

# *Improved access to raw diffraction data and its impact on crystallographic education and teaching*

*John R. Helliwell<sup>1</sup> and Brian McMahon<sup>2</sup>*



- 1. School of Chemistry, University of Manchester, Manchester, M13 9PL, UK*
- 2. IUCr 5 Abbey Square, Chester CH1 2HU, UK*

IUCr Diffraction Data Deposition Working Group



# *Overall vision*

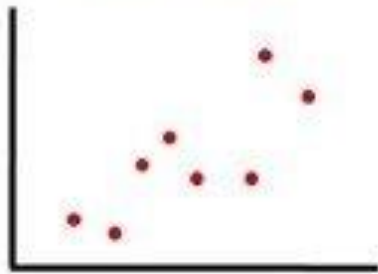
- *Raw diffraction images are expected to be useful for improving numerous research applications that crystallographers engage with and which therefore impacts on education....*

*Nb the **philosophical view** of the  
importance of access to raw  
diffraction data; namely analysis  
through one's own eyes not the lens  
of someone else*

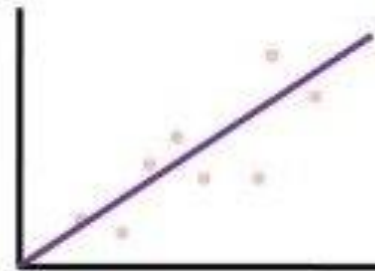
# *Data seen through the lens of others!*

## PhD tips # 22

Actual data



How you saw it:  
Perfectly linear, as expected



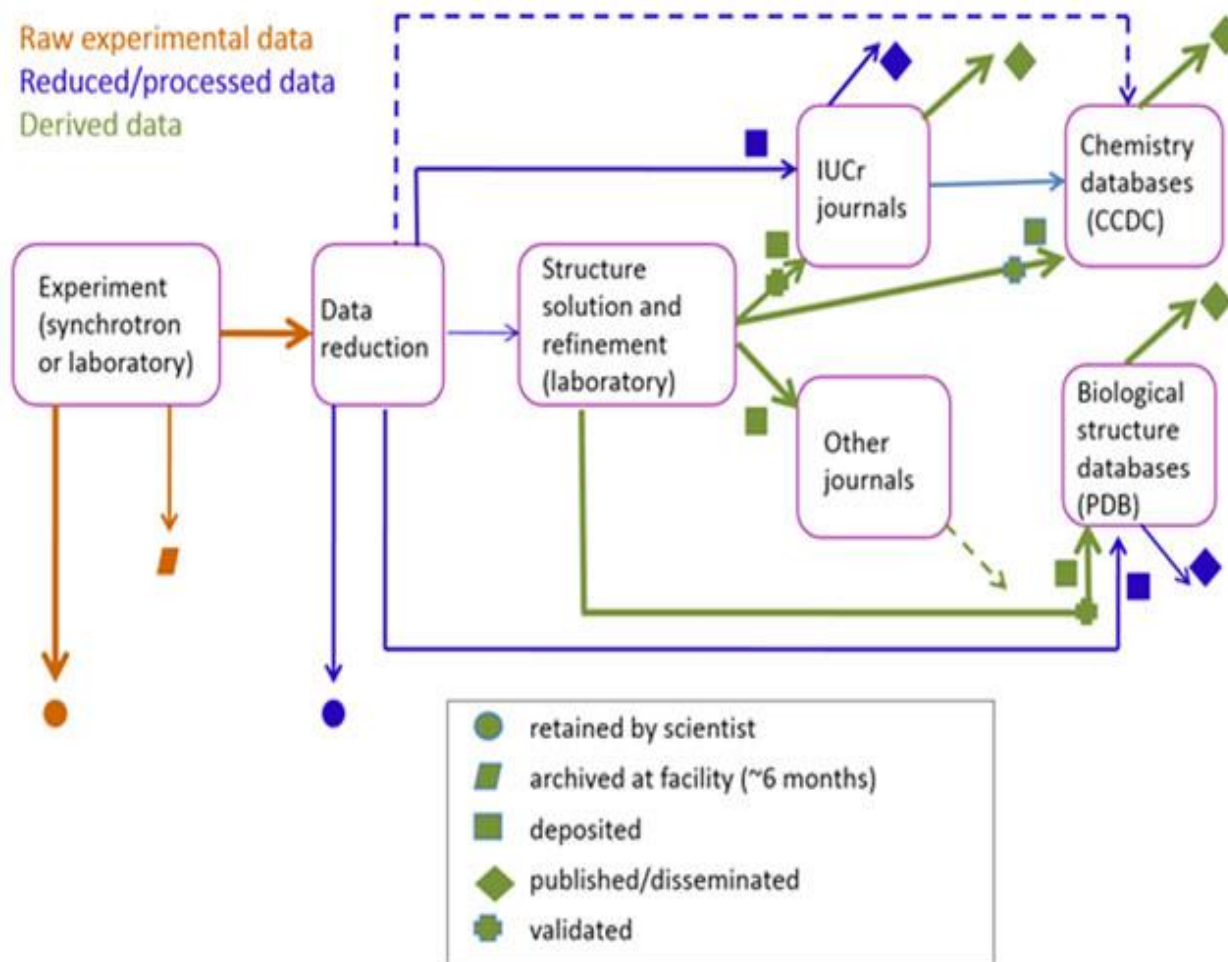
How your supervisor saw it:  
Perfectly exponential, amazing new stuff



How the referee saw it:  
Pure noise



# *A coherent information flow in crystallography. CIF ontologies characterize data at every stage of the information processing life cycle*



*”Ideally, the full scientific record should provide access to the raw data.....the IUCr is beginning to consider longer-term approaches to archiving the raw data”*

*P. Strickland, B. McMahon and J. R. Helliwell, Learned Publishing 21 (2008) 63.*

# *The consequences are*

*Researchers will learn the protocols associated with archived, open, raw diffraction data, to accompany their processed diffraction data and derived coordinates that are expected to be found in the curated databases.*

*Referees of submitted articles can assess the words, the coordinates, the processed diffraction data and the finally the raw data if they deem it to be necessary.*

*Funding agencies are looking at an Open Science protocol to improve speed of discovery for societal challenges, from the start of a funded project. Eg the EU OPEN\_AIRE Open Science Cloud,*

*So, routine access to raw data needs to impact crystallographic teaching such as for our important European Crystallography School. Crystallographers in general likewise need to know:-*

- Their funding agency and/or employer's needs for a Research Data Management Plan;*
- Their option(s) for storing their raw data;*
- The metadata that needs to be provided (general and specific) to allows its re use;*
- When to request access to raw data for optimal refereeing;*
- For special research programs tackling urgent health problems 'Open Science' special research grant terms and conditions may require 'publication of the raw data before analysis';*



## ***Raw diffraction images offer the opportunity of***

- *Analysing data at higher resolution than used in the original work*
- *Serving as benchmarks in developing improved methods of analysis*
- *Checking the interpretation of the symmetries of the crystals*
- *Analysing diffraction from multiple lattices present in the crystals*
- *Guarding against incorrect unit cell parameters leading to missed lattice layers*
- *Analysing the diffuse scattering that reflects correlated motions or disorder of atoms in the crystals*

# *The practicalities of archiving of our raw data*

- *The IUCr's Diffraction Data Deposition Working Group has for over 4 years examined the issues and prospects for linking raw diffraction data sets to publications.*
- *Dealing with our raw data which is definitely 'Big Data' territory!*

# *Members of the DDDWG 2011 to 2017*

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

# Significant pioneering developments

- Australian synchrotron MX raw data archive  
[https://store.synchrotron.org.au/public\\_data/](https://store.synchrotron.org.au/public_data/)
- The USA NIH funded various structural genomics projects with raw data archives
- The Institut Laue Langevin and ISIS are exemplars at preserving all data and with dois  
eg

Science & Technology Facilities Council		ISIS Data Journal
The archive for ISIS research data		
ISIS	ISIS Data	

RB920486

**Investigation title:** Electric field effect on the interfacial uncompensated spins in the Co/BiFeO<sub>3</sub>/STO exchange bias system.

**Release date:** Fri Jul 26 09:06:29 BST 2013

**Creator:** Dr Nina-Juliane Steinke

**DOI:** 10.5286/ISIS.E.24079627

**Date of Experiment:** Fri Jul 23 08:52:43 BST 2010

**Publisher:** STFC ISIS Facility

**Data format:** RAW/Nexus  
Select the data format above to find out more about it.

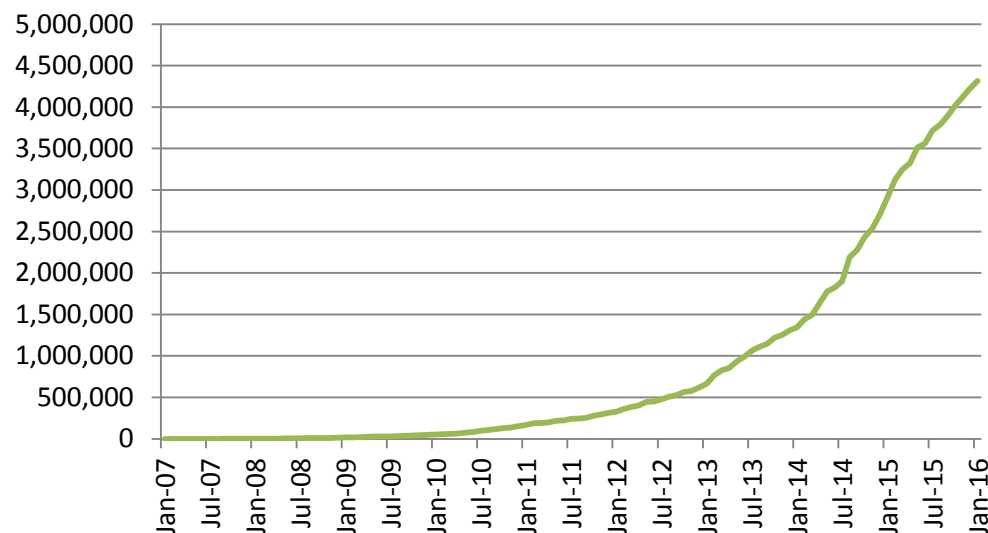
**Data Citation**  
The recommended format for citing this dataset in a research publication is as:  
[author], [date], [title], [publisher], [doi]

 **DOWNLOAD**  
download the dataset


  
Data collected  
CRISP line  
at the ISIS

*A pioneer SR facility retaining all data since start up*

Total Data On Diamond Archive (Gb)



diamond.ac.uk/Users/UserGuide/Data-User-Guide/Accessing-Data/Data-Policy.html#Owns

 diamond

About Us For Users Industry Public Science Beamlines Careers

For Users

Users User Guide Applying for Beamtime Diamond Users Committee FAQ New User Administration

Users / User Guide / Your Data: User Guide / Accessing Your Data / Experimental Data Management Policy

In This Section

Experimental Data Management Policy

*Users conducting Peer Reviewed Research will own the Experimental Data that they produce.*

*“...Following the initial 30 day storage period, Diamond will create a single archive copy of the Experimental Data on tape.”*

*“...Users of Diamond Facilities are responsible for meeting any third-party data management obligations that may be applicable.”*

# *Options and recent developments*

- *University data repositories accept the raw data and organise doi registration;*
- *EC's Zenodo science data archive (free of charge!);*
- *ESRF Data Archive (“every raw data set measured with a registered doi”);*
- *Institut Laue Langevin Data Archive;*
- *National initiatives (eg the RepOD (<https://repod.pon.edu.pl/>) open science repository in Poland).*

# ESRF DATA POLICY SINCE LAST PROPOSAL ROUND

- **ESRF Council officially adopted a Data Policy (1/12/2015)**
- **ESRF is custodian of data and metadata**
- **ESRF to collect high quality metadata to facilitate reuse of data**
- **ESRF will keep raw (or reduced) data for 10 years + metadata for ever**
- **Data will be registered in a data catalogue (icat) + published with a Digital Object Identifier (DOI)**
- **Principal investigators have exclusive access to data during the embargo period (3 years but can be extended)**
- **Data will be made public after the embargo period under CC-BY**
- **Data Policy will be implemented on all beamlines by 2020**

<http://www.esrf.eu/home/UsersAndScience/UserGuide/esrf-data-policy-implementation.html>

# Context

‘FAIR Data’ (Findable, Accessible, Interoperable and Reusable)

[www.nature.com/scientificdata](http://www.nature.com/scientificdata)

## SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Research data

» Publication

characteristics

### Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*<sup>#</sup>

#### Box 2 | The FAIR Guiding Principles

##### To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

##### To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

##### To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

##### To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards



*Two specific examples.....*

*A synchrotron data set and a home  
laboratory data set.....*

[Search resources](#)

[Academic support](#)

[Our services](#)

[The John Rylands Library](#)

[Special Collections](#)

[About Us](#)

[My account](#)

▲ [The University of Manchester Library](#)

▲ [Search resources](#)

Manchester eScholar

- ▶ [Search](#)
- ▶ [Browse by](#)
- ▶ [Help](#)
- ▶ [About](#)
- ▶ [Contact us](#)

**FEEDBACK AND  
ENQUIRIES**



## 4XAN\_Carboplatin\_NaBr\_Diamond\_I04

Tanley, Simon

[Experimental data]. 2014. The University of Manchester.

### Access to files

- [carboplatin\\_NaBr\\_diamond\\_1\\_1-180.ZIP](#) (x-zip-compressed)
- [carboplatin\\_NaBr\\_diamond\\_1\\_181-360.ZIP](#) (x-zip-compressed)

### Abstract

Carboplatin is a second-generation platinum anticancer agent used for the treatment of a variety of cancers. Previous X-ray crystallographic studies of carboplatin binding to histidine (in hen egg-white lysozyme; HEWL) showed the partial conversion of carboplatin to cisplatin owing to the high NaCl concentration used in the crystallization conditions. HEWL co-crystallizations with carboplatin in NaBr conditions have now been carried out to confirm whether carboplatin converts to the bromine form and whether this takes place in a similar way to the partial conversion of carboplatin to cisplatin observed previously in NaCl conditions. Here, it is reported that a partial chemical transformation takes place but to a transplatin form. Thus, to attempt to resolve purely carboplatin binding at histidine, this study utilized co-crystallization of HEWL with carboplatin without NaCl to eliminate the partial chemical conversion of carboplatin. Tetragonal HEWL crystals co-crystallized with carboplatin were successfully obtained in four different conditions, each at a different pH value. The structural results obtained show carboplatin bound to either one or both of the N atoms of His15 of HEWL, and this particular variation was dependent on the concentration of anions in the crystallization mixture and the elapsed time, as well as the pH used. The structural details of the bound carboplatin molecule also differed between them. Overall, the most detailed crystal structure showed the majority of the carboplatin atoms bound to the platinum centre; however, the four-carbon ring structure of the carboplatin dicarbonylate moiety (CDDC) remained elusive.

### Related resources

Full-text held externally  
DOI: [10.15127/1.266906](#)  
<http://scripts.iucr.org/cgi-bin/paper?S2053230X16000777>

University researcher(s)

Academic department(s)

[Faculty of Life Sciences' website](#)

[Search resources](#)

[Academic support](#)

[Our services](#)

[The John Rylands Library](#)

[Special Collections](#)

[About Us](#)

[My account](#)

▲ The University of Manchester Library

▲ Search resources

Manchester eScholar

- ▶ [Search](#)
- ▶ [Browse by](#)
- ▶ [Help](#)
- ▶ [About](#)
- ▶ [Contact us](#)

FEEDBACK AND  
ENQUIRIES



## HEWL\_cisplatin\_5percentDMSO\_RT: 4g4a

Tanley, Simon

[Experimental data] version online. 2012. The University of Manchester.

### Access to files

[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_01\\_0001to01\\_0347.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_02\\_0001to02\\_0200.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_03\\_0001to03\\_0235.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_04\\_0001to04\\_0303.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_05\\_0001to05\\_0314.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_06\\_0001to06\\_0348.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_07\\_0001to07\\_0200.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_07\\_0201to07\\_0377.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_08\\_0001to08\\_0300.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_08\\_0301to08\\_0584.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_09\\_0001to09\\_0310.zip](#) (zip)  
[HEWL\\_cisplatin\\_5percentDMSO\\_RT\\_09\\_0311to09\\_0658.zip](#) (zip)

### Abstract

Abstract: The anticancer complexes cisplatin and carboplatin are known to bind to both the N and the N atoms of His15 of hen egg-white lysozyme (HEWL) in the presence of dimethyl sulfoxide (DMSO). However, neither binds in aqueous media after 4 d of crystallization and crystal growth, suggesting that DMSO facilitates cisplatin/carboplatin binding to the N atoms of His15 by an unknown mechanism. Crystals of HEWL cocrystallized with cisplatin in both aqueous and DMSO media, of HEWL cocrystallized with carboplatin in DMSO medium and of HEWL cocrystallized with cisplatin and N-acetylglucosamine (NAG) in DMSO medium were stored for between seven and 15 months. X-ray diffraction studies of these crystals were carried out on a Bruker APEX II home-source diffractometer at room temperature. Room-temperature X-ray diffraction data

### Related resources

Full-text held externally

DOI: [10.15127/1.215887](#)

DOI: [doi:10.1107/S1744309112042005](#)

DOI: [doi:10.1107/S1744309112042005](#)

University researcher(s)

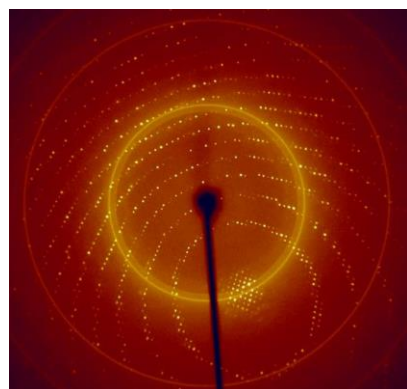
Academic department(s)

Faculty of Engineering and Physical Sciences' website

Faculty of Life Sciences' website

School of Chemistry's website

*Each entry also includes  
a variety of metadata*



# wwPDB Deposition & Annotation System

## Identifying Primary Data



wwPDB Deposition: D\_8000200025 -- Requested ID: PDB

FAQ

Tutorial

Welcome to the Worldwide Protein Data Bank

### Navigation

- ✓ Instructions
- ✓ Communication
- ✓ Re-upload files
- ✓ Upload summary
- Admin
  - ✓ Contact information
  - ✓ Grant information
  - ✓ Release status
  - ✓ Entry title & author
  - ✓ Citation information
- Macromolecules
  - ✓ 1) Interleukin-1 beta
- Data collection
  - ✓ Crystal Information
  - ✓ Collection Source
  - ✓ Software Used
  - ✓ Collection Statistics
- Refinement
  - ✓ Refinement
  - ✓ Ligands
  - ✓ Biological assembly
  - Validation reports
  - Summary & conditions

Log out

Database ID:



Database name:



Details:



### Related external experimental data sets

DOI for the related experimental data set:



doi:10.000/100/da.dat

DOI for additional metadata describing the related data set:



doi:10.000/100/md.cif

The type of experimental data:



diffraction image data



DOI for data set

DOI for descriptive metadata

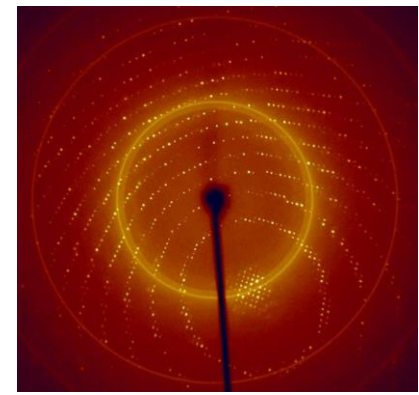
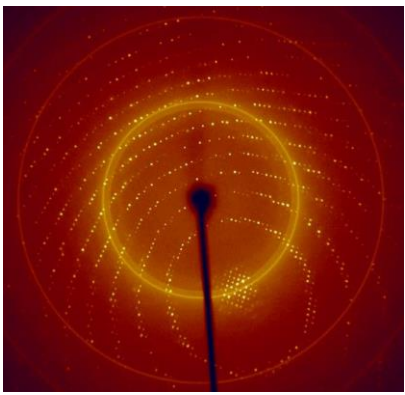
Continue to next section

*I have archived 34 such platin protein raw diffraction data sets at the University of Manchester Library Data Archive making a fully open access complete suite of our raw data along with SFs, coordinates (held at the PDB) and our IUCr Journals publications within this medical related research theme.....*

*Further actions of JRH as a researcher; inviting other platins with proteins and nucleic acids researchers to contribute to this archive.....*

# *In conclusion*

- New modus operandi for *published funded research*; access to all data and should be open access.
- A *limited time for funded researchers to analyse their data and publish*; typically 3 years. Then the raw data should be put on open access excepting exceptional circumstances can sometimes apply.
- Policy makers are now discussing new ways to '*speed up science and discovery for tax payers to reap quicker benefits*'; *Open Science*. New rules of conduct for such 'Open Science' funded research would be essential!
- Next actions of the IUCr DDDWG;
  - articles are in preparation from ECM29 Rovinj Workshop;
  - with IUCr COMCIFS a checkcif for raw data;
  - planned workshop on *metadata for raw data at* ACA New Orleans;
  - Sessions/talks at IUCr Hyderabad.....



*Join in with the raw diffraction data  
revolution.....learn the new protocols for  
all your data*

*Thankyou*

