



ECM28 Lunchtime Open Forum on Diffraction Data Deposition (DDD)

John.R. Helliwell

**Chair of the IUCr Working Group on DDD and IUCr
Representative to CODATA**

Held at:-

The University of Warwick, Coventry, England
Thursday, August 29, 2013

Context:-

- This Forum for open discussion is organised by the DDD Working Group, appointed by the IUCr Executive Committee to define the need for and practicalities of routine deposition of our primary experimental data;
- It will take the form of a short review of progress during the Working Group's two years of activity;
- It is an opportunity for input from the community represented at ECM28;
- A summary of Public input thus far is at the IUCr Forum:-
- <http://forums.iucr.org/>

The IUCr Diffraction Data Deposition Working Group

Established by IUCr Summer 2011



Terms of Reference

- It is becoming increasingly important to deposit the raw data from scattering experiments;
- A lot of valuable information gets lost when only structure factors are deposited.
- A number of research centres, *e.g.* synchrotron and neutron facilities, are fully aware of the need and have established detector working groups addressing this issue.

Membership

Full members

Steve Androulakis <i>Australia</i>	Brian McMahon <i>UK</i>
Sol Gruner <i>USA</i>	Tom Terwilliger <i>USA</i>
John R. Helliwell , Chair <i>UK</i>	John Westbrook <i>USA</i>
Loes Kroon-Batenburg <i>The Netherlands</i>	Heinz-Josef Weyer <i>Switzerland</i>

By invitation

Chairs and delegates of IUCr Commissions

Consultants

Currently five specialists in data archiving, software development and macromolecular crystallography

Why publish data?

Some reasons:

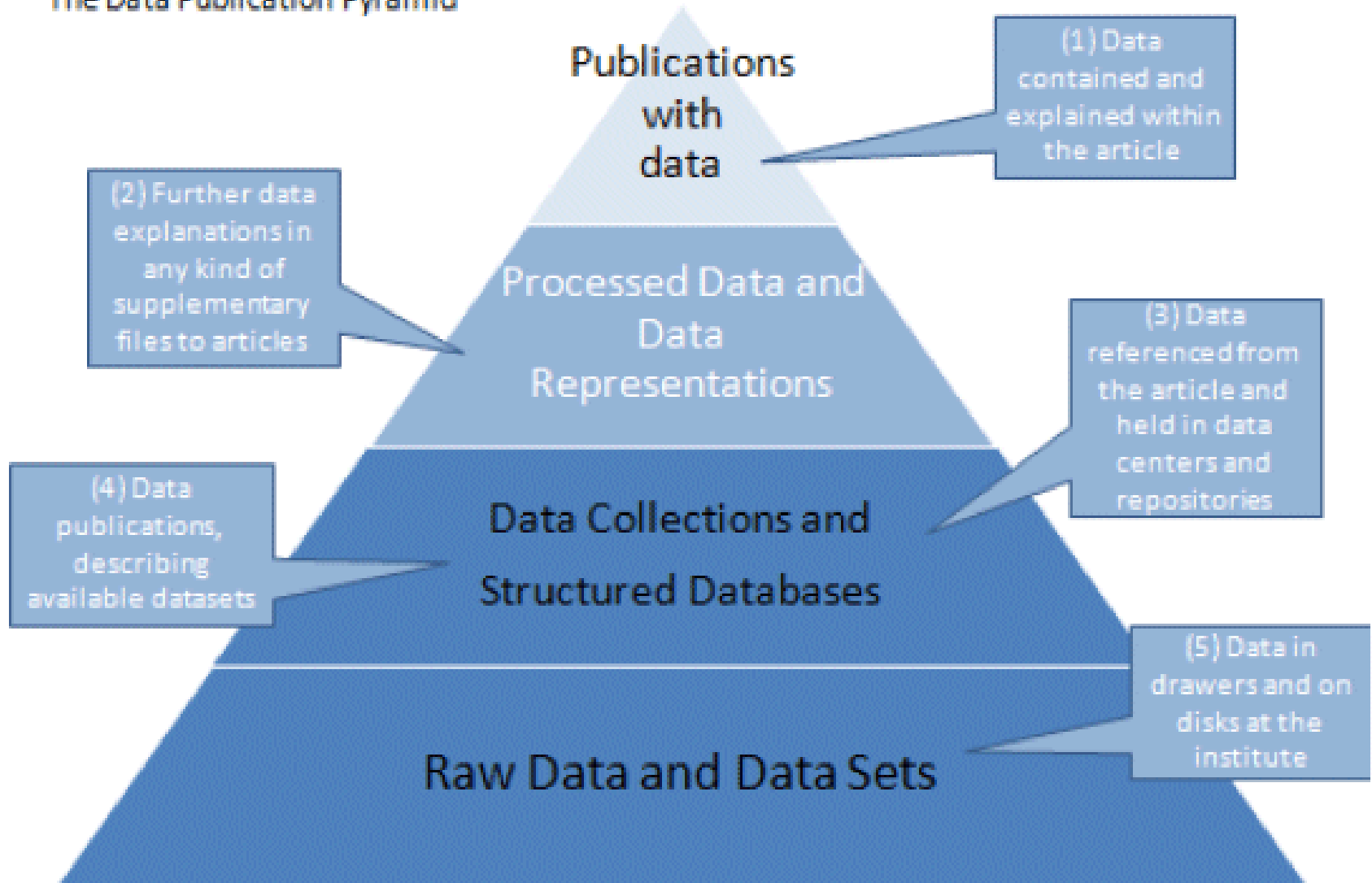
- Verify or support the validity of deductions from an experiment
- Safeguard against error or fraud
- Allow other scholars to conduct further research based on experiments already conducted
- Allow reanalysis at a later date, especially to extract 'new' science as new techniques are developed
- Provide example materials for teaching and learning
- As a mechanism for long-term preservation of experimental results

Why publish raw crystallographic diffraction data?

- Just what is the symmetry layout of a crystal namely its space-group symmetry?
- Just what is the diffraction resolution limit?
- The diffuse scattering may be significant and yield details of conformational mobility and or flexibility;
- Raw data availability can be used by developers to improve software ie our data processing tools;
- Raw data being an obligatory requirement for publication could serve to prevent fraud;
- Structure determination cannot proceed and needs a wider community effort eg if the diffraction is from an awkward composite of crystals.

In general the data challenge is:-

The Data Publication Pyramid



Current perspective

- There is enthusiasm and encouragement to archive more than derived or processed data in many areas of science besides our own.
- The crystallographic community prides itself in making its processed data accompany its publications; indeed it has been obligatory these last 10 years or so.
- We have three practical options in the near future to extend these principles to our raw data;
 - via a local Data Archive
 - via synchrotron or neutron facility data storage
 - or via the corresponding author setting up a personal weblink to datasets underpinning publications on their personal websites.

The future as seen by the particle physicists

- Use cloud storage;
- Our reaction as crystallographers:-
- Does this mean using commercial data storage suppliers like Google?
- So, do we feel comfortable trusting our data to a commercial agent?
- Cost issues also need to be evaluated carefully, but look promisingly, ie relatively, cheap;

Initial recommendations to the IUCr Executive Committee by the IUCr DDD WG in December 2012:-

- 1) *Authors **should** provide a permanent and prominent link from an article to the raw data sets underpinning a journal publication*

(with a view to making this a formal requirement on authors at such time as the community has adopted raw data deposition as a routine procedure)

- 2) *Commissions should be charged with the task of defining experimental metadata relevant to their scientific fields in order to harmonise raw data archiving at disparate facilities*

“should” changed to “may” by IUCr Exec at its meeting held Dec 2012 in Adelaide.

Some thorny questions/points:-

- Do people actually request or air a view wishing to have access to raw data, whether published or unpublished?
- How long should the raw data be available? In perpetuity in the case of publication?
- After a time period, without a publication, should raw data derived from public funding be mandated for release? Some research fields operate such a mandate after 3 years (eg space research);
- Local data archivists, rather than those at a specialised centralised repository, may be inexperienced at checking that depositors give all necessary metadata thus rendering the raw data of limited future use by other researchers;

A relatively easier question

- Would there be a way of annotating raw data sets that are not linked to a publication in a kind of annotation data base? E.g. structure not solved; indexing not succeeded; multiple lattices; diffuse scattering..... It would be sufficient to have one or two images in the data base. The data base then links to where the raw data reside (probably local archive).