An IUCr XXI Congress Workshop on
'Crystallographic Data Publication'

The methods and approaches to publishing scientific research output (i.e. making it available to the public domain) are currently undergoing a seed-change. New initiatives include such processes as author self-deposition of reprints in Institutional Repositories, author-pays open access journals and open discussion of results in Wiki's or Blogs, and these are gathering momentum. On one hand this movement has been slow to appear in some disciplines (eg Chemistry), yet is widely accepted and routine in others (eg Physics and Biology). The field of crystallography is a data-rich subject that supports a vast number of conventional journal articles across a very broad range of disciplines, and seeds of change are very much evident here.

Recent developments include:

1. The IUCr prepared an extensive report to the Science and Technology Committee in 2004 regarding electronic publishing, data and open access (http://www.iucr.org/iucr/stcttee04.html) and a position paper (http://www.iucr.org/iucr/gicsi/positionpaper.html).

2. IUCr were awarded three rounds of funding to enable the content of their journals provided by UK authors to be made publicly available through Open Access (JISC funded) and examine the culture changes and impact on their business model.

3. From Jan 2008 Acta Cryst Section E will become Open Access (Author pays).

4. Numerous successful projects aimed at authors self-publishing data (eCrystals, ReciprocalNet, Crystallography Open Database) have been created.

The indications are therefore that the time is now right to explore these new routes to publishing, with a particular emphasis on the crystallographic raw and results data, as an alternative to conventional journal articles. This is especially relevant in the light of a recognised data deluge (F.H. Allen, *Crystallogr. Reviews*, **2004**, 10, 3-15), where it is commonly recognised that approximately only 20% of small-molecule crystal structures determined as a result of publicly funded research reach the public domain. This is a direct consequence of the way in which data is traditionally published.

The principal contributions to this workshop would be from those responsible for recent developments and innovations in the area and are outlined below. The primary aims of the workshop will be to:

1. Raise awareness in the crystallographic community of these new innovations and the opportunities they offer for data publication, and invite participation.

2. Enable projects and interested parties working in the area to compare approaches and develop a unified way forward for the future of data publication

The topics to be covered as sessions of the workshop include:

1. Self publication by an author: Institutional data repositories (eg eCrystals project), Open databases (e.g. COD) and consortia (eg ReciprocalNet)

2. Author pays Open Access (eg Acta Cryst E, Chemistry Central)

3. Interoperability and linking between data sources

4. Interactions between data sources and data centres, publishers & learned societies

5. Data discovery and linking to journal articles and other sources of electronic information openly available

6. Support, training, raising awareness and advocacy for data publication

7. Future sustainability and preservation of openly available data, -policy and finance.

8. Ownership of data

The workshop would be led by the eCrystals project (see Appendix 1 for a full outline) and its supporting partners. The primary contact is Dr Simon Coles (School of Chemistry, University of Southampton, UK; s.j.coles@soton.ac.uk) and co-organisers will be Prof M.B. Hursthouse (Southampton), Dr J.C. Huffman (ReciprocalNet, Indiana), Dr. F.H. Allen (CCDC) and representatives from IUCr Chester (B McMahon) & the Protein Data Bank (J. Westbrook). A proposed date for the workshop is 23rd August and it would be most useful if the venue were the same as the main Congress meeting. We expect that something in the region of 200 participants would attend and we would require projection, public announcement and internet facilities.

Principal contributors would include:

The **eCrystals** project outlined below.

The **International Union of Crystallography** (http://www.iucr.org/) has an interest in all forms of data publication and the preservation of the arising open data.

The **Cambridge Crystallographic Data Centre** (http://www.ccdc.cam.ac.uk/) as an operator of the world repository for all carbon containing crystallographic data published in journal articles (the CSD).

The **Inorganic Crystal Structure Database (ICSD)** is the official database for all chemical crystal structures that do not contain carbon. Operated and sold under license by FIZ Karlsrhure: (http://www.fiz-karlsruhe.de/icsd.html?&L=0)

The **Protein Data Bank (**http://www.pdb.org/pdb/home/home.do) is an open source database of biological crystal structures and the official collection of all published data.

The **ReciprocalNet** (http://www.reciprocalnet.org) where members of a consortium share their crystallographic results data.

The **Crystallography Open Database** (http://www.crystallography.net) where CIF files can be uploaded to an open repository.

Further initiatives that will be contacted or represented include:

SPECTRa http://www.lib.cam.ac.uk/spectra/

CrystalEye http://wwmm.ch.cam.ac.uk/crystaleye/

MMSN http://mmsn.net.au/

CrystalGrid http://www.crystalgrid.org

Chem Refer http://www.chemrefer.com/

CiteSeer http://citeseer.ist.psu.edu/

eMolecules http://www.emolecules.com/

ChemSpider http://www.chemspider.com/

PubChem http://pubchem.ncbi.nlm.nih.gov/

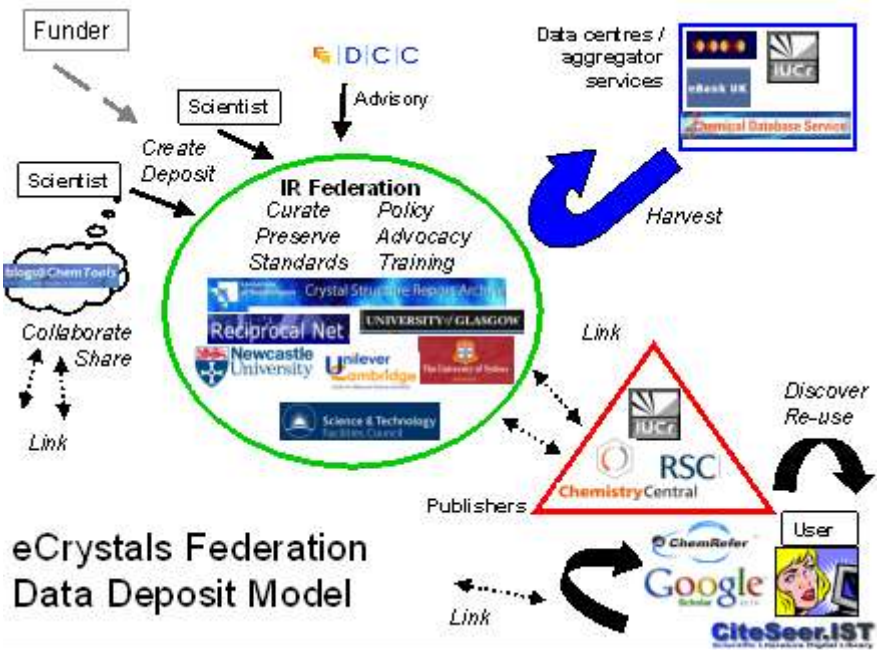
RSC Prospect http://www.rsc.org/Publishing/Journals/ProjectProspect/

Appendix 1: The eCrystals project

The **eCrystals Federation**, is aimed at establishing a solid foundation of crystallography data repositories across an international group of partner sites, with metadata harvested by a number of aggregrator services (CCDC, CDS) and builds on the previous work of the **eBank Project (**http://www.ukoln.ac.uk/projects/ebank-uk/**)**. Partners are working together to enable data repositories operating on different platforms (EPrints, DSpace, Fedora & ReciprocalNet), investigate aggregation issues arising from harvesting metadata from repositories within the networked information environments in other countries (EU, USA & Australia) and enable the Federation of institutional archives to interoperate with international subject archives (IUCr and CCDC) and other third party harvesters. The project is also addressing the areas of data curation, preservation and provenance.

This very successful collaboration, now formally includes participation from the wider chemistry community (Unilever Centre at Cambridge, the Science & Technology Facilities Council STFC, Glasgow University, Newcastle University, University of Sydney, ReciprocalNet at Indiana), digital libraries (UKOLN, StORe & ARCHER), digital preservation experts (Digital Curation Centre with STFC), publishers (International Union of Crystallography, IUCr, Royal Society of Chemistry RSC & Chemistry Central), the Chemical Database Service (CDS) and the key learned society & subject oriented bodies in the field (IUCr & the Cambridge Crystallographic Data Centre, CCDC).

Figure 1: The eCrystals Federation Data Deposit Model.



The eCrystals project is led by the University of Southampton, with core partners at UKOLN (University of Bath), the Digital Curation Centre and the Unilever Centre (University of Cambridge). **Supporting** partners are:

The International Union of Crystallography (http://www.iucr.org/), a body serving the world community of crystallographers, promoting international standardisation and cooperation in crystallography and enabling international publication of crystallographic research; **The Cambridge Crystallographic Data Centre** (http://www.ccdc.cam.ac.uk/), operator of the world repository for all small-molecule organic and metal-organic crystallographic data published in journal articles; **The Royal Society of Chemistry** (http://www.rsc.org/), the largest organisation in Europe for advancing the chemical sciences supported by a worldwide network of members and an international publishing business; **The Chemical Database Service** (http://www.cds.dl.ac.uk), provider of a single interface to search all available crystallographic databases (CSD, ICSD, CrystMet); **Chemistry Central** (http://www.chemistrycentral.com), an emergent Open Access series of journals, covering all aspects of the chemistry domain; **StORe** (http://jiscstore.jot.com/WikiHome), a JISC Digital Repositories Programme project, involved in linking 'source' and 'output' repositories; **CLADDIER** (http://claddier.badc.ac.uk/trac), a JISC Digital Repositories Programme project, concerned with

linking data to publications; **The Research Information Network** (http://www.rin.ac.uk/), an organisation funded to assess and coordinate the provision of research information in the UK; **ReciprocalNet** (http://www.reciprocalnet.org/), a well established consortium of partners (US based but also including USyd and NCS) sharing and publishing crystallographic data; **ARCHER** (http://archer.edu.au/), a DEST project based on data repositories supporting the research process; **The University of Sydney** (http://mmsn.net.au/), a lead partner in the Molecular and Materials Structure Network, (an Australian Research Council funded project); **The Science and Technology Facilities Council** (http://www.scitech.ac.uk), a multidisciplinary research organisation supporting scientists and engineers world-wide through operation of large-scale research facilities and provision of strategic advice to the government on their development; **University of Glasgow** (http://www.chem.gla.ac.uk/xtal/)**,** a UK Chemistry department with a strong crystallographic section that supports a wide range of research; **Newcastle University** (http://www.ncl.ac.uk/xraycry/), a well established crystallography laboratory with world leading experience in operating a synchrotron based national service and a strong background in data-based publishing.