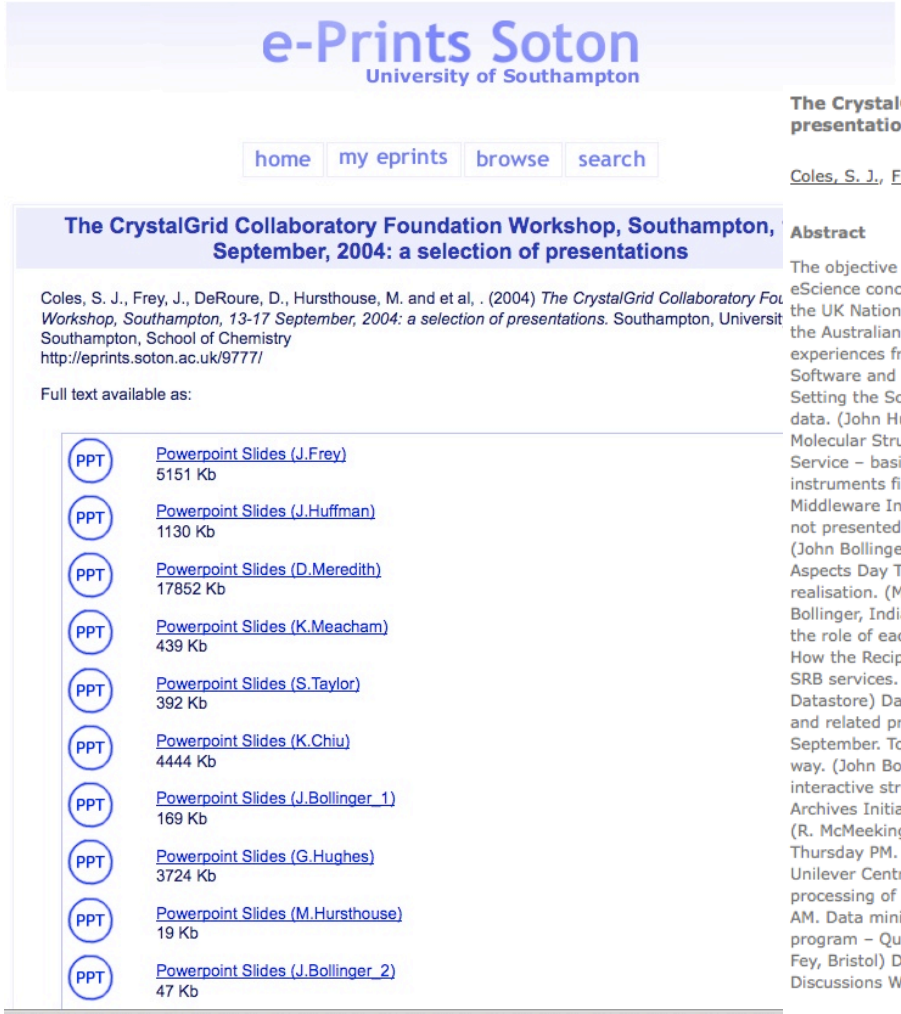
Back in 2004…

- A few crystallography groups working on ‘e-Science’ approaches to crystallographic data handling (principally Indiana, Southampton & Sydney)
- Identified a need for exchange of practice, standards, technologies, etc.



e-Prints Soton
University of Southampton

home | my eprints | browse | search

The CrystalGrid Collaboratory Foundation Workshop, Southampton, September, 2004: a selection of presentations

Coles, S. J., Frey, J., DeRoure, D., Hursthouse, M. and et al, . (2004) *The CrystalGrid Collaboratory Fou Workshop, Southampton, 13-17 September, 2004: a selection of presentations.* Southampton, Universit Southampton, School of Chemistry
http://eprints.soton.ac.uk/9777/

Full text available as:

- Powerpoint Slides (J.Frey) 5151 Kb
- Powerpoint Slides (J.Huffman) 1130 Kb
- Powerpoint Slides (D.Meredith) 17852 Kb
- Powerpoint Slides (K.Meacham) 439 Kb
- Powerpoint Slides (S.Taylor) 392 Kb
- Powerpoint Slides (K.Chiu) 4444 Kb
- Powerpoint Slides (J.Bollinger_1) 169 Kb
- Powerpoint Slides (G.Hughes) 3724 Kb
- Powerpoint Slides (M.Hursthouse) 19 Kb
- Powerpoint Slides (J.Bollinger_2) 47 Kb

The CrystalGrid Collaboratory Foundation Workshop, Southampton, 13-17 September, 2004: a selection of presentations (2004)

Coles, S. J., Frey, J., DeRoure, D., Hursthouse, M., Et Al, .

Abstract

The objective of the workshop was to discuss and plan the establishment of a global collaboratory for the development and application of eScience concepts and technology to the field of chemical crystallography and dependent or related subjects. The lead roles are taken by the UK National Crystallography Service (Southampton, UK), Indiana University Molecular Structure Center (Bloomington, IN, USA) and the Australian Molecular & Materials Structure Network (Sydney, Australia). A five day schedule of talks and discussions drew on work and experiences from numerous and varied groups: Monday 13th, Tuesday 14th September. Topics:- Grid Infrastructure, Grid Services – Software and Hardware; Security, Licence Keys, Automation etc. Monday AM. 1. Introduction and Welcome. (Mike Hursthouse, Soton). 2. Setting the Scene (just brief introductory outlines) The Comb-e-Chem Project. (Jeremy Frey, Soton) Crystallography: it’s all about the data. (John Huffman, Indiana) An overview of why and how the various past and present informatics projects in the Indiana University Molecular Structure Center have developed. The e-HTPX Project. (Dave Meredith, Daresbury Lab.) Monday PM. 3. Realisations The NCS Service – basic approaches. (Ken Meacham & Steve Taylor, IT-Innovation) The Semantic Grid. (Dave DeRoure, Soton ECS) Making instruments first-class members of the Grid. (Ken Chiu, Indiana). A discussion of the progress and goals of the Common Instrument Middleware Initiative (CIMA) project. Tuesday Continuation and further development of Monday topics. ECSES and NCS Service demos (if not presented Monday). (Ken Meacham, IT-Innovation) A distributed architecture for crystallography data, metadata, and applications. (John Bollinger, Indiana) A discussion of the purpose and design of the Reciprocal Net software suite. Wednesday 15 September. Data Aspects Day The e-Lab concept – The Smart Tea Exemplar. (Gareth Hughes, ECS Soton) The Crystallographic e-Lab – requirements and realisation. (Mike Hursthouse, Soton) Crystallographic Metadata. (Simon Coles, Soton) Data and metadata in the Reciprocal Net. (John Bollinger, Indiana) A few words about how the Reciprocal Net software currently classifies information into data and metadata, and about the role of each category in the Reciprocal Net system. Information management in a crystallography laboratory. (John Huffman, Indiana) How the Reciprocal Net site software is useful for tracking the information required for efficiently managing a crystallography laboratory. SRB services. (Peter Berrisford, RAL Data management group) The Atlas Datastore: Data archival and retrieval. (David Corney – Atlas Datastore) Data storage and archiving (John Huffman, Indiana) A brief discussion of the data archiving strategy planned for the SCrAPS and related projects, to be enabled via the CIMA project. Other speakers may contribute – to be decided on the fly. Thursday 16 September. Topics: Dissemination of Results. Harvesting and Aggregation Thursday AM. Disseminating crystallography results the Indiana way. (John Bollinger / John Huffman, Indiana) A presentation of the data dissemination aspects of the Reciprocal Net software, including interactive structure visualization, data tables, and automated graphics generation. Also a few words about exposing metadata via Open Archives Initiative protocols. eCrystallographyDataReports and the eBank project. (Simon Coles, Soton) The Chemical Database Service. (R. McMeeking, CDS) Publishing and the IUCr. (Peter Strickland, Brian McMahon, IUCr) Publishing and the CCDC. (Owen Johnson, CCDC) Thursday PM. Use/reuse of data. Data Base Aspects. CIF2CML. Data and software sharing in molecular science. (Peter Murray-Rust, Unilever Centre, Cambridge) Design of a GRID enabled database system to facilitate reuse, provenance tracking and automated processing of chemical information. (Rob Gledhill, Soton and Comb-e-Chem) Other speakers may be added. Friday 17 September Friday AM. Data mining, pattern searching, structure descriptors Management of data in the PDB (John Westbrook, Rutgers, US) The XPAC program – Quantification of Solid State Structure Similarity. (Thomas Gelbrich, Soton) Development of a Ligand Knowledge Base. (Natalie Fey, Bristol) Development of Molecular Geometry Knowledge Bases from the Cambridge Structural Database. (Steph Harris, Bristol) Discussions Wrap-up. Future pathways, prospects, recommendations.

Florence IUCr Congress 2005

Crystallographic Software Fayre
at the
IUCr Congress, Florence, Italy
Wednesday 24th August until Tuesday 30th August 2005

Formal Software Demonstrations and Bookings

[Crystallographic Software Fayre Homepage | What's New | Formal Software Demonstrations | Maps]

Sunday 28th August - All Afternoon	**Dr. Simon Coles (presenting)** School of Chemistry University of Southampton www: http://www.crystalgrid.org/ **Grid computing and crystallography** **CrystalGrid** is an informal consortium of academic groups from the UK, USA and Australia that are working with eScience and Grid computing in crystallography. This approach includes new ways to monitor or control diffraction experiments remotely, automated or remote data processing and novel approaches to data publication and sharing. The UK partners in CrystalGrid have obtained funding from the EPSRC to host a series of workshops around the world to promote understanding in this area and enable 'like minded' researchers to meet and exchange ideas. This session will comprise about 8 parallel demonstrations or poster presentations of systems developed by projects such as CombeChem, EBank, MMSN, CIMA, ReciprocalNet, CCP4, eHTPX, eCCP& COD. This is to be run as a mixer session so that congress delegates can visit at any point in the afternoon to discuss ideas or simply find out what it is all about.

Second CrystalGrid workshop (2007)

The Second International
CrystalGrid Workshop on
Crystallographic Data Management
April 26-28, 2007

Welcome
Workshop Overview
Registration
Local Arrangements
Documents

Managing Data...

Second International CrystalGrid Workshop on Data Management for Crystallography
April 26-28, 2007
Indiana University, Bloomington IN

Workshop Hosts
The Molecular Structure Center
Chemistry Department, Indiana University

Co-chairs
John C. Huffman, Indiana University, USA
Michael Hursthouse, University of Southampton, UK

Co-Sponsors
School of Informatics, Indiana University
Pervasive Technology Labs, Indiana University

The workshop will discuss two primary themes of standards for data and metadata, and data stewardship. The workshop will provide an opportunity for participants to see how other laboratories are organizing the deluge of data being generated by modern diffractometer systems. With the increased emphasis by funding agencies on cyber-enhanced instrumentation and remote access and collaboration, there are significant opportunities for laboratories to share instrumentation and expertise that were previously unavailable. Agreement on standards will lead to enhanced opportunities for collaborations. An expected outcome of the workshop will be a report enumeratng best practices and recommended standards in data management for diffraction methods. Participants will also identify areas where targeted research would lead to significant improvements in information management in crystallography as input to decsison making at funding agencies and commercial vendors.

Announcements
Program

NSF & UK e-Science programme funded.

Discussion Topics:

- Data and metadata formats
- Data lifecycles and stewardship
- Exchanging and using data
- Funding and training in crystallographic data management

http://www.crystalgrid.org/

Areas Addressed by CrystalGrid2007

Identified opportunities and needs for exchange of ideas and approaches to:

- Develop a standards-based infrastructure for archival of raw, processed and results data
- Derive operational and archival data formats and metadata schema
- Building community interest and involvement in the process of developing standards and building infrastructure for data management
- Secondary use and re-use of data sets for research, education, training and public science outreach activities

IUCrXXI Congress Satellite Workshop on “New Routes to Data Publication”

CrystalGrid aims to foster collaborative research in crystallographic informatics and e-Science, specifically including issues of data acquisition, representation, management and exchange.

Around 50 attendees:

- Small molecule diffraction
- Macromolecular crystallography
- Materials science
- Instrument & detector design
- Computer science & Informatics
- Research & Education networking
- Digital libraries
- Intellectual property law & Digital rights management

From:

- Universities
- National Laboratories & Synchrotrons
- International science organisations
- Instrument manufacturers
- Software providers

Operational & Archival Data Formats & Metadata Schema

- A coherent set of data formats (or variants of a single format) should be adopted for representing crystallographic data at all points of the lifecycle
  - Primary representation should be imgCIF family of formats
  - Further media, text and archival formats for requirements not met by imgCIF for data generated during the crystallographic experiment
- Metadata schema are required to describe ‘collections’ of data generated by a crystallographic experiment
  - Should be compatible with Open Archive Initiative (OAI)

Operational & Archival Data Formats & Metadata Schema

Specific areas for development:

- Translation of vendor specific formats to imgCIF
- Inclusion of imgCIF capability in community developed software
- Format for containing entire datasets
- Explore OAI compatibility & develop extensions to describe crystallographic data
- Standards for managing identity, privacy & access security at file, sample & archive levels
- Mechanisms to create and use persistent identifiers to refer to data sets and components thereof
- Standardisation of provenance information across imgCIF, CIF & NeXus

Standards-based infrastructure for archival of raw, processed and results data

- Promote archival and community stewardship of data as good scientific practice
- Need for standards and reusable software based on them for development of archives of raw, processed and published data
- Research & development into organisation, access protocols, archiving processes & workflows surrounding archives
- Standards for citation of datasets
- Community-wide policies on data accessibility
- Approaches to assess cost allocation for archive infrastructure and recurrent costs for its maintenance

Community Interest & Involvement in Standards, Infrastructure & Data Management

- CrystalGrid2007 workshop representation was broad (small molecule, macromolecular, neutron, etc)
- A 'broad & deep' effort must be mounted and sustained to develop necessary data management standards, practices & infrastructure
- Convene working groups to develop standards
- Further (international) workshops for broad community involvement
- Involvement of institutional policy makers and information managers
- Multi-institutional research projects, conducted by multi-disciplinary teams, to address issues arising from workshop

Secondary Use & Reuse of Data Sets for Research, Education, Training & Public Outreach

- Primary and derived data sets, process information and software can be used to add value to data already collected
  - Validation, instrument design, software improvement, data mining & starting points or additional support for related or tangential research
- Access to data & analytical techniques to support educational programs and professional training in chemistry, materials science & physics
- Enable public understanding & involvement in science

Need to develop systems that link operational day-to-day acquisition and archival processes for structural science to subject databases, information providers, publishers and educational programs