



# *Quality structural data in modern scientific culture*

***John R. Helliwell***

*School of Chemistry, University of Manchester,  
Manchester M13 9PL, England*

*John.helliwell@manchester.ac.uk*



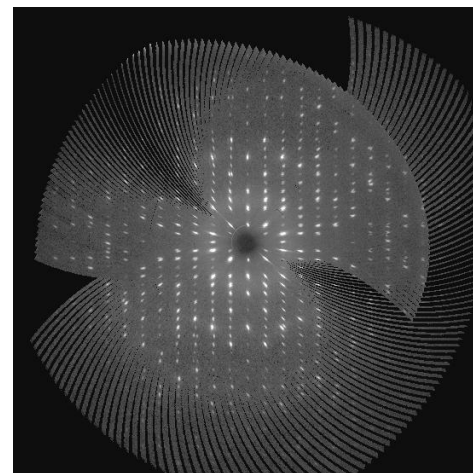
# IUCr activities relating to data

- 1991 Crystallographic Information Framework
- 1998 *checkCIF*
- 2011 Raw diffraction data deposition (the DDDWG)
- 2017 Committee on Data (CommDat)
  - *Terms of reference: CommDat will advise the IUCr Executive Committee on all aspects of data with respect to policy and actions to be taken*

# Our data zoo

Data can mean any or all of:

1. raw measurements from an experiment
2. processed numerical observations
3. derived structural information



```
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 0 772.37 856.47 28.20 o
4 0 0 1445.15 1446.80 39.55 o
6 0 0 1130.79 1097.08 30.62 o
8 0 0 1347.13 1490.27 55.41 o
10 0 0 3273.01 3545.64 154.91 o
12 0 0 48.20 40.50 4.56 o
14 0 0 79.87 63.02 7.91 o
2 1 0 2093.70 1975.83 47.36 o
3 1 0 33795.10 34884.29 1287.71 o
4 1 0 2298.16 2035.72 38.24 o
5 1 0 9.73 36.06 5.59 o
6 1 0 449.80 506.89 11.92 o
7 1 0 1.81 7.91 5.59 o
8 1 0 43.36 28.81 6.79 o
9 1 0 64.18 48.51 6.02 o
10 1 0 1412.22 1628.54 45.96 o
11 1 0 242.68 279.96 9.70 o
12 1 0 14.96 10.52 3.84 o
13 1 0 16.87 15.76 4.56 o
14 1 0 16.46 7.91 7.91 o
15 1 0 0.00 3.95 5.59 o
0 2 0 2443.71 2679.14 61.27 o
1 2 0 23397.80 23770.90 546.30 o
2 2 0 20572.37 19502.51 520.01 o
3 2 0 8854.88 8282.53 169.57 o
```

Table 1. Selected geometric parameters ( $\text{\AA}$ ,  $^\circ$ )

Fe1—C9	2.030 (4)	Fe1—C7	2.049 (4)
Fe1—C5	2.036 (3)	Fe1—C11	2.053 (4)
Fe1—C12	2.038 (4)	S1—C1	1.693 (3)
Fe1—C13	2.038 (4)	N1—C1	1.315 (4)
Fe1—C4	2.038 (3)	N2—C1	1.345 (4)
Fe1—C8	2.041 (4)	N2—N3	1.387 (4)
Fe1—C10	2.042 (4)	N3—C2	1.290 (4)
Fe1—C6	2.048 (4)		
C1—N2—N3	118.2 (3)	N2—C1—S1	120.1 (2)
C2—N3—N2	116.9 (3)	N3—C2—C4	115.6 (3)
N1—C1—N2	117.0 (3)	N3—C2—C3	125.2 (3)
N1—C1—S1	122.8 (2)	C4—C2—C3	119.3 (3)

Each includes **metadata** ie also *are* data!

*Data*



Subjective



Objective

The article 'narrative'  
(Very subjective)

Coordinates  
kbytes

Intensities  
Mbytes

Diffraction images;  
Gbytes or even Tbytes

# The element of trust

## Trust in

- Data transmission/exchange
  - Crystallographic Information File (1991)
- Data consistency
  - checkCIF for derived (coordinate) data (1998)
  - checkCIF including structure factors (2007)
- Data provenance
  - Diffraction data deposition (2011-2017)
- The science is in the data

# Association of Learned and Professional Society Publishers Award 2006: IUCr

## IUCr WINS PRESTIGIOUS INTERNATIONAL AWARD FOR PUBLISHING INNOVATION



The IUCr has won the 2006 Award for Publishing Innovation of the Association of Learned and Professional Society Publishers (ALPSP).

The Award, for Data Exchange, Quality Assurance and Integrated Data Publication (CIF and *checkCIF*), recognizes the involvement of the IUCr in the development of the Crystallographic Information Framework and its applications, for example

- standard data definitions for crystallographic information archive and interchange
- submission format for structure report articles in crystallographic journals
- standard format for depositing supplementary structural data accompanying publications
- automated checking of the integrity and self-consistency of crystal structure models (the web *checkCIF* service)
- use of *checkCIF* as a peer review tool
- dissemination of crystal structural models in online publications and automated visualization.

The judges felt that in developing CIF and *checkCIF*, the IUCr has established an important example of data quality assurance with potential applications in other scientific, medical and indeed social sciences publishing.

The crystallographic information file (CIF) and associated data dictionaries allow the seamless transfer of information from experimental apparatus, through computation analysis, to database deposition and publication. CIF also allows the definition of quality standards for data deposition and publication and the



Brian McMahon (left), IUCr R&D Officer, receives the award from Bernard Donovan, ALPSP President, at the Association's annual dinner on 14 September 2006 at the Armourers' Hall in London, UK.



# CODATA Prize 2014: Sydney R. Hall, IUCr



## Research Papers

### The Implementation and Evolution of STAR/CIF Ontologies: Interoperability and Preservation of Structured Data

**Authors:** Sydney R. Hall , Brian McMahon

## Abstract

The global application of the Crystallographic Information Framework (CIF) to the molecular structure domain has been successful over the past 20 years. It is used widely by molecular science journals and databases for submission, validation and deposition. This paper will give an overview of the CIF implementation, highlighting its particular successes and occasional failures. It will also recommend criteria for the application of an ontology-based data management system to any information domain. The paper will conclude with some details of the latest STAR data definition language and the importance of methods to the preservation of derivative data items.

**Keywords:** [crystallography](#), [crystallographic information file](#), [data exchange standard](#), [machine readable dictionaries](#), [ontologies](#)

Hall, S.R. & McMahon, B. (2016). The Implementation and Evolution of STAR/CIF Ontologies: Interoperability and Preservation of Structured Data. *Data Science Journal*. **15**, p.3.

# Leading the way

“The requirement from academic journals that authors provide data in support to their papers has proven to be potentially culture-changing, as has been the case in crystallography.”

“Many data standards are maintained by international scientific unions (*e.g.* the International Union of Crystallography) ....”



The European Union Report 2018



# DDDWG (2011-2017)

## Membership

- John R Helliwell and Brian McMahon (UK), Chair and Co-Chair;
- Steve Androulakis (Australia)
- Sol Gruner (USA)/D. Marian Szebenyi (USA)
- Loes Kroon-Batenburg (Netherlands)
- Tom Terwilliger (USA)
- John Westbrook (USA)
- Heinz-Josef Weyer (Switzerland) †
- Edgar Weckert (Germany)

# CommDat Membership

- J.R. Helliwell (Chair, UK)
- B. McMahon (Secretary, UK)
- H.J. Bernstein (USA)
- S. Coles (UK)
- K. Dziubek (Poland)
- A. Goetz (France)
- S. Kabekkodu (USA)
- Loes Kroon-Batenburg (Netherlands)
- W. Minor (USA)
- A. Sarjeant (USA)
- J. Westbrook (USA)
- L. Van Meervelt (Belgium)
- J. Hester (Australia) (COMCIFS liaison)

## ***Consultants***

*S. Androulakis (Australia), M.P. Blakeley (France), G. Bricogne (UK), S. Grazulis (Lithuania), B. Matthews (UK), D. Szebenyi (USA), E.F. Weckert (Germany), J. Trewhella (Australia)*

## *IUCr and CommDat's take-home message*

The IUCr (representing the community of crystallographers and structural scientists) maintains the need for the highest quality of data management at all stages, from experimental data collection, through reduction and analysis, to publication and database deposition