



The Exemplary Crystallography and Structural Databases

This excellent educational course will describe the diverse and thriving ecosystem of crystallographic and structural databases.

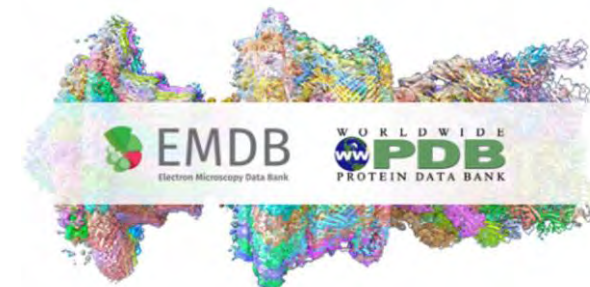
John R Helliwell

School of Chemistry, University of Manchester, UK

john.helliwell@manchester.ac.uk



Crystallography Open Database



Talk contents

- Crystallographers and structural scientists have long accepted the maxim “*Take nobody’s word for it*” because “*the science is in the data*”.
- In the first crystal structure report (*Bragg 1913*) *the diffraction data were included*.
- They have been among the first scientific communities to *exploit technology developments for expanding their data archiving scope*.
- This *ecosystem is underpinned by data exchange standards*, the crystallographic information framework “*cif*”, and the desire to ensure the highest achievable quality, making a strong world-wide data infrastructure.
- There is then an *exemplary provision for molecular structure scientists of collections of their data in these databases*, be they for *biology, chemistry or materials science*, which is widely admired.
- Indeed, the *use of these vast collections of data is extremely broad*, and yield strong basic science and high societal impacts, which the course contributors will each describe.

A couple of core references and websites

- Bragg, W.L. (1913) *The structure of some crystals as indicated by their diffraction of X-rays* Proc. R. Soc. London, Ser. A89, 248–277.
- National Academies of Sciences, Engineering, and Medicine 2019. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25303>.
- The International Science Council <https://council.science/> accords great importance to data and has a dedicated committee on data, known simply as CODATA <https://codata.org/>, with many decades of experience of good policy and practice.
- The International Union of Crystallography is proactive in promoting community discussion and data standards including within its Journals <https://www.iucr.org/resources/data>



Crystallographers, as well as astronomers and particle physicists, have always greatly valued their data



Olga Kennard
1st IUCr Representative to ICSU's CODATA

Provenance as the basis for trust in crystallography also defines best practice

- In practice there must be:
 1. Trust in the crystallography
 - Articles, database entries, the data sets
 2. Trust in the process of crystal structure analysis
 - Metadata for the experiment & the workflows to deriving the model
 3. Trust in the measurement at source
 - Raw diffraction data as the ground truth

Reproducibility of the study can lead to replicability by other studies^{\$}

^{\$} National Academies of Sciences, Engineering, and Medicine. 2019.
Reproducibility and Replicability in Science, Washington, DC, USA.

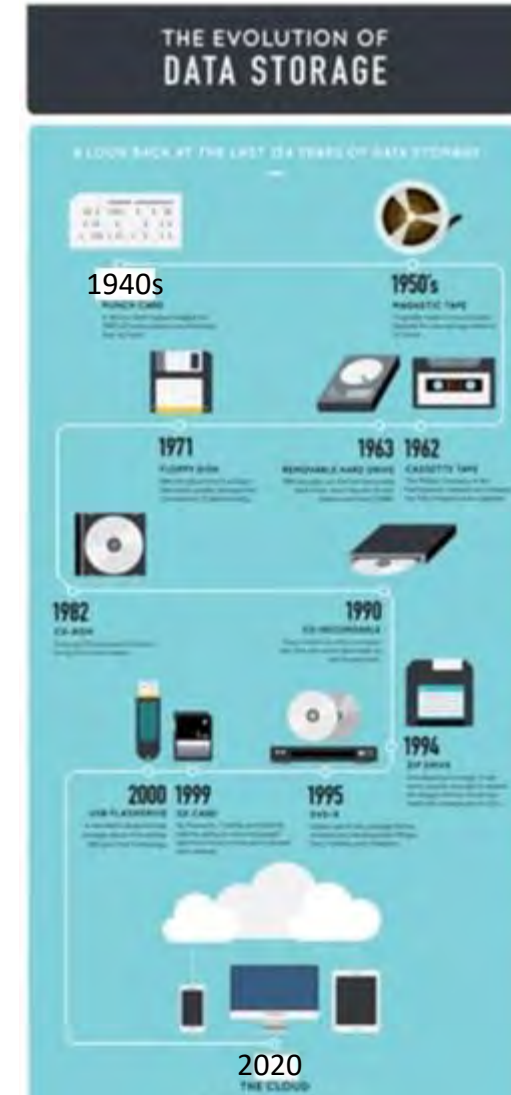
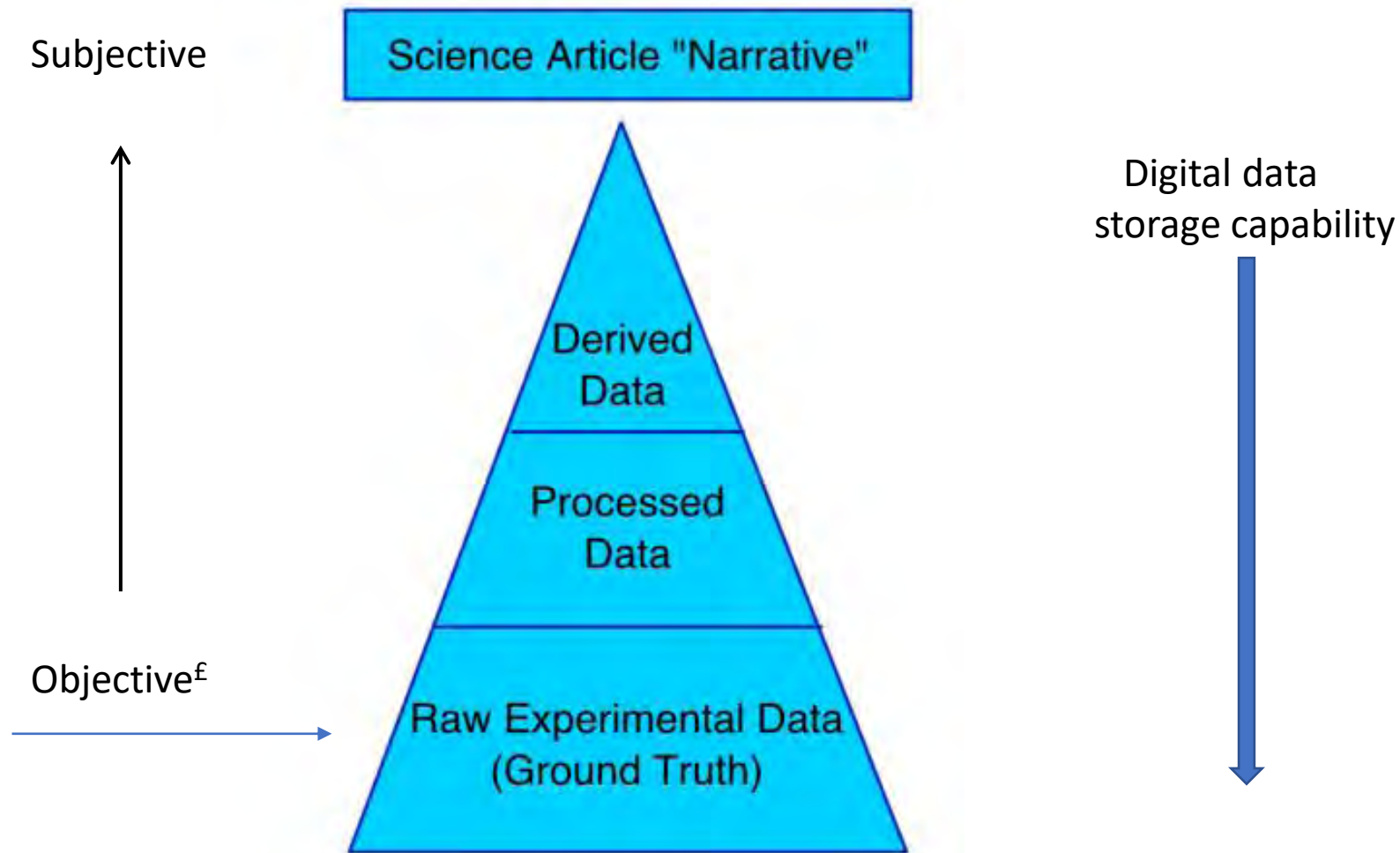
Data storage, and checking, facilitates trust in a study

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

- A reader should not just believe the authors' words in a publication. So the *underpinning data are paramount*.
- The *archiving of processed diffraction data and final derived molecular coordinates in crystallography have achieved an exemplary state of the art* relative to most fields via our databases.
- *What about the raw (unprocessed) experimental data archiving?* This is the raw data archiving challenge, in size and need for recording its correct metadata.
 - *In the USA* there are two resources: for MX there is the University of Virginia proteindiffraction.org and at Harvard University for all structural biology methods the SBDaGrid. You will have lectures on these.
 - *Elsewhere in the World:* PDBj is pioneering its XRDa linked to PDBj. In Europe popular archives are Zenodo, which will become the European Open Science Cloud, and the European Photon and Neutron Open Science Cloud (Led by ESRF and the Institut Laue Langevin) as well as the University research data archives especially in the UK.

Today we can start to include our raw data as part of our preserved workflow



[£] Nb an instrument must be calibrated by a person and this leaves some degree of subjectivity

Some history; let's go back to the first crystal structures.....

The diffraction data were included from the outset by the Braggs in their scientific records



William Lawrence
Bragg

Experimental diffraction data 1913

Bragg, W. L. (1913). *The Structure of Some Crystals as Indicated by their Diffraction of X-rays*. Proc. R. Soc. London Ser. A, 89, 248-277; contained a lot of raw data (Laue patterns)

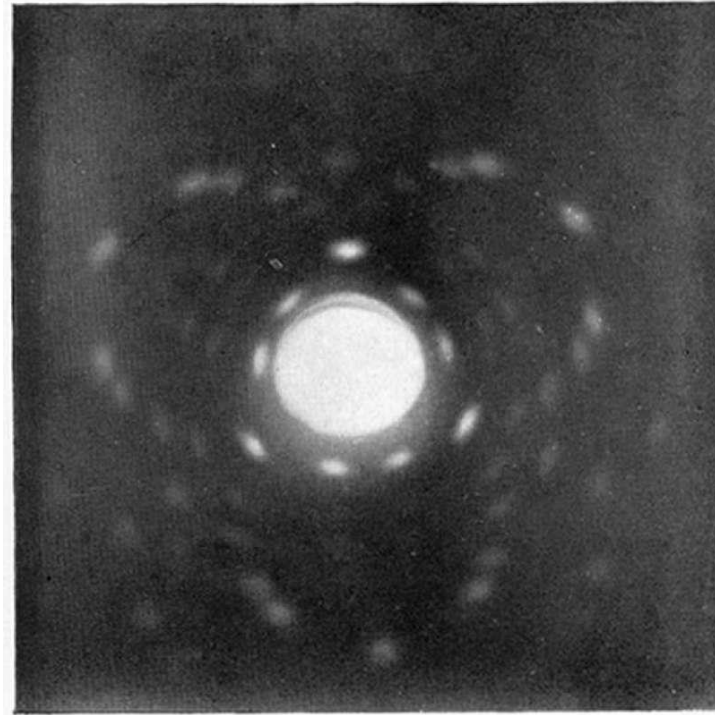


FIG. 11.—Fluorspar.

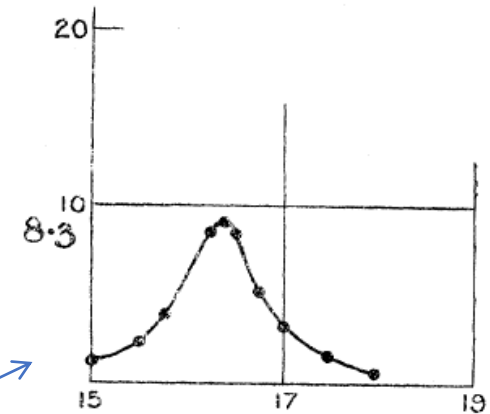


FIG. 1.—Regular reflection from cleavage face of rock-salt, glancing angle 8.3° .

**W. H. Bragg and W. L. Bragg *The Reflection of X-rays by Crystals*
Proc. R. Soc. Lond. A 1913 88, 428-438 using the WHB X-ray spectrometer**



Sir William Henry Bragg

Processed diffraction data 1913

Now assume a face centred lattice

	Calcd	Actual value found
$d_{100} = \frac{a \cdot 951}{2} = 3.04 \times 10^{-8}$	$\sin \theta_{100} = \frac{\lambda}{2d} = .0947$ $\theta = 5.43$	5.35
$d_{1\bar{1}0} = \frac{a \cdot 779}{2} = 2.48 \times 10^{-8}$	$\sin \theta_{1\bar{1}0} = .1161$ $\theta = 6.67$	6.6
$d_{110} = \frac{a \cdot 601}{2} = 1.917 \times 10^{-8}$	$\sin \theta_{110} = .1502$ $\theta = 8.63$	8.70
$d_{111} = \frac{a \cdot 437}{2} = 1.79 \times 10^{-8}$	$\sin \theta_{111} = .2064$ $\theta = 11.92$	11.77
$d_{2\bar{1}\bar{1}} = \frac{a \cdot 206}{2} = 1.43 \times 10^{-8}$	$\sin \theta_{2\bar{1}\bar{1}} = .2064$ $\theta = 11.92$	11.65

From the Braggs' notebook, held at the [University of Leeds Library](#), showing observed angular locations of diffracted beams from different crystal planes, and their calculated values.

Plane.	Calculated Spacing.	Calculated Angle.	Observed Angle.
100	$d_{(100)} = 3.04.$	$5^{\circ} 26'$	$5^{\circ} 21'$
$1\bar{1}0$	$d_{(1\bar{1}0)} = 2.48.$	$6^{\circ} 40'$	$6^{\circ} 36'$
110	$d_{(110)} = 1.92.$	$8^{\circ} 38'$	$8^{\circ} 42'$
111	$d_{(111)} = 2.79.$	$5^{\circ} 55'$	$5^{\circ} 46'$
$2\bar{1}\bar{1}$	$d_{(2\bar{1}\bar{1})} = 1.43.$	$11^{\circ} 35'$	$11^{\circ} 39'$

A similar published tabulation from their [1915 book X-rays and Crystal Structure](#).



William Lawrence
Bragg

Derived data:

Sodium chloride crystal structure coordinates 1913

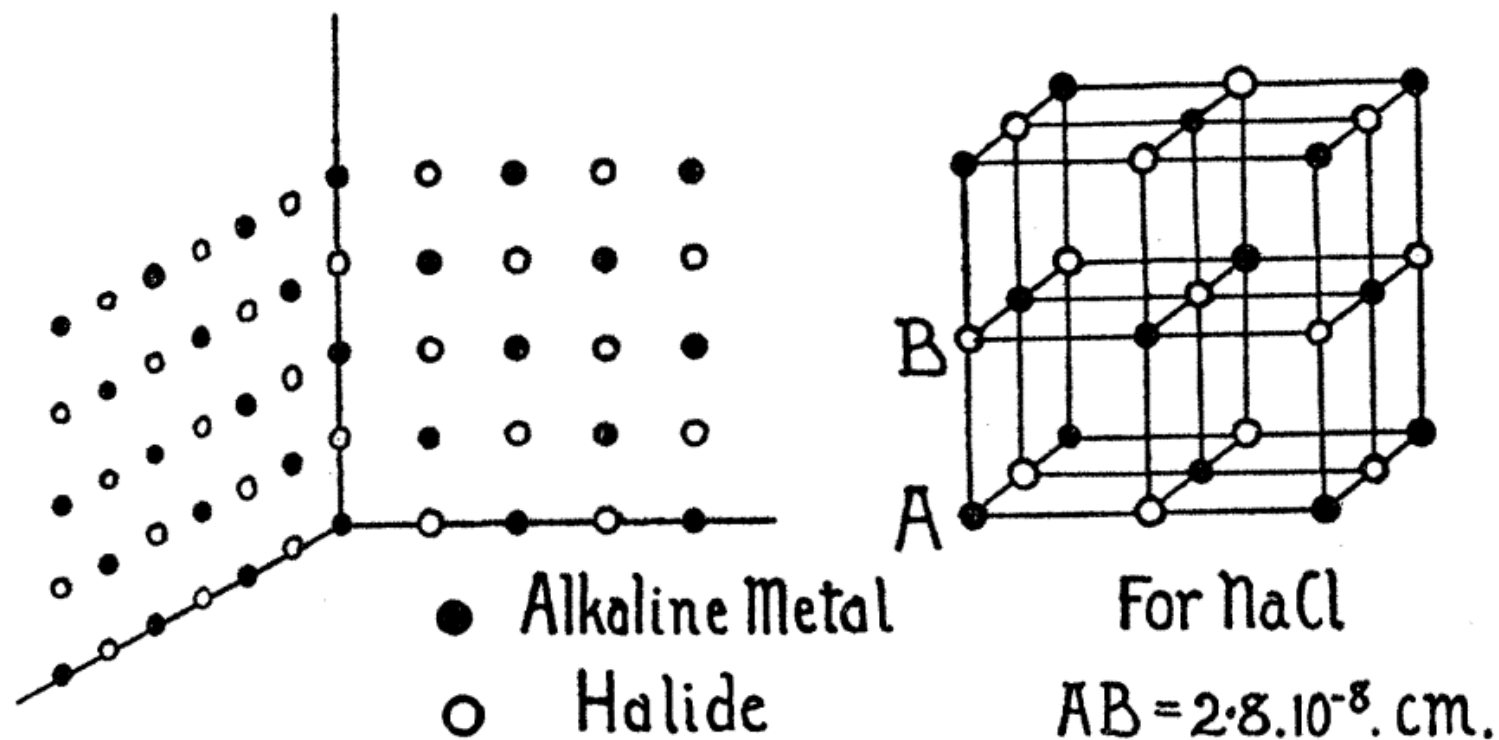


FIG. 10.