Overview of the diffraction data deposition activities of the crystallographic community

John R Helliwell and Brian McMahon

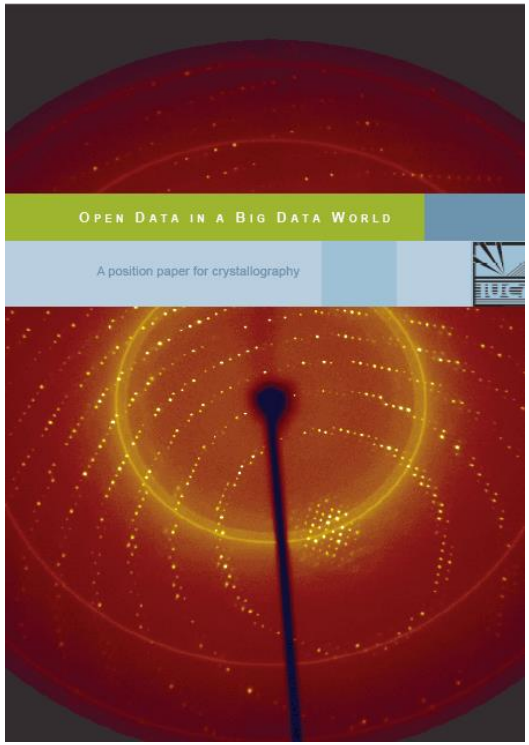
MANCHESTER 1824
The University of Manchester

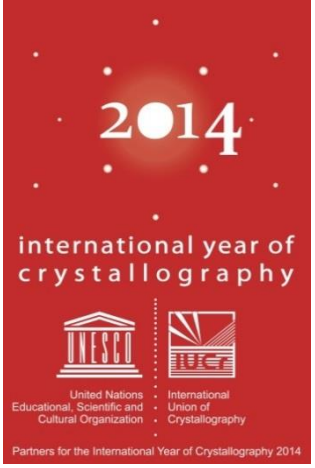
IUCr Diffraction Data Deposition Working Group
IUCr

This Session

- *Includes several major databases encompassing many fields of the IUCr's Commissions*
- *Several other related databases omitted simply through time pressure, e.g.*
  - *Inorganic Crystal Structure database (ICSD)*
  - *Metals Database (CRYSTMET)*
  - *Bilbao Crystallographic Server (symmetry data)*
  - *Incommensurate modulated structures (ICSDB)*
  - *American Mineralogist Crystal Structure database*
- *These opening remarks include a short summary of the IUCr's raw data deposition activities*

Crystallography embraces open data





Let's recall that :-
2014 was proclaimed the International Year of Crystallography

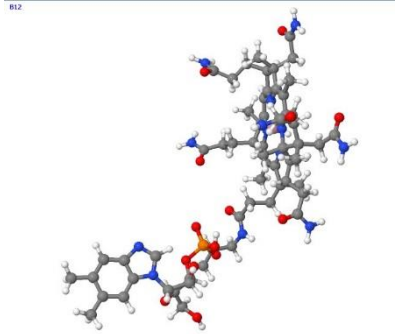
United Nations

"Crystallography has an important place as we work for inclusive sustainable development – policies that are good for people and the planet"

Ban Ki-Moon, UN Secretary-General • IYCr 2014 Opening Ceremony

The Nobel Prize in Chemistry 1964
Dorothy Crowfoot Hodgkin

The Nobel Prize in Chemistry 1964

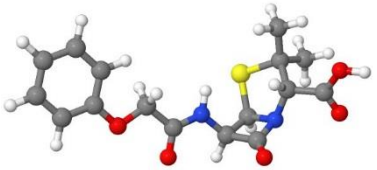


vitamin B12 (cobalamin)

Dorothy Crowfoot Hodgkin

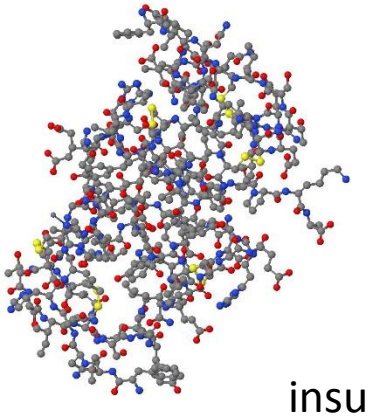
Prize share: 1/1

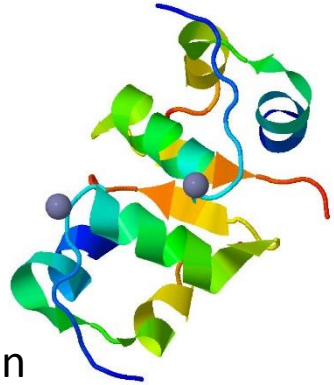
Dorothy Hodgkin (1910-1994): still a major inspiration today



penicillin

The Nobel Prize in Chemistry 1964 was awarded to Dorothy Crowfoot Hodgkin *"for her determinations by X-ray techniques of the structures of important biochemical substances"*.





insulin

"The Nobel Prize in Chemistry 1964". *Nobelprize.org.* Nobel Media AB 2014. Web. 22 Oct 2014. <http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1964/>

Benefits of retaining derived data

Crystal and molecular structures: the main output from crystal structure determination experiments

- *Scientific record*
- *Database-driven discovery*
- *Protein-ligand interactions*
- *New pathways to synthesis, manufacturing, energetics…*
- *Identification/indexing (e.g. forensic science)*

Benefits of retaining processed data

Experimental data, calibrated and reduced to provide the basis for deriving structural models

- *Structure validation*
- *Re-refinement*
- *Systematic bias, methods development*
- *Guard against structures associated with incorrect data sets*
- *Help guard against 'bad apples' in the databases (Minor et al. 2016)*

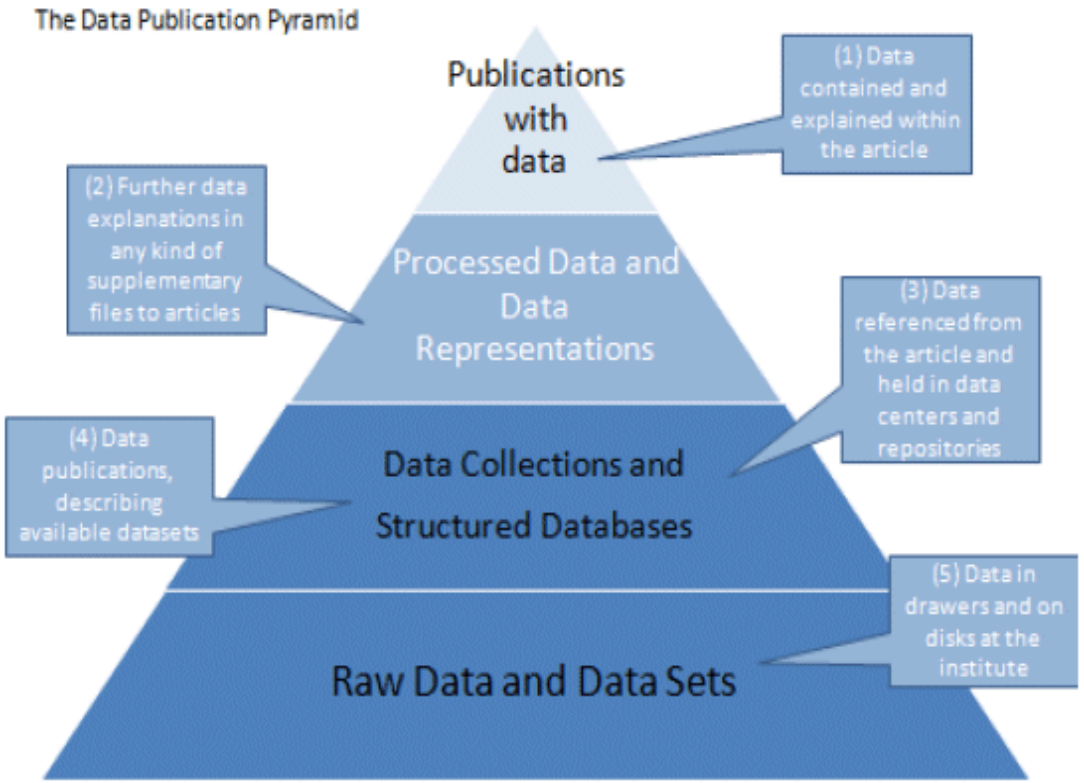
Benefits of retaining raw data

Experimental data fresh from the detector or measuring apparatus: minimal correction/interpretation

- *Methods development*
- *Benchmarking of software algorithms*
- *Maximise reproducibility (with sufficient metadata)*
- *Data analysis at higher resolution*
- *Understanding of multiple crystal lattices*
- *Diffuse scattering: correlated motions/disorder*

The philosophical view of the importance of access to raw diffraction data

Analysis through one's own eyes – not the lens of someone else



"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.

IUCr has set up a Diffraction Data Deposition Working Group (DDDWG) and whose members 2011 to 2017 are:-

- *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)*

Overview

Towards a future structural science based on raw data archiving: the IUCr's Diffraction Data Deposition Working Group (DDDWG)

- ***IUCr Commissions*** are actively working on

  "defining their commission's metadata for raw diffraction data"

  namely:-

  Commission on EXAFS;
  Commission on Small angle scattering;
  Commission on High pressure;
  Commission on Biological Macromolecules.

- ***The ICDD has been active on the harnessing of raw powder diffraction data sets*** for some time and reported to us at ECM30 in Rovinj that they now have incorporated 10,000 raw powder diffraction data sets into their powder diffraction file. ***The Commission on Powder Diffraction*** is planning further work on neutron powder diffraction raw data and will liaise with the Commission on Neutron Scattering as appropriate.
- The ***Commission on Structural Chemistry*** had enthusiastic participants in Madrid, Bergen and Rovinj DDDWG events.

Status mid-2016

- *Several domain-specific raw data repositories established*
  - *Integrated Resource for Reproducibility in Macromolecular Crystallography*
  - *Structural Biology Data Grid*
- *University digital library/data archiving services (U. Manchester)*
- *General public data repositories (Zenodo; ResearchGate; Dryad)*
- *Moves towards archiving at experimental facilities (ESRF; ISIS, Rutherford Appleton Laboratory; Diamond)*
- *IUCr journals accept DOIs of archived raw data sets as acceptable supplementary information*
- *IUCr initiatives to characterize and collect essential experimental metadata for raw data*

Let this Session at SciDataCon2016 commence!