externals)

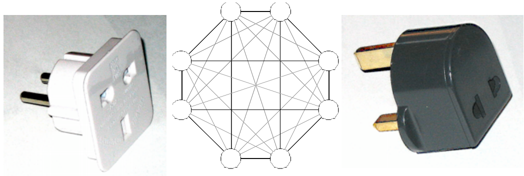
Not workable:

- **Might suppress new ideas and good science**
- **There is no agreement on the "perfect" format**
  - **What is perfect for the internals of one project might be imperfect for the internals of another project**

A Solution: Focus on Interchange

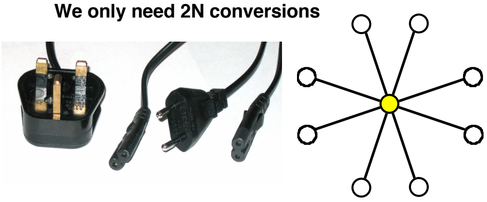
Include adapters with/in each system to handle many formats

For N systems with N-1 foreign formats, need N*(N-1) adapters



A Better Solution

Provide a common format for interchange (e.g. CIF)
- **Instead of N*(N-1) conversions**
- **We only need 2N conversions**



The Basics of imgCIF

There are multiple types of CIF
- **DDL1 CIFs (e.g. coreCIF, pdCIF)**
- **DDL2 CIFs (e.g. mmCIF, imgCIF)**
- **DDL3 is coming**

CIF Dictionaries define the terms that can be used and their relationships.

Users can add terms of their own, but you should not use an existing term with a meaning that conflicts with the meaning in a dictionary or in a way that could be confused with terms that have been officially adopted.

For all CIFs:

- **Information is organized into blocks of data**
- **Each block of data is managed essentially in terms of tables**
  - **Tables are called "categories" or "loops"**
  - **The column headings are called tags" or "data names"**
  - **Some tables have only one row of data**
    - **then each tag can be put with its value**
  - **Some tables have multiple rows of data**

- **A given tag can appear only once in a block**

- **DDL1 CIFs treat all categories similarly**
- **DDL2 CIFs explicitly state relationships**
  - **e.g. parent-child relationships**

imgCIF is a DDL2 dictionary that extends the macromolecular CIF (mmCIF) dictionary.

CIF Syntax

A collection of data blocks

Each data block contains data names (tags) and their values

White space delimits tokens

Tags start with a leading underscore ("_") to distinguish them from values
- **Values that might be confused with data names or keywords or that contain whitespace are quoted**

Quoting
- **single quote (single line only)**
- **double quote (single line only)**
- **semicolon in column 1 (multiple lines OK)**

- **terminal quote mark must be followed by whitespace**

Characters with special meaning
- **Underscore**
- **Quote marks**
- **Period (“.”) or question mark (“?”) (null value)**
- **Hash mark (“#”) (comment)**

Reserved words
- **"global_", "data_", "loop_", "stop_", and "save_"**

In addition to the underscore, and the three quote marks, three other characters have special meaning: the period ("."), the question mark ("?") and the hash mark ("#"). The period is used when no value is specified. The question mark is used when a value is desired but not available. The hash mark indicates that the remaining characters on that line are part of a comment.

There are a small number of reserved words:
"global_", "data_", "loop_", "stop_", and "save_".
The last two reserved words are not used by CIF but are reserved to prevent conflict with the language from which CIF is derived (STAR). "global_" and "data_" mark the start of a data block.

"data_" should be followed immediately with the name of the block, without intervening whitespace.

If "loop_" appears, it is followed by a sequence of tags without intervening data values. Those tags are considered as the column headings of a table. These are followed by rows of data values corresponding to those column headings.

Outside of a table, tags and data values appear in simple alternation. Within a data block a given tag may appear only once.

The meaning of a CIF document is not altered by changing the order of presentation of data blocks nor is it altered by changing the order of presentation of tags within a block.

There are two styles of CIF in use for crystallography: DDL1 and DDL2.

DDL1 CIF (e.g. coreCIF, pdCIF)

Partial example of a small molecule coordinate list [Longridge 98]

```
loop_
_atom_site_label
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_U_iso_or_equiv
_atom_site_adp_type
_atom_site_calc_flag
_atom_site_refinement_flags
_atom_site_occupancy
_atom_site_disorder_assembly
_atom_site_disorder_group
_atom_site_type_symbol
Fe1 1 0 1 .0084(2) Uani d S 1 . . Fe
Na1 .50907(11) .13980(8) 1.09450(9) .0185(3) Uani d . 1 . . Na
Na2 .89904(10) .37128(8) 1.21657(9) .0171(3) Uani d . 1 . . Na
C1 .7997(2) -.01740(18) 1.0419(2) .0110(4) Uani d . 1 . . C
N1 .6788(2) -.02885(18) 1.0696(2) .0166(4) Uani d . 1 . . N
C2 .9306(3) -.01004(16) .8075(3) .0130(4) Uani d . 1 . . C
```

DDL2 CIF (e.g. mmCIF, imgCIF)

Partial example of a macromolecular CIF (1CRN) as converted to mmCIF by the program pdb2cif [Bernstein et al. 98]

```
loop_
_atom_site.label_seq_id
_atom_site.group_PDB
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.auth_seq_id
_atom_site.label_alt_id
_atom_site.cartn_x
_atom_site.cartn_y
_atom_site.cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.footnote_id
_atom_site.label_entity_id
_atom_site.id
1
ATOM N N THR * 1 . 17.047 14.099 3.625 1.00 13.79 . 1 1
1
ATOM C CA THR * 1 . 16.967 12.784 4.338 1.00 10.80 . 1 2
```

ImgCIF Binary Data

```
_array_structure.id ARRAY1
_array_structure.encoding_type "signed 32-bit integer"
_array_structure.compression_type packed
_array_structure.byte_order little_endian
_array_data.array_id ARRAY1
_array_data.binary_id 1
_array_data.data
;
--CIF-BINARY-FORMAT-SECTION--
Content-Type: application/octet-stream;
conversions="x-CBF_PACKED"
Content-Transfer-Encoding: BINARY
X-Binary-Size: 3745758
X-Binary-ID: 1
X-Binary-Element-Type: "signed 32-bit integer"
Content-MD5: 1zsJjWPfol2GYI2V+QSXrw==
’ ∏P«q «q FA• f¡Æ• àR˜u<ˇ›k2˙b |5ß …
```

Background

Imagine trying to do electronics without any agreement as to how to draw schematics.

That is how we have managed massive amounts of data in structural biology.

Some points of coherence in the chaos:
- BNL Protein Data Bank format
  - [Bernstein et al. 1977]
  - *de facto* standard for macromolecular coordinates
- CCDC's Cambridge Structural Database
  - [Allen et al. 1973]
  - de facto standard for small molecules

Background (cont.)

In general
  Each vendor of data collection equipment
  Each group maintaining a software package
has their own approach to
  representing and storing raw data

This causes delays and confusion -- friction -- in working with data in so many different formats

---

Is the Problem Real?

Date: Sat, 06 Jan 2007 16:32:12 -0500
From: Arun Malhotra <malhotra@miami.edu>
To: CCP4 Bulletin Board <ccp4bb@dl.ac.uk>

“One common source of errors is changes made in reflection or coordinate files… For example, only recently, someone in my lab just manually edited out a hkl format file to change a few reflections from the exponential format into the standard F format, so that it could be converted into the mtz format. … Format conversions (hkl to mtz or vice-versa) or simple manual edits to coordinate files are very common, and are fertile places for mistakes to creep in. Once such mistakes are made, they are not often easy to catch since there is no easy way to compare files.…”

Is the Problem Real?

Date: Sun, 07 Jan 2007 11:25:26 +0900
From: Charlie Bond <Charles.Bond@uwa.edu.au>
To: ccp4bb@dl.ac.uk

“… Often at the synchrotron one is in a tired hurry to get an image indexed and processed. If the wrong parameters are used (eg the ones from the home lab with a bit of editing), a dataset in the wrong hand can be quickly produced

“Increasingly beamlines automatically prepare the correct parameter files for you, but it is cases where images are difficult to process (low resolution, disorder) that processing may occur later at home and the details of the beamline may be disregarded.

“Correct me if I'm wrong, but even the deposition of images would not help this as the critical information is the geometry of the beamline set up which is probably not recor[d]ed with the images.

…”

Frequently Asked Questions

Can I change it?

Yes, please do. We would appreciate:
- **New ideas**
- **New items for the dictionary**
- **New support software**
- **Bug fixes and improvements**
- **Translations to and from other presentations**

Please don’t use existing terms in ways that conflict with their meanings; define a new term with a new name instead

---

Frequently Asked Questions (Cont.)

The BIG Frequently Asked Question Can I make proprietary software using imgCIF and CBFlib?

Yes, the API in CBFlib is available under the LGPL.

If you change CBFlib itself, you must publish the changed source code under the LGPL, but even if you change CBFlib, you do not have to make your program into an open source program.

How to Make Changes to the imgCIF Dictionary

1. Get the best current version of the dictionary from the IUCr

2. Check that what you propose is not already there, or if there is at least an appropriate category

3. To avoid conflicts with others doing the same thing, get a prefix from Brian McMahon (bm@iucr.org)

4. If you are going to be sending files to other people, discuss your new definition with them and, please, on the imgcif-l list

5. If this will remain just a local change, use it in good health

6. If you think this should be added to the main dictionary for community use, please say so on the imgcif-l list, and, if appropriate, on other lists

7. If there is sentiment to add it to the main imgCIF dictionary, we will post a revised dictionary for comments, and then, if the dictionary working group agrees, forward the dictionary to COMCIFS for adoption

How to Use and Make or Propose Changes to CBFlib

Use:

1. **Download the package (source or binary)**
2. **If source, build for your machine**
3. **If you need help building, contact** yaya@dowling.edu
4. **If you are using the utilities, install them in your favorite location for binaries and use them**
5. **If you are building an application against the API, install the library in your favorite location and use it**

Changes:

1. **Changes in your own programs that just use the API:**
   Just do it (LGPL)
2. **Changes to the API or Program**
   Do it, but follow the GPL/LGPL rules on changes
   (making source available, carrying the license forward)

Credit

We would appreciate a credit and knowing about changes.
Please cite [Bernstein, Ellis 2005] (see below)

Where to Find imgCIF Information

IUCr Crystallographic Information Framework:
International Tables, Volume G
http://www.iucr.org/iucr-top/cif/index.html
official copies of dictionaries and stable releases of software
Image CIF/Crystallographic Binary File (imgCIF/CBF)
http://arcib.dowling.edu/CBF
http://www.bernstein-plus-sons.com/software/CBF
development versions of dictionary and software
http://www.iucr.org/iucr-top/cif/cbf/imgcif-l
http://scripts.iucr.org/mailman/listinfo/imgcif-l
imgCIF discussion list (please join)
Management of Experimental Data in Structural Biology (MEDSBIO)
http://www.medsbio.org
A broader perspective (imgCIF, NeXus, …) concentrating on interfaces
http://www.medsbio.org/meetings
information on this workshop and future ones of interest
http://scripts.iucr.org/pipermail/medsbio-l/
http://scripts.iucr.org/mailman/listinfo/medsbio-l
MEDSBIO discussion list (please join)
Protein Data Bank
http://www.pdb.org
Information on dictionaries and file format, BioSync, etc.

References

[Allen et al. 1973] Allen, F. H., Kennard, O., Motherwell, W. D. S., Town, W. G. & Watson, D. G. (1973), "Cambridge Crystallographic Data Centre. ii. Structural Data File", *J. Chem. Doc.* 13, 119 – 123.

[Berman et al. 2000] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000), 'The Protein Data Bank', Nucleic Acids Research 28, 235 – 242.

[Bernstein et al. 1977] Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977), 'The protein data bank: a computer based archival file for macromolecular structures', *J. Mol. Biol.* 112, 535 – 542.

[Bernstein 2005] Bernstein, H. J. (2005) "The Classification of Image Data", chapter 3.7 in "*International Tables For Crystallography, Volume G: Definition and Exchange of Crystallographic Data*," S. R. Hall and B. McMahon, eds. International Union of Crystallography, Springer, Dordrecht, NL, pp. 199 – 205.

References (Cont.)

[Bernstein, Bernstein 2005] Bernstein, H. J., Bernstein, F. C. (2005) "Databanks of Macromolecular Structure", chapter 4 in "*Database Annotation in Molecular Biology*," A. M. Lesk, ed., Wiley, Chichester, UK, pp. 63 –79.

[Bernstein, Ellis 2005] Bernstein, H. J., Ellis, P. J. (2005). "CBFlib: An ANSI C Library for Manipulating Image Data", chapter 5.6 in "*International Tables For Crystallography, Volume G: Definition and Exchange of Crystallographic Data*," S. R. Hall and B. McMahon, eds., International Union of Crystallography, Springer, Dordrecht, NL, pp. 544 – 556."

[Bernstein, Hammersley 2005] Bernstein, H. J., Hammersley, A. P. (2005) "Specification of the Crystallographic Binary File (CBF/imgCIF)", chapter 2.3 in "*International Tables For Crystallography, Volume G: Definition and Exchange of Crystallographic Data*," S. R. Hall and B. McMahon, eds., International Union of Crystallography, Springer, Dordrecht, NL, pp. 37 – 43.

[Bernstein et al. 98] Bernstein, H. J., Bernstein, F. C., Bourne, P. E. (1998) "CIF Applications. VIII. pdb2cif: Translating PDB Entries into mmCIF Format", J. Appl. Crystallography, Vol.31, pp. 282 -- 295.

References (Cont.)

[Bray, Paoli, Sperberg-McQueen 98] Bray, T., Paoli, J., Sperberg, C. M., eds, "Extensible Markup Language (XML)", W3C Recommendation 10-Feb-98, REC-xml-19980210, http://www.w3.org/TR/1998/REC-xml-19980210

[CCP4 1994] COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4. 1994. ``The CCP4 Suite: Programs for Protein Crystallography''. *Acta Cryst.* D50, 760-763.

[Gewirth 2003] Gewirth, D. (2003). "*THE HKL MANUAL, A Description of the Programs Denzo, XDisplayF, Scalepack An Oscillation Data Processing Suite for Macromolecular Crystallography*," 6th ed. (written with the cooperation of the program authors Zbyszek Otwinowski and Wladek Minor, revised and updated by Wladyslaw Majewski", http://www.hkl-xray.com/hkl_web1/hkl/manual_online.pdf

[Hall, Allen, Brown 1991] Hall, S. R., Allen, F. H. & Brown, I. D. (1991), 'The crystallographic information file (CIF): a new standard archive file. for crystallography', *Acta Cryst.* A47, 655 – 685.

References (Cont.)

[Hall, McMahon 2005] Hall, S. R. & McMahon, B. (2005), *Volume G: Definition and Exchange of Crystallographic Data, International Tables For Crystallography*, Springer: Dordrecht, chapter 1.1. Genesis of the Crystallographic Information File

[Hammersley, Bernstein, Westbrook 2005] Hammersley, A. P., Bernstein, H. J., Westbrook, J. D. (2005). "Image Dictionary (imgCIF)", chapter 4.6 in "*International Tables For Crystallography, Volume G: Definition and Exchange of Crystallographic Data*," S. R. Hall and B. McMahon, eds., International Union of Crystallography, Springer, Dordrecht, NL, pp. 444 – 458.

[Klosowski et al. 1998] Klosowski, P.; Koennecke, M.; Tischler, J. Z.; Osborn, R. (1998). "NeXus: A common format for the exchange of neutron and synchrotron data", *Physica B: Physics of Condensed Matter*, 241,1-4, pp. 151-153.

[Otwinowski, Minor 1997] Otwinowski, Z., Minor, W. (1997). "Processing of X-ray Diffraction Data Collected in Oscillation Mode", *Methods in Enzymology, 276: Macromolecular Crystallography*, part A, pp. 307 -- 326.

References (Cont.)

[Powell 2001] Powell, H. (2001), “Recent improvements to Mosflm - version 6.11”, *CCP4 Newsletter on Protein Crystallography*, 39, 18. See http://www.ccp4.ac.uk/newsletters/newsletter39/18_mosflm.html.

[Westbrook et al. 2003] Westbrook, J., Feng, Z., Chen, L., Yang, H., Berman, H. M.. (2003) “The Protein Data Bank and structural genomics,” *Nucleic Acids Research,* Vol. 31, No. 1 489-4

ADDITIONAL READING

[BIOXHIT 2004]. “Bioxhit: biocrystallography (X) on a highly integrated technology platform for European structural genomics,” EU Genomics News, No. 3, November 2004. See http://icarus.embl-hamburg.de/bioxhit/index.html

[Fitzgerald et al. 2005] Fitzgerald, P. M. D., Westbrook, J. D., Bourne, P. E., Mcmahon, B., Watenpaugh, K. D. (2005). “Macromolecular dictionary (mmCIF)”, chapter 4.5 in “International Tables For Crystallography, Volume G: Definition and exchange of crystallographic data,” Vol. G, S. R. Hall and B. McMahon, eds., International Union of Crystallography, Heidelberg: Springer, pp. 444 – 458.

[Szebenyi, Arvai, Ealick, Laluppa, Nielsen, 1997] Szebenyi, D. M. E., Arvai, A., Ealick, S., Laluppa, J. M., Nielsen, C. (1997) “A System for Integrated Collection and Analysis of Crystallographic Diffraction Data”, J. Synchrotron Rad. 4, 128-135. For adxv see http://www.scripps.edu/~arvai/adxv.html