Championing data standards in chemical crystallography with CIF

Kamil Filip Dziubek
Institute of Mineralogy and Crystallography
University of Vienna, Austria

IUCr Representative to CODATA
IUCr Committee on Data (CommDat) member

International Union of Crystallography: a scientific union

The International Union of Crystallography (IUCr) is a non-profit scientific union serving the worldwide interests of crystallographers and other scientists employing crystallographic methods

- 55 National Committees make up the General Assembly
- The scientific work is conducted through Commissions and Committees

> Paul Ewald of the Polytechnic Institute of Brooklyn, President of the International Union of Crystallography (a scientific union, not a labor union)
>
> Elizabeth A. Wood, *Crystals and Light: An Introduction to Optical Crystallography*, D. Van Nostrand, Princeton 1964.

- The **Committee on Data (CommDat)** work with the IUCr's Commissions, including the Commission on Journals, having a coordinating and advisory role regarding data
- CommDat exist alongside, and have a formal relationship with, the Committee for the Maintenance of the CIF Standard (COMCIFS)

- The **Committee for the Maintenance of the CIF Standard (COMCIFS)** oversees the development of the CIF and reports to the Executive Committee of the IUCr
- Drafts of new definitions and official dictionaries are submitted to COMCIFS for technical validation and ultimately for approval

IUCr International Union of Crystallography

Simon Coles
CommDat chair

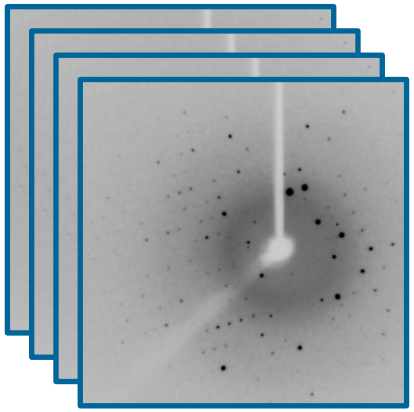
James Hester
COMCIFS chair

75 IUCr 1948 – 2023

Flavors and data cycle of crystallographic data

Crystallographic data lifecycle

- Raw data images collected in a diffraction experiment
- Processed data (unit-cell parameters and indexed Bragg reflections with associated integrated intensities)
- Crystal structure (symmetry, unit-cell parameters, atomic coordinates, ADPs)



Raw (primary) data

```
loop_
 _refln_index_h
 _refln_index_k
 _refln_index_l
 _refln_F_squared_calc
 _refln_F_squared_meas
 _refln_F_squared_sigma
 _refln_observed_status
   2    0    0      25185.30      25637.90      369.48 o
   4    0    0       1787.58       1693.17       29.40 o
   6    0    0         59.02         71.69        4.22 o
   8    0    0         27.82         14.56        3.68 o
  10    0    0        483.46        450.47       23.58 o
   1    1    0       7357.92       7392.33       96.02 o
   2    1    0         27.32         30.41        1.75 o
   3    1    0       1335.00       1608.41       31.10 o
   4    1    0       9037.50       8887.28       88.78 o
   5    1    0       4968.33       5067.08       84.21 o
   6    1    0         29.52         35.79        2.56 o
   7    1    0          4.53          6.29        2.14 o
   8    1    0          1.53          3.02        2.39 o
   9    1    0         62.03         58.07        4.39 o
```

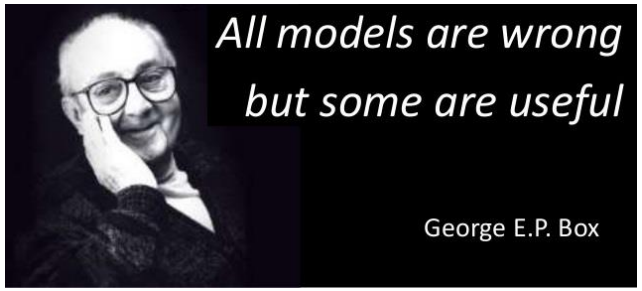
Processed data

```
loop_
 _atom_site_label
 _atom_site_type_symbol
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_U_iso_or_equiv
 _atom_site_adp_type
 _atom_site_occupancy
 _atom_site_symmetry_multiplicity
 _atom_site_calc_flag
 _atom_site_refinement_flags
 _atom_site_disorder_assembly
 _atom_site_disorder_group
C1 C 0.10650(11) 0.6028(2) 0.15602(9) 0.0265(4) Uani 1 1 d . . .
C2 C 0.07667(12) 0.6156(2) 0.09181(10) 0.0346(5) Uani 1 1 d . . .
H2 H 0.0274 0.6713 0.0837 0.042 Uiso 1 1 calc R . .
C3 C 0.11916(13) 0.5466(3) 0.03925(10) 0.0414(5) Uani 1 1 d . . .
H3 H 0.0986 0.5546 -0.0049 0.050 Uiso 1 1 calc R . .
C4 C 0.19043(12) 0.4672(3) 0.05067(10) 0.0390(5) Uani 1 1 d . . .
H4 H 0.2188 0.4203 0.0144 0.047 Uiso 1 1 calc R . .
C5 C 0.22184(11) 0.4543(2) 0.11478(10) 0.0326(5) Uani 1 1 d . . .
H5 H 0.2715 0.3992 0.1223 0.039 Uiso 1 1 calc R . .
C6 C 0.18010(10) 0.5223(2) 0.16750(9) 0.0253(4) Uani 1 1 d
```

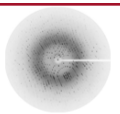
Derived (structural) data

Crystal structures are models!

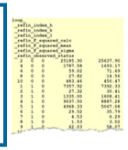
- The actual raw data in X-ray crystallography are diffraction data leading to electron density
- Atom positions that make up a crystal structure are a model that attempts to explain the raw data in as complete a fashion as possible



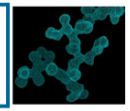
Crystal → **Data collection** → *Diffraction images*
Measure diffraction intensities from a crystal



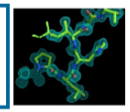
Diffr. images → **Data processing** → *Intensities*
Calculate magnitude and uncertainties of intensities
Reduce the data to the proper symmetry group



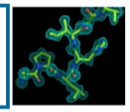
Intensities → **Phasing** → *Density map*
Retrieve phases to calculate electron density map



Density map → **Model building** → *Model*
Build and adjust a model in the el. density map



Model → **Refinement** → *Refined model*
Fit model parameters values to diffraction data



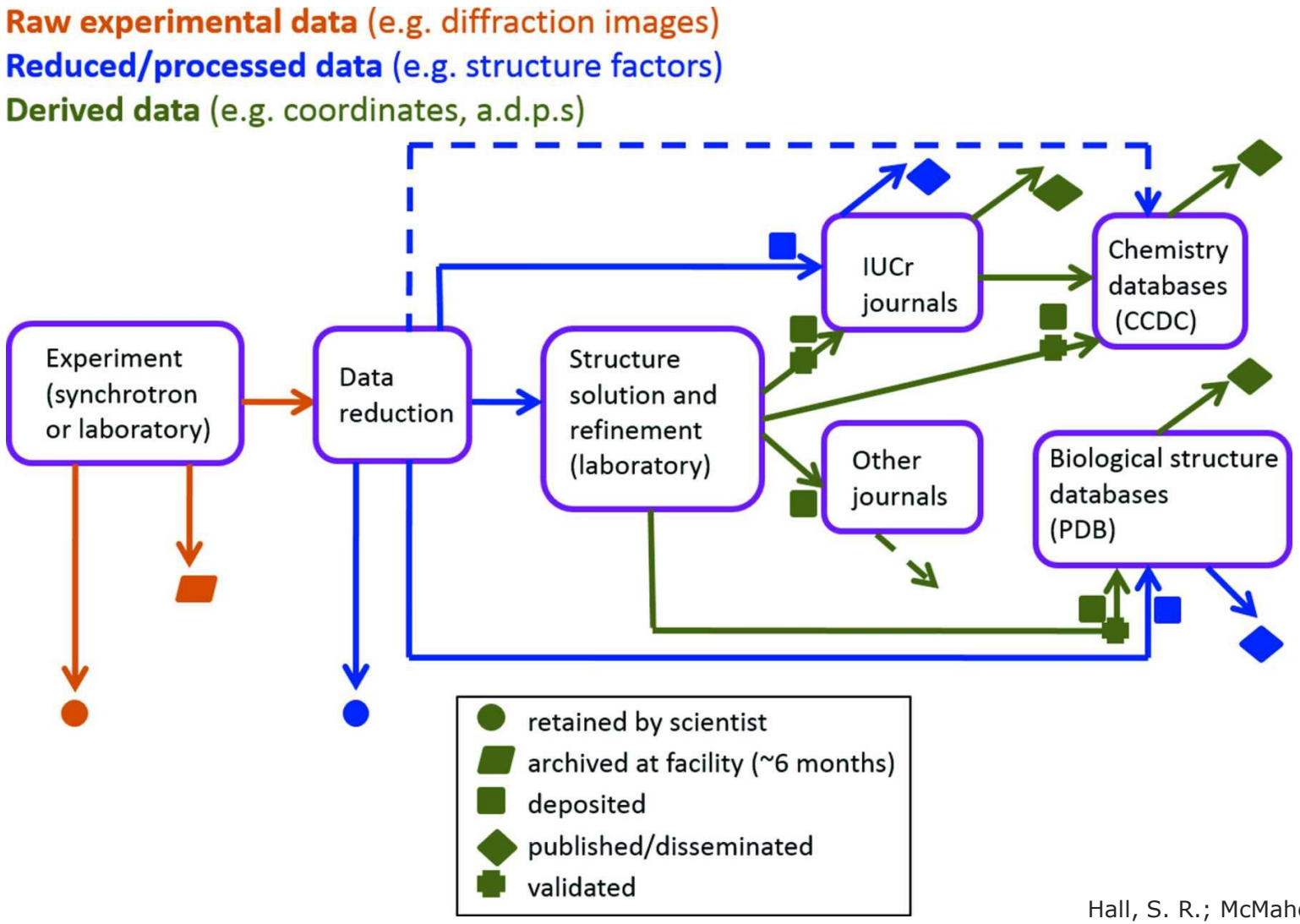
Refined model → **Validation** → *Final model*
Fit model parameters values to diffraction data

checkCIF

Feedback exists but relies on user
Automatic feedback is used/implemented
Feedback proposed, not used/implemented
Automatic feedback is proposed/possible

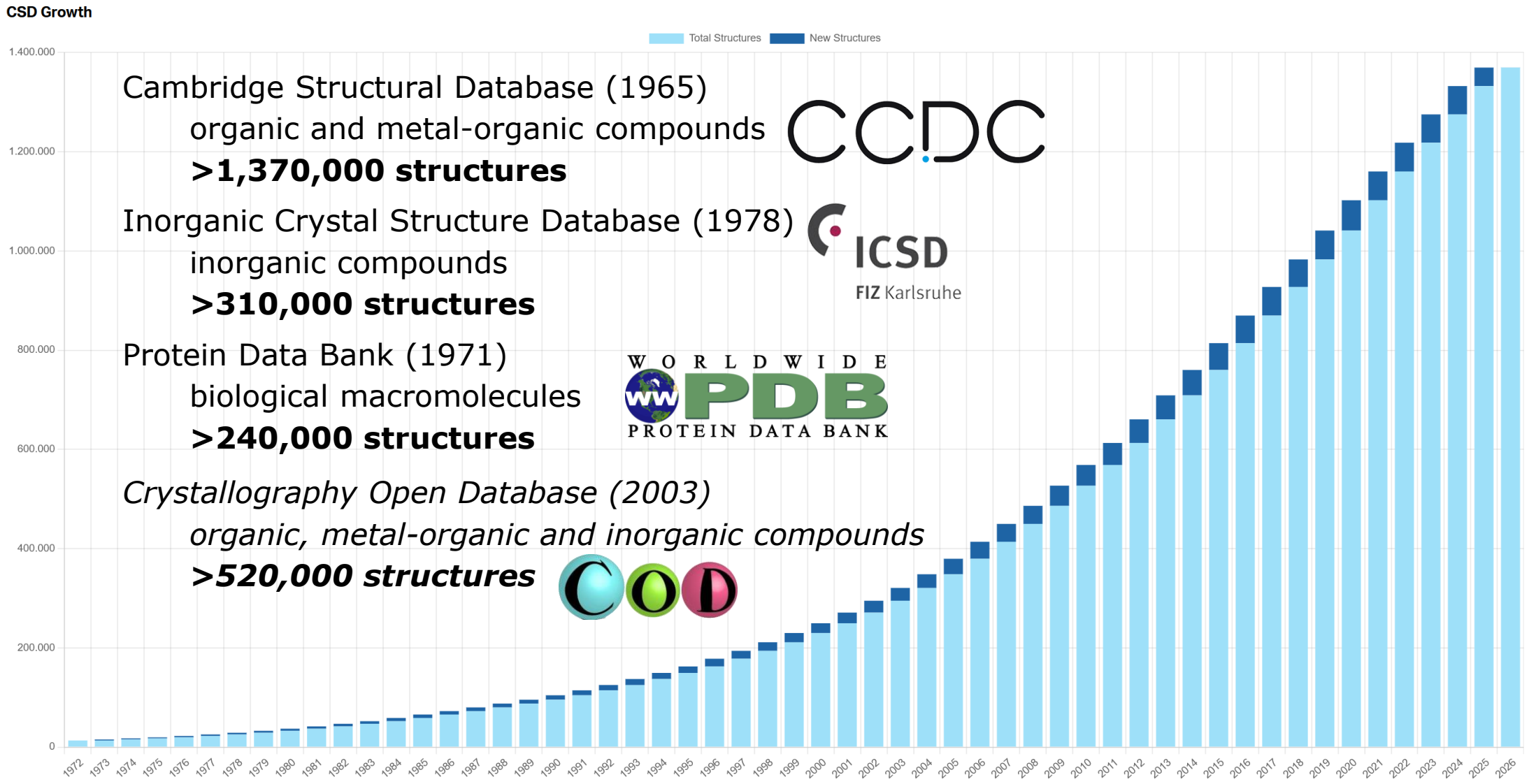
Lamzin, V. L.; Perrakis, P., *Nat. Struct. Biol.* **2000**, *7*, 978–981.

Coherent information flow in crystallography



Hall, S. R.; McMahon, B. *Data Sci. J.* **2016**, *15*, 3.

Databases in the data deluge era



Cambridge Structural Database (1965)
organic and metal-organic compounds
>1,370,000 structures

Inorganic Crystal Structure Database (1978)
inorganic compounds
>310,000 structures

Protein Data Bank (1971)
biological macromolecules
>240,000 structures

Crystallography Open Database (2003)
organic, metal-organic and inorganic compounds
>520,000 structures

Image credit: www.ccdc.cam.ac.uk

In 1991…

"All the News That's Fit to Print"

The New York Times

Late Edition

VOL.CXLI... No. 48,826 NEW YORK, THURSDAY, DECEMBER 26, 1991 50 CENTS

GORBACHEV, LAST SOVIET LEADER, RESIGNS; U.S. RECOGNIZES REPUBLICS' INDEPENDENCE

RETAILERS REPORT SALES FELL SHORT OF DIM FORECASTS

Last-Minute Buying Spree Fails to Carry Merchants Ahead of Last Year's Receipts

Communist Flag Is Removed; Yeltsin Gets Nuclear Controls

BUSH LAUDS VISION OF SOVIET LEADER

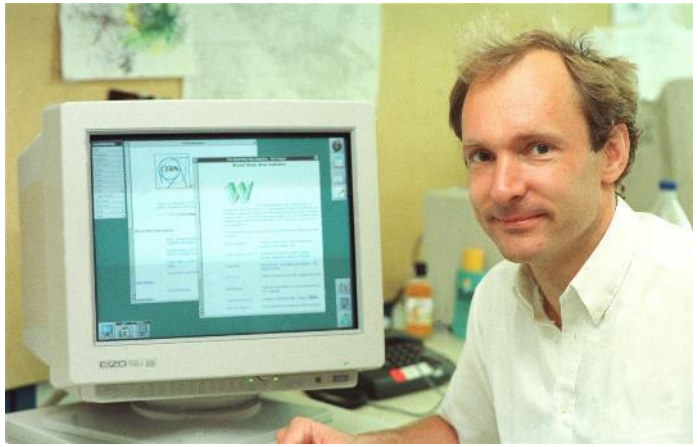
In Farewell, President Cites Gorbachev's Historic Role

The Soviet State, Born of a Dream, Dies

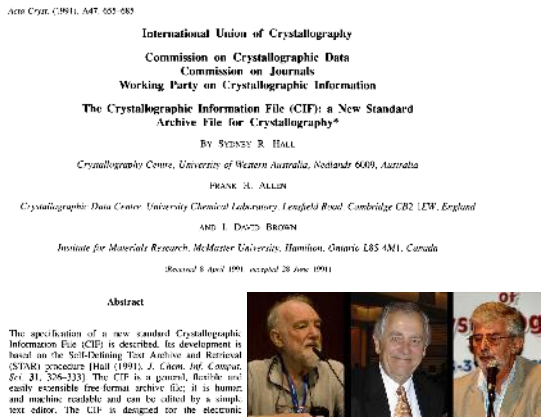
End of an Empire

…USSR passes into history

…Operation Desert Storm began



…World Wide Web publicly debuts
(HTML, XML and Java were yet to come!)



International Union of Crystallography

Commission on Crystallographic Data
Commission on Journals
Working Party on Crystallographic Information

The Crystallographic Information File (CIF): a New Standard Archive File for Crystallography*

By Sydney R. Hall

Crystallography Centre, University of Western Australia, Nedlands 6009, Australia

Frank H. Allen

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

and I. David Brown

Institute for Materials Research, McMaster University, Hamilton, Ontario L8S 4M1, Canada

Abstract

…IUCr introduced the Crystallographic Information File

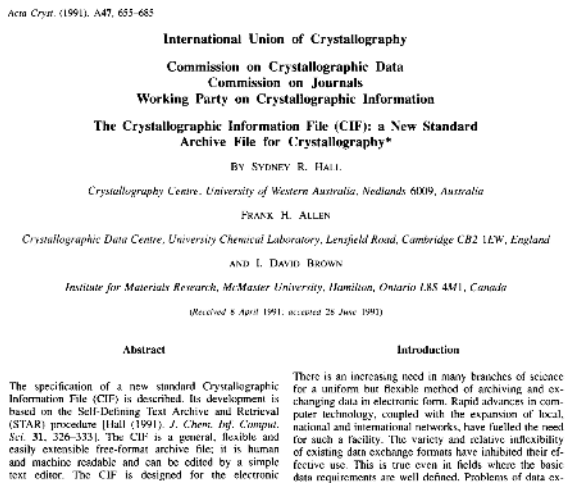
Image credits: Wikimedia Commons, NYT archive, IUCr Newsletter

What is CIF and how is it used?

Crystallographic Information Framework (CIF) is a data exchange format for crystallographic and related structural science data

- The CIF format and extensible CIF dictionaries are defined, adopted and curated by the IUCr
- Until 2005 the acronym stood for 'Crystallographic Information File', but the name was modified in recognition of its application across crystallography and all related structural science fields
- CIF offers domain-specific ontology (collection of data identifiers, attributes and relationships)
- Simple syntax well suited to archive and exchange purposes (human- and machine-readable)
- Adopted by major databases (CSD, ICSD, COD, RRUFF, Mindat) and required as ESI in most journals
- Can also be adopted for the hosting structure factors / reflection intensity data
- imgCIF/CBF can handle binary data (including raw data images) and is isomorphous to CIF
- Canonical information on CIF contained in International Tables Vol. G

```
loop_
 _atom_site_label
 _atom_site_type_symbol
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_U_iso_or_equiv
 _atom_site_adp_type
 _atom_site_occupancy
 _atom_site_symmetry_multiplicity
 _atom_site_calc_flag
 _atom_site_refinement_flags
 _atom_site_disorder_assembly
 _atom_site_disorder_group
C1 C 0.10650(11) 0.6028(2) 0.15602(9) 0.0265(4) Uani 1 1 d . . .
C2 C 0.07667(12) 0.6156(2) 0.09181(10) 0.0346(5) Uani 1 1 d . . .
H2 H 0.0274 0.6713 0.0837 0.042 Uiso 1 1 calc R . .
C3 C 0.11916(13) 0.5466(3) 0.03925(10) 0.0414(5) Uani 1 1 d . . .
H3 H 0.0986 0.5546 -0.0049 0.050 Uiso 1 1 calc R . .
C4 C 0.19043(12) 0.4672(3) 0.05067(10) 0.0390(5) Uani 1 1 d . . .
H4 H 0.2188 0.4203 0.0144 0.047 Uiso 1 1 calc R . .
C5 C 0.22184(11) 0.4543(2) 0.11478(10) 0.0326(5) Uani 1 1 d . . .
H5 H 0.2715 0.3992 0.1223 0.039 Uiso 1 1 calc R . .
C6 C 0.18010(10) 0.5223(2) 0.16750(9) 0.0253(4) Uani 1 1 d . . .
```


Acta Cryst. (1991). A47, 655–685

International Union of Crystallography

Commission on Crystallographic Data
Commission on Journals
Working Party on Crystallographic Information

The Crystallographic Information File (CIF): a New Standard Archive File for Crystallography*

BY SYDNEY R. HALL

Crystallography Centre, University of Western Australia, Nedlands 6009, Australia

FRANK H. ALLEN

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

AND I. DAVID BROWN

Institute for Materials Research, McMaster University, Hamilton, Ontario L8S 4M1, Canada

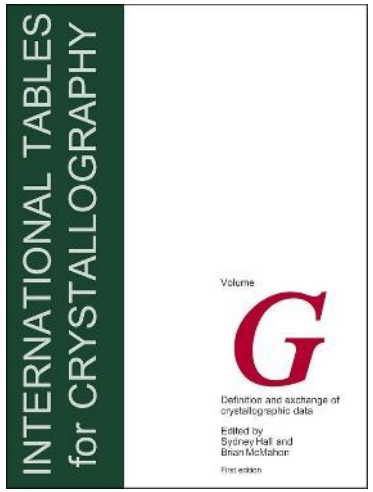
(Received 8 April 1991, accepted 28 June 1991)

Abstract

The specification of a new standard Crystallographic Information File (CIF) is described. Its development is based on the Self-Defining Text Archive and Retrieval (STAR) procedure [Hall (1991). *J. Chem. Inf. Comput. Sci.* **31**, 326–333]. The CIF is a general, flexible and easily extensible free-format archive file; it is human and machine readable and can be edited by a simple text editor. The CIF is designed for the electronic

Introduction

There is an increasing need in many branches of science for a uniform but flexible method of archiving and exchanging data in electronic form. Rapid advances in computer technology, coupled with the expansion of local, national and international networks, have fuelled the need for such a facility. The variety and relative inflexibility of existing data exchange formats have inhibited their effective use. This is true even in fields where the basic data requirements are well defined. Problems of data ex-




```
loop_
 _refln_index_h
 _refln_index_k
 _refln_index_l
 _refln_F_squared_calc
 _refln_F_squared_meas
 _refln_F_squared_sigma
 _refln_observed_status
   2   0   0      25185.30      25637.90        369.48 o
   4   0   0       1787.58       1693.17         29.40 o
   6   0   0         59.02         71.69          4.22 o
   8   0   0         27.82         14.56          3.68 o
  10   0   0        483.46        450.47         23.58 o
   1   1   0       7357.92       7392.33         96.02 o
   2   1   0         27.32         30.41          1.75 o
   3   1   0       1335.00       1608.41         31.10 o
   4   1   0       9037.50       8887.28         88.78 o
   5   1   0       4968.33       5067.08         84.21 o
   6   1   0         29.52         35.79          2.56 o
   7   1   0          4.53          6.29          2.14 o
   8   1   0          1.53          3.02          2.39 o
   9   1   0         62.03         58.07          4.39 o
  10
```

Hall, S. R.; Allen, F. H.; Brown, D. *Acta Crystallogr.* **1991**, A*47*, 655–685.

Wide applicability of STAR

CIF is based on the Self-Defining Text Archive and Retrieval (STAR) procedure

- STAR language has the capacity to handle the more complex data representations found in various disciplines such as imaging, quantum chemistry, botany, *etc*.
- Non-CIF STAR applications include
  - NMR imaging field (Biological Magnetic Resonance Data Bank)
  - Molecular Information File (MIF)
  - QCHEM ontology for quantum chemistry
  - Botanical ontologies (Florabase, Western Australian Herbarium), *e.g.* plant naming hierarchies lend themselves to the use of 'definition' methods to connect these hierarchies

326 *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 326–333

The STAR File: A New Format for Electronic Data Transfer and Archiving

SYDNEY R. HALL
Crystallography Centre, University of Western Australia, Nedlands 6009, Australia

Received October 2, 1990

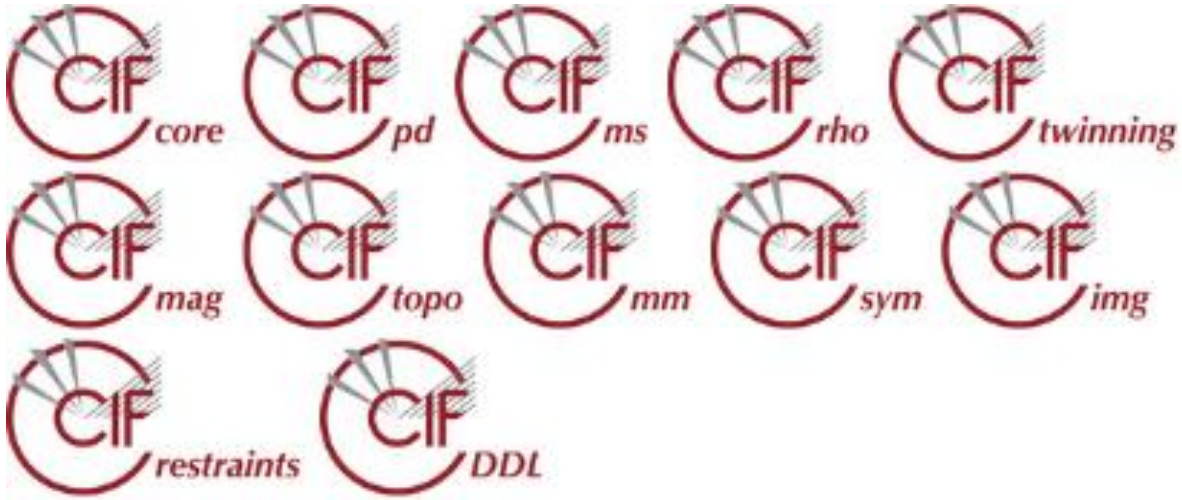
A new type of format is proposed for the computer archiving and electronic transmission of text and numerical data. The Self-defining Text Archive and Retrieval (STAR) File uses standard ASCII text to specify both the data structure and the information. The syntax of this file is simple, and it may be easily interpreted visually or by computer. The STAR format is the basis for the Crystallographic Information File (CIF), which has been adopted by the International Union of Crystallography for the submission of data and text to crystallographic journals and data bases.

Sydney R. Hall

Extensibility of CIF: dictionaries

CIF can host not only structural data

- Formally, the CIF approach makes no distinction between 'data' and 'metadata' and is adaptable and extensible to any domain of structural science
- CIF can host more than just data (symmetry, unit-cell parameters, atomic coordinates, ADPs); it can also host structural, experimental, and processing details, as well as provenance, authorship, journal details, *etc*.
- CIF dictionaries provide a formal taxonomy of crystallographic terms and ideas that can be discipline-specific
  - Core dictionary (coreCIF) *„set of data names designed to cover the requirements of archiving and exchanging raw and processed data and derived structural results"*
  - Powder dictionary (pdCIF)
  - Modulated and composite structures dictionary (msCIF)
  - Electron density dictionary (rhoCIF)
  - Twinning dictionary
  - Magnetic dictionary (magCIF)
  - Topology dictionary (topoCIF)
  - Macromolecular dictionary (mmCIF)
  - Symmetry dictionary (symCIF)
  - Image dictionary (imgCIF)
  - Restraints dictionary



Data flavors: the raw, the cooked and the medium-rare

- Q: Which of the data types – (1) raw, (2) processed or (3) derived data – can be described using the CIF format?
- A: All of them! However, raw data requires imgCIF/CBF ontologies, the particular formats describing diffraction images.

- Q: Do the major structural databases accept raw diffraction data?
- A: Currently not, but most large-scale photon facilities host and make raw diffraction data publicly available after the embargo period. The in-house diffractometer data can be uploaded to free data-sharing platforms (*e.g.* Zenodo).

- Q: And how about the processed data (structure factors)?
- A: From 2011 the Cambridge Structural Database (CSD) and the Inorganic Crystal Structure Database (ICSD) strongly encourage the inclusion of structure factor data along with CIFs, in line with the recommendations by the IUCr. This service is also available in the Crystallography Open Database (COD). Besides, many journals accept CIF files with structure factors as Electronic Supplementary Information.

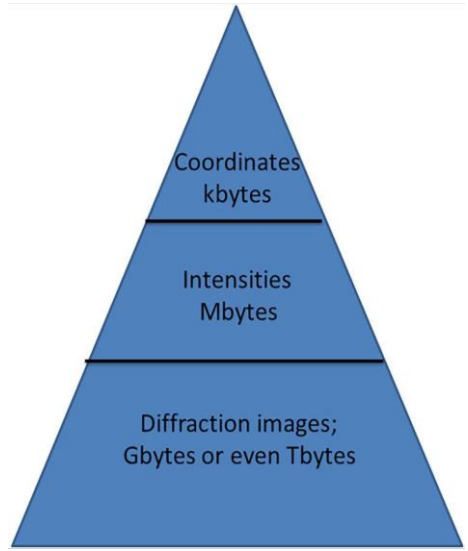
Raw diffraction data reuse: the good, the bad and the challenging
A Satellite Workshop to the XXVI IUCr Congress

Organised by Loes Kroon-Batenburg, Selina Storm, John R. Helliwell and Brian McMahon under the auspices of the IUCr Committee on Data

The raw, the cooked and the medium-rare: unmerged diffraction data as a rich source of opportunities for data re-use and improvements in methods and results

G. Bricogne, C. Flensburg, R. H. Fogh, P. A. Keller, I. J. Tickle and C. Vonrhein

Tuesday 22 August 2023 — International Convention Centre, Melbourne, Victoria, Australia



The Big Data pyramid, after: J. R. Helliwell, *Struct. Dyn.* **2019**, *6*, 054306.

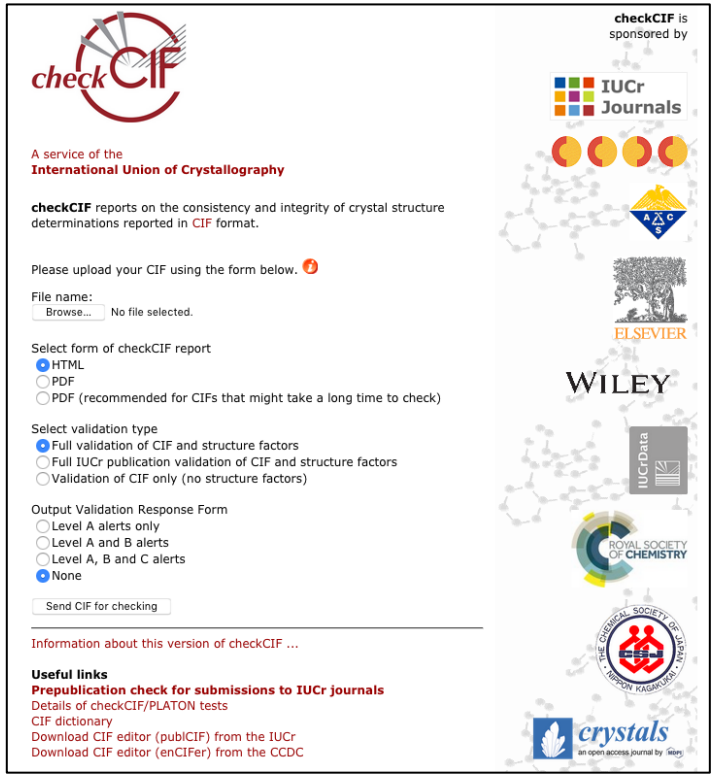
The element of trust: data validation

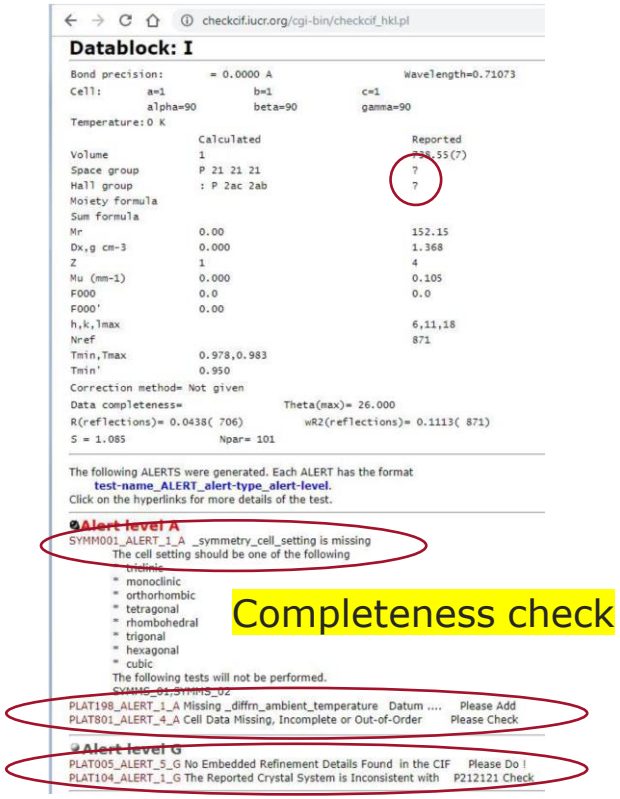
checkCIF is the IUCr service that reports on the completeness, consistency and integrity of CIF files

- Began as IUCr journals in-house suite for validation and consistency checks
- Consolidated with Ton Spek's PLATON (tool for analyzing geometrical calculations, e.g. bond lengths, angles, torsions and various structural tests, *e.g.* for missing symmetry, voids in the lattice) and now offered as public service
- Adopted by many journal publishers and databases (checkCIF reports required!)

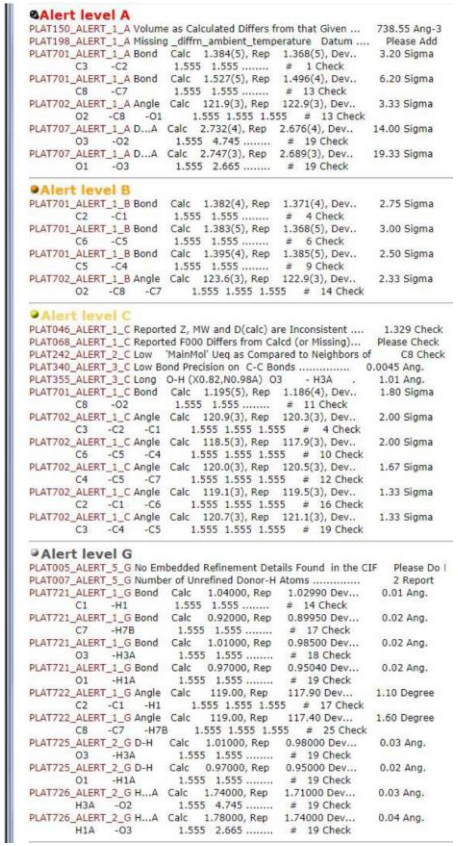
PLATON
A Multipurpose Crystallographic Tool

Anthony L. Spek





Completeness check



A typo in a cell parameter (b = 9.3092 Å instead of 9.0392 Å) generates inconsistencies and anomalies in bond geometry

Consistency check

Three prongs of the *check*CIF approach

FAIR principles do not, in themselves, cover the crucial aspects of intrinsic data quality and FAIR data are not *per se* high quality data! Three C's of *check*CIF*:

- **Completeness**: Including minimal metadata (complete crystal description, details of the diffraction experiment, structure solution and refinement strategy), required for reproducibility
- **Correctness**: Integrity and internal self-consistency (*e.g.* space group must agree with lattice parameters)
- **Context**: Comparison with similar structures in the wider universe of knowledge (*e.g.* tests for missed symmetry, missed twinning, solvent accessible voids, and mis-assigned atom types)

*McMahon, B. *The vital role of Crystallographic Information Files in chemical and biological crystallography to underpin the databases' validation reports*, Data Science Skills in Publishing Workshop, 2019, Vienna.

*check*CIF reports and alerts

The automated *check*CIF report contains three types of alerts:

- ALERT level A = In general: serious problem
- ALERT level B = Potentially serious problem
- ALERT level C = Check and explain

research communications

CRYSTALLOGRAPHIC COMMUNICATIONS
ISSN 2056-9890

checkCIF validation ALERTS: what they mean and how to respond

Anthony L. Spek*

Acta Cryst. (2020). E76, 1–11 https://doi.org/10.1107/S2056989019016244 1

A. L. Spek, *Acta Crystallogr. E* **2020**, *76*, 1–11.

Try to eliminate the problems, but do not change the CIF or alter your refinement just to make *check*CIF alerts vanish! Sometimes alerts cannot be eliminated, but instead have valid reasons for their presence. In these cases validation reply form (vrf) statements can be inserted into a CIF file, which acknowledge and account for checkCIF alerts.

```
# start Validation Reply Form
_vrf_ATOM007_Sb3N5_35_GPa
;
PROBLEM: _atom_site_aniso_label is missing
RESPONSE: Due to incompleteness of the high pressure dataset for a sample
in a diamond anvil cell ADPs of all atoms were refined
in isotropic approximation.
;
_vrf_PLAT029_Sb3N5_35_GPa
;
PROBLEM: _diffrn_measured_fraction_theta_full value Low .       0.462 Why?
RESPONSE: This is a high pressure data set, a substantial part of
a reciprocal space was shaded by the diamond anvil cell.
;
_vrf_PLAT911_Sb3N5_35_GPa
;
PROBLEM: Missing FCF Refl Between Thmin & STh/L=    0.600          80
Report
RESPONSE: Certain part of the reflections is missing due to shading
by the diamond anvil cell.
;
# end Validation Reply Form
```

The following ALERTS were generated. Each ALERT has the format
test-name_ALERT_alert-type_alert-level.
Click on the hyperlinks for more details of the test.

Alert level A

ATOM007_ALERT_1_A _atom_site_aniso_label is missing
Unique label identifying the atom site.

Author Response: Due to incompleteness of the high pressure dataset for a sample in a diamond anvil cell ADPs of all atoms were refined in isotropic approximation.

PLAT029_ALERT_3_A _diffrn_measured_fraction_theta_full value Low . 0.462 Why?

Author Response: This is a high pressure data set, a substantial part of a reciprocal space was shaded by the diamond anvil cell.

Alert level B

PLAT911_ALERT_3_B Missing FCF Refl Between Thmin & STh/L= 0.600 80 Report

Author Response: Certain part of the reflections is missing due to shading by the diamond anvil cell.

CIF and *check*CIF: Editors' requirements

ROYAL SOCIETY OF CHEMISTRY

Membership ▼ Publishing ▼ Policy and campaigning ▼ Standards and recognition ▼ Fundi

Home > Publishing > Publish with us > Publish a journal article > Data sharing

Data sharing

Guidance for best practice and reproducibility of experimental data.

Recommended repositories

A data repository is an external storage space for researchers to deposit datasets associated with their research. Data should be submitted to a discipline-specific, community-recognised repository where possible, or alternatively to an institutional repository, or a generalist repository if no subject discipline repository is available for the given data type.

The choice of repository is the author's decision, provided it is in line with institutional or funder guidelines. The exception to this is small molecule crystal data, which must be deposited with the Cambridge Crystallographic Data Centre (CCDC).

Data availability statements

- Crystallographic data for [compound number] has been deposited at the [name of repository, such as CCDC / ICSD / PBD] under [accession number] and can be obtained from [URL of data record, format https://doi.org/DOI].

CIF and *check*CIF: Editors' requirements

X-Ray crystallography

These guidelines provide details for the presentation of single crystal and powder diffraction data; they apply to submissions to any of our journals.

Small molecule single crystal data

Authors should present their crystal data in a CIF (Crystallographic Information File) format and deposit this with the Cambridge Crystallographic Data Centre (CCDC) before submission. Data will be held in the CCDC's confidential archive until publication of the article, but it will be made accessible to reviewers and the publisher assigned to review the data.

At the point of publication, any deposited data will be made publicly available through the CCDC Access Structures service. In addition, organic and metal-organic structures will be curated into the Cambridge Structural Database, and inorganic structures will be curated into the Inorganic Crystal Structures Database (FIZ Karlsruhe).

Upon deposition, each data set is assigned a Digital Object Identifier (DOI), so that the crystal structure is unambiguously identified and registered.

Include CCDC or ICSD numbers in the manuscript prior to submission as part of a Data Availability Statement. During submission authors will be asked to cite CCDC or ICSD reference numbers; CIFs should not be submitted with the manuscript. Any revised CIFs should be deposited directly with the CCDC before the revised manuscript is submitted to us.

CheckCIF

In addition, authors are required to provide a checkCIF report for their reported crystal data. The checkCIF report can be obtained via the International Union of Crystallography's (IUCr) free checkCIF service, or as part of the CCDC deposition process. Any 'level A' alerts in the report should be explained in the submission details for the article or an explanation provided within the submitted CIF. Authors should submit the checkCIF reports to the Royal Society of Chemistry along with the manuscript files.

If the editor deems it necessary during the peer-review process, the crystallography associated with the manuscript may undergo specialist crystallographic assessment, in which case a report will be provided along with the other reports from reviewers. Any points raised in this assessment should be attended to and all revised CIFs should be deposited with the CCDC prior to uploading the revised manuscript.

For recommended information to include in your CIF, please see the CCDC CIF deposition guidelines. If SQUEEZE or MASK procedures are used, this should be noted in the CIF file.

We encourage authors to include hkl data in the deposited CIF file. Alternatively authors can submit hkl data and the structure files (.fcf) separately during deposition with the CCDC. Raw data accompanying a structure should be made available by the authors for the review process, on request.

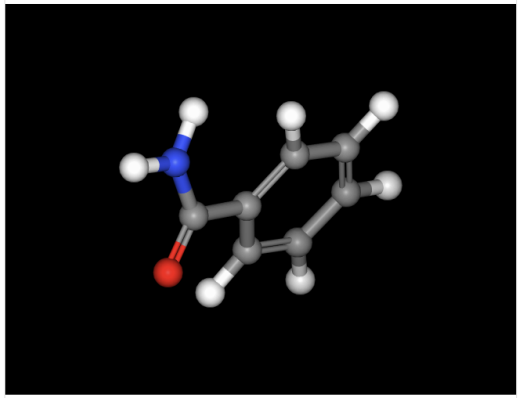
Raw data: cross-referencing between CCDC and external databases (*e.g.* Zenodo)

DOI: 10.5517/ccdc.csd.cc2jmntn

LOHZER : New Structure undergoing enhancement
Space Group: P $2_1/c$ (14), **Cell:** *a* 5.56(4)Å *b* 5.04(3)Å *c* 21.72(4)Å, *α* 90° *β* 90.100(11)° *γ* 90°

3D viewer

Ball and Stick | No Labels



No Packing | H | DISORDER

Chemical diagram

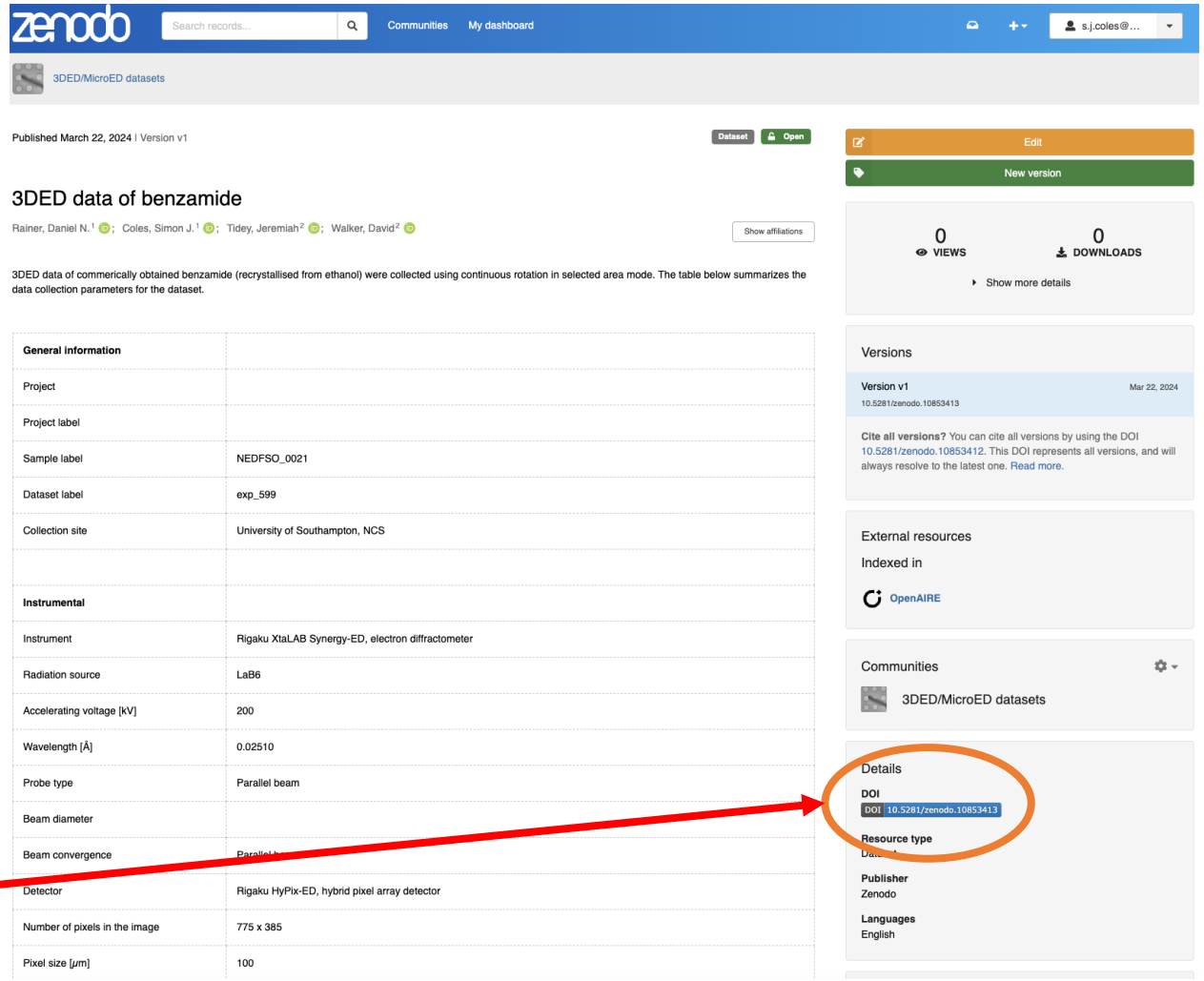
View group symbols key

Additional details

Deposition Number	2342602
Data Citation	Daniel N. Rainer, Simon J. Coles, Mark E. Light, Jeremiah P. Tidey, David Walker CCDC 2342602: Experimental Crystal Structure Determination, 2024, DOI: 10.5517/ccdc.csd.cc2jmntn
Deposited on	22/03/2024

Raw data DOI(s)

DOI	10.5281/zenodo.10853412



zenodo | Search records... | Communities | My dashboard | s.j.coles@...

3DED/MicroED datasets

Published March 22, 2024 | Version v1 — Dataset — Open

3DED data of benzamide

Rainer, Daniel N.[1]; Coles, Simon J.[1]; Tidey, Jeremiah[2]; Walker, David[2] — Show affiliations

3DED data of commerically obtained benzamide (recrystallised from ethanol) were collected using continuous rotation in selected area mode. The table below summarizes the data collection parameters for the dataset.

General information	
Project	
Project label	
Sample label	NEDFSO_0021
Dataset label	exp_599
Collection site	University of Southampton, NCS
Instrumental	
Instrument	Rigaku XtaLAB Synergy-ED, electron diffractometer
Radiation source	LaB6
Accelerating voltage [kV]	200
Wavelength [Å]	0.02510
Probe type	Parallel beam
Beam diameter	
Beam convergence	Parallel
Detector	Rigaku HyPix-ED, hybrid pixel array detector
Number of pixels in the image	775 x 385
Pixel size [μm]	100

Edit | New version

0 VIEWS | 0 DOWNLOADS

Show more details

Versions

Version v1 — Mar 22, 2024 — 10.5281/zenodo.10853413

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.10853412. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

Indexed in

OpenAIRE

Communities

3DED/MicroED datasets

Details

DOI

DOI 10.5281/zenodo.10853413

Resource type

Publisher
Zenodo

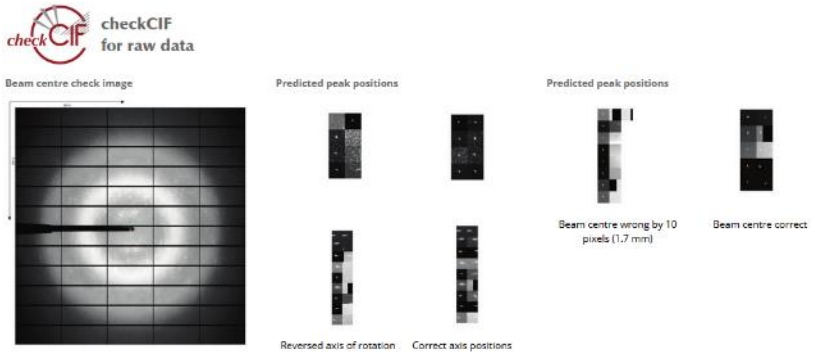
Languages
English

Courtesy of Simon Coles

Continuing FAIRification of raw data within CIF

A brief history of raw data deposition

- Launch of *IUCrData* for peer-reviewed short structure reports (2016)
- A Gold Standard for metadata to describe an MX raw data (2020)
- Raw Data Letters, a new section of *IUCrData* for authors to describe unprocessed diffraction images (2022)
- *check*CIF for raw data (2022)




research papers

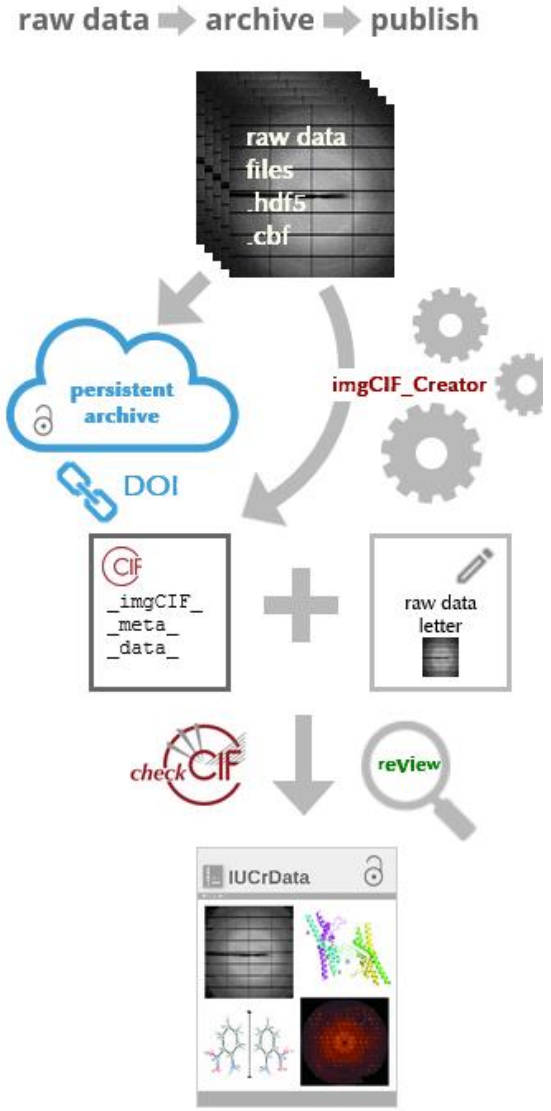
IUCrJ
ISSN 2052-2525
BIOLOGY | MEDICINE

Received 6 February 2020
Accepted 26 June 2020

Gold Standard for macromolecular crystallography diffraction data

Herbert J. Bernstein,[a]* Andreas Förster,[b] Asmit Bhowmick,[c] Aaron S. Brewster,[c] Sandor Brockhauser,[d,e,f] Luca Gelisio,[g] David R. Hall,[h] Filip Leonarski,[i] Valerio Mariani,[g] Gianluca Santoni,[j] Clemens Vonrhein[k] and Graeme Winter[h]

784 https://doi.org/10.1107/S2052252520008672 IUCrJ (2020). 7, 784–792



EDITOR

Loes Kroon-Batenburg
Bijvoet Centre for Biomolecular Research, Utrecht University, The Netherlands
chemical crystallography, macromolecular crystallography, raw data archiving

CO-EDITORS

Miguel Aranda
University of Malaga, Spain
Rietveld method, powder diffraction

Elena Boldyreva
Novosibirsk State University, Russia
high pressure

Aaron Brewster
Lawrence Berkeley National Lab, USA
serial crystallography computational methods development

Simon Coles
University of Southampton, UK
chemical crystallography, structural chemistry, databases

John R. Helliwell
University of Manchester, UK
macromolecular crystallography, data science

New challenge: Crystal Structure Prediction standards and CIF dictionary

Number of theoretically calculated crystal structures far exceeds that of experimentally determined structures!

- In one year a draft of the final dictionary will be communicated to the IUCr and the finalized version will be officially published
- In the meantime feedback, input and approval from community is required
- Available at: github.com/COMCIFS/Structure_Prediction_Dictionary

CSP Data Standards

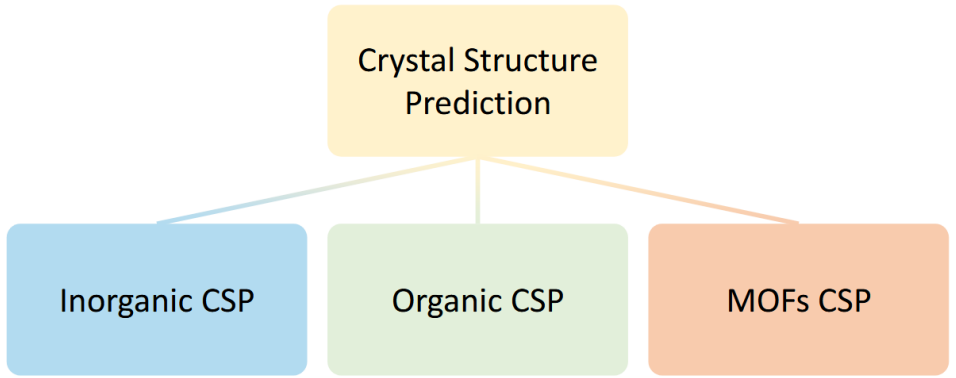
Scope of the project:

- An inclusive initiative that establishes **community standards** for predicted crystal structures
- Enables effective publication of results and easy **search and retrieval** across resources
- Supports levels of **reproducibility** deemed appropriate by different communities
- **Flexible structure** for future expansion as CSP methods are continuing to evolve

IUCr International Union of Crystallography

CCDC

TCOD




James Hester


Nicholas Francia


Ian Bruno

Potential of CIF to host non-crystallographic (spectroscopic) data

research papers

JOURNAL OF APPLIED CRYSTALLOGRAPHY
ISSN 1600-5767

Raman Open Database: first interconnected Raman–X-ray diffraction open-access resource for material identification

Yassine El Mendili,[a]* Antanas Vaitkus,[b] Andrius Merkys,[b] Saulius Gražulis,[b] Daniel Chateigner,[a] Fabrice Mathevet,[a] Stéphanie Gascoin,[a] Sebastien Petit,[a] Jean-François Bardeau,[c] Marco Zanatta,[d] Maria Secchi,[e] Gino Mariotto,[d] Arun Kumar,[d] Michele Cassetta,[d] Luca Lutterotti,[e] Evgeny Borovin,[e] Beate Orberger,[f] Patrick Simon,[g] Bernard Hehlen[h] and Monique Le Guen[i]

Received 23 November 2018
Accepted 28 March 2019

Edited by A. Borbély, Ecole National Supérieure des Mines, Saint-Etienne, France

Keywords: Raman spectroscopy; open databases; combined Raman–X-ray diffraction; DDLm dictionary; CIF2.

[a]Normandie Université, CRISMAT-ENSICAEN, UMR6508 CNRS, Université de Caen Normandie, 6 Boulevard Maréchal Juin, 14050 Caen, France, [b]Vilnius University, Institute of Biotechnology, Sauletekio av. 7, LT-10257 Vilnius, Lithuania, [c]Institut des Molécules et Matériaux du Mans, UMR6283 CNRS, Le Mans Université, Avenue Olivier Messiaen, 72085 Le Mans, France, [d]Department of Computer Science, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy, [e]Department of Industrial Engineering, University of Trento, via Sommarive 9, 38123 Trento, Italy, [f]GEOPS-Paris Sud, Université Paris-Saclay, UMR8148 (CNRS-UPS), Bâtiment 504, 91405 Orsay, France, [g]CEMHTI, UPR CNRS 3079, Université d'Orléans, 1D Avenue de la Recherche Scientifique, 45071 Orléans Cedex 2, France, [h]Laboratoire Charles Coulomb, UMR5521 CNRS, Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France, and [i]ERAMET, 1 Avenue Albert Einstein, 78190 Trappes, France. *Correspondence e-mail: yassine.elmendili@ensicaen.fr

Detailed crystallographic information provided by X-ray diffraction (XRD) is complementary to molecular information provided by Raman spectroscopy. Accordingly, the combined use of these techniques allows the identification of an unknown compound without ambiguity. However, a full combination of Raman and XRD results requires an appropriate and reliable reference database with complete information. This is already available for XRD. The main objective of this paper is to introduce and describe the recently developed Raman Open Database (ROD, http://solsa.crystallography.net/rod). It comprises a collection of high-quality uncorrected Raman spectra. The novelty of this database is its interconnectedness with other open databases like the Crystallography Open Database (http://www.crystallography.net/cod and Theoretical Crystallography Open Database (http://www.crystallography.net/tcod/). The syntax adopted to format entries in the ROD is based on the worldwide recognized and used CIF format, which offers a simple way for data exchange, writing and description. ROD also uses JCAMP-DX files as an alternative format for submitted spectra. JCAMP-DX files are compatible to varying degrees with most commercial Raman software and can be read and edited using standard text editors.

q2xafs workshop

Journal of Synchrotron Radiation
ISSN 0909-0495

Received 28 March 2012
Accepted 26 August 2012

Towards data format standardization for X-ray absorption spectroscopy

B. Ravel,[a]* J. R. Hester,[b] V. A. Solé[c] and M. Newville[d]

[a]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA, [b]Bragg Institute, ANSTO, Locked Bag 2001, Kirrawee DC, NSW 2232, Australia, [c]European Synchrotron Radiation Facility, 6 rue Jules Horowitz, BP 220, 38043 Grenoble Cedex 9, France, and [d]Center for Advanced Radiation Studies, University of Chicago, Building 434A, Argonne National Laboratory, Argonne, IL 60439, USA. E-mail: bravel@bnl.gov

A working group on data format standardization for X-ray absorption spectroscopy (XAS) has recently formed under the auspices of the International X-ray Absorption Society and the XAFS Commission of the International Union of Crystallography. This group of beamline scientists and XAS practitioners has been tasked to propose data format standards to meet the needs of the world-wide XAS community. In this report, concepts for addressing three XAS data storage needs are presented: a single spectrum interchange format, a hierarchical format for multispectral X-ray experiment, and a relational database format for XAS data libraries.

Keywords: XAFS; standardization; data formats.

Synergies between IUCr and other scientific standard organizations

CIF integrates standards developed within IUPAC

- Wavelength description uses the IUPAC nomenclature
- Enantioexcess is as per IUPAC recommendation
- InChi and InChIKey can be embedded in CIF
- Systematic chemical name and chemical formula follow IUPAC standards
- Topological CIF dictionary includes recommendations of IUPAC on nomenclature for coordination polymers

Liaisons between CIF and CDIF

- Physical Sciences Data Infrastructure (PDSI) at Southhampton managed by Simon Coles strives to introduce CDIF (the CODATA Cross Domain Interoperability Framework) that integrates a wide range of resources across chemistry (with crystallography and CIF at the core)

Coalition for the Sustainability of Digital Data Standards in the Chemical Sciences

IUCr (representing the CIF standard) actively participates in the DigSustain meetings from the beginning

DigSustain1: Sustainable Business Modeling for Digital Standards Development, Cambridge, UK, March 2024

DigSustain2: Digital Data Standards Sustainability in Chemical Sciences, Delitzsch, Germany, April 2025

DigSustain3: Coalition for the Sustainability of Digital Data Standards in the Chemical Sciences Planning Meeting, London, UK, November 2025

Stay tuned for more details in the talk of Wendy Patterson!

Where do we go from here?

Impact of CIF can be assessed by its popularity

- '*The single most important impact of CIF has been the ability to easily* ***access data globally across journals, databases and laboratories***'*
- CIF is now one of Universal Data Languages
- Peter Murray-Rust called CIF '***the best working data system in the scientific world***' (2014)
- 'Technology lock-in': foundational technology choice can become so embedded that it stifles future alternatives

Progress and adoption can be measured only indirectly

- Efficiency to submission–deposition processes
- Better understanding of particular data items across laboratories and databases
- Adoption in programming libraries and end-user software
- '*Anyone can place their own tag–value items into a CIF, but until their definitions are accepted into the global ontology they can only be for local use*'*
- Official global ontology is the responsibility of the IUCr COMCIFS group, who moderates and maintain the standard
- Important to keep momentum and train new generations
- **The art of understanding and being understood**

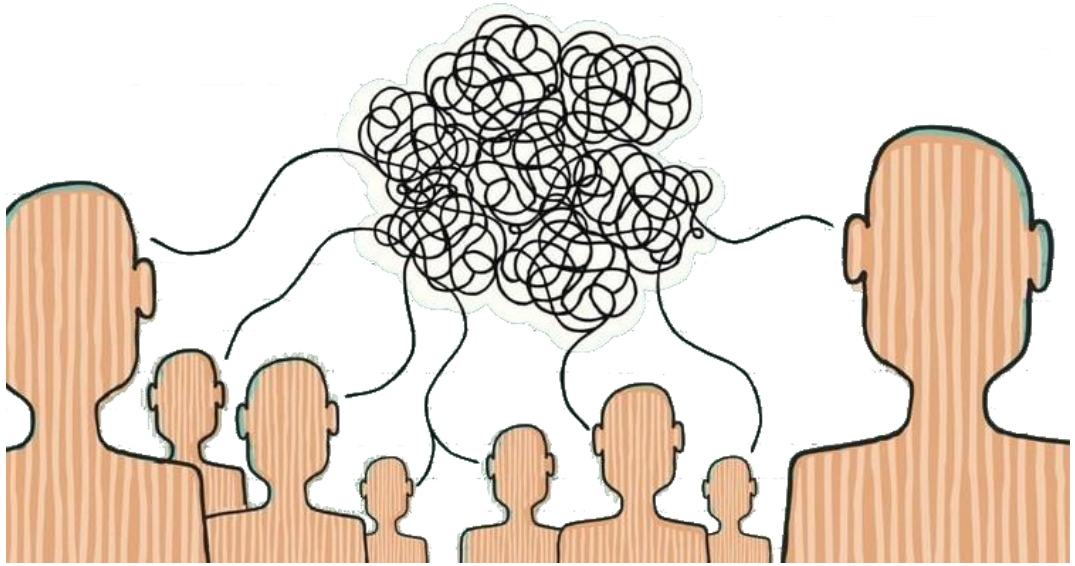


Image credit: iStock

*S. R. Hall, B. McMahon, *Data Science Journal* **2016**, *15*, 1–15.

Thank you for listening!

Acknowledgments

Simon Coles

James Hester

John R. Helliwell

Committee on Data (CommDat)

Membership

- S. Coles (Chair, UK)
- H.J. Bernstein (USA)
- A. Brink (South Africa)
- I. Bruno (UK)
- S. Coles (UK)
- K. Dziubek (Austria)
- A. Goetz (France)
- S. Kabekkodu (USA)
- L.M.J. Kroon-Batenburg (The Netherlands)
- G. Kurisu (Japan)
- W. Minor (USA)
- S. Storm (Germany)
- L. Van Meervelt (Belgium)
- J. Hester (Australia) (COMCIFS liaison)

Consultants

- S. Androulakis (Australia)
- M.P. Blakeley (France)
- G. Bricogne (UK)
- S. Grazulis (Lithuania)
- J.R. Helliwell (UK)
- B. Matthews (UK)
- A. Sarjeant (USA)
- D. Szebenyi (USA)
- E.F. Weckert (Germany)
- J. Trewhella (Australia)