Introduction, Welcome and Update on behalf of the IUCr Diffraction Data Deposition Working Group (DDDDWG)

from

John R Helliwell & Brian McMahon

A Hearty Big Thankyou to Our Sponsors &
To the Croatian Association of Crystallographers &
All Our Speakers from Around the World!
Contents

• 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
• Brief recap of 2011 to 2014; our DDDWG Report to the IUCr Montreal 2014 General Assembly
• The challenge of the sheer volume of our raw diffraction data
• The challenge of achieving a new-user depth to raw diffraction data metadata description i.e. can a new-user successfully understand and process any raw data set he/she wishes to analyse?
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
• analysing the diffuse scattering that reflects correlated motions or disorder of atoms in the crystals
Benefits of retaining derived data

- 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

- Structure validation
- Re-refinement
- Systematic bias, methods development
- Guard against structures associated with incorrect data sets
IUCr Diffraction Data Deposition Working Group (DDDWG) report to the IUCr General Assembly Montreal

John R Helliwell on behalf of the DDDWG
Recommendations from the DDDWG for the upcoming Triennium

• IUCr Commissions to define their metadata;
• J. Appl. Cryst. to introduce a ‘Difficult Raw Data’ Section (Loes Kroon-Batenburg);
• A centralised crystallographic repository of raw data set metadata should be scoped, including a search interface, leading to a pilot service;
• With a viable pilot metadata registry authors should provide a permanent and prominent link from an article to their raw data sets underpinning a journal publication.
Issues for the IUCr

• The IUCr’s science involves ‘Big data’ up towards the level of the data-deluge of the Square Kilometre Array radio telescope; we may have to consider subsets of data retention or limited time periods for retention;

• Rights of access to publicly funded, but unpublished, crystallographic research data after e.g. 3 to 5 years.
Members of the DDDWG 2011 to 2014

- John R Helliwell and Brian McMahon (UK), Chair and Co-Chair;
- Steve Androulakis (Australia)
- Sol Gruner (USA)
- Loes Kroon-Batenburg (Netherlands)
- Tom Terwilliger (USA)
- John Westbrook (USA)
- Heinz-Josef Weyer (Switzerland)
Members of the DDDWG 2014 to 2017

- John R Helliwell and Brian McMahon (UK), Chair and Co-Chair;
- Steve Androulakis (Australia)
- Dolothea Szebenyi (USA)
- Loes Kroon-Batenburg (Netherlands)
- Tom Terwilliger (USA)
- John Westbrook (USA)
- †Heinz-Josef Weyer (Switzerland)
The Cloud to help solve the raw data storage challenge?

Cloud cover

Opposition to storing vast scientific data sets on cloud-computing platforms is weakening.

08 July 2015
Charge to the Workshop Participants

• **Define** your Metadata or at the least Define the Challenges you face

• **We have provided a template form** for you to supply information about metadata for your specific research field / IUCr Commission

• **We consider the challenge of aligning scientific metadata with generic standards** like the ‘**Dublin Core of Metadata descriptors**’

• **We have to understand each other’s fields; this will assist the core challenge** of ‘**seeing a data set through a new-user’s eyes**’
Our Workshop Sessions

• Session I: Introduction
• Session II: Diffraction images - what can we get out?
• Session III: Metadata for diffraction images and other experimental methods
• Session IV: Data in the Wider World - From Laboratory to Database
• Session V: What new metadata items are needed?
• Session VI: Metadata schemas
Let’s go to it!