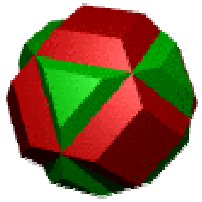
eBank UK



R4L Repository for the Laboratory

Metadata and Standards for Crystallographic Data Dissemination:

Implications for Raw Data Management and Archival

Simon Coles

School of Chemistry,

University of Southampton, U.K.

s.j.coles@soton.ac.uk

EPSRC

Engineering and Physical Sciences Research Council

JISC

Current Situation – Data Management

"Data from experiments conducted as recently as six months ago might be suddenly deemed important, but those researchers may never find those numbers – or if they did might not know what those numbers meant"

"Lost in some research assistant's computer, the data are often irretrievable or an undecipherable string of digits"

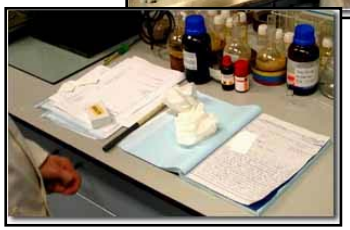
"To vet experiments, correct errors, or find new breakthroughs, scientists desperately need better ways to store and retrieve research data"

"Data from Big Science is … easier to handle, understand and archive. Small Science is horribly heterogeneous and far more vast. In time Small Science will generate 2-3 times more data than Big Science."

'Lost in a Sea of Science Data' S.Carlson, The Chronicle of Higher Education (23/06/2006)

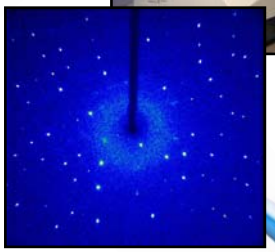
Experimental Metadata Generation

Synthesis

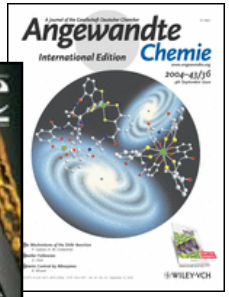


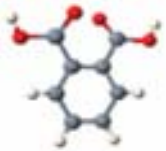
Preparation

Data Collection

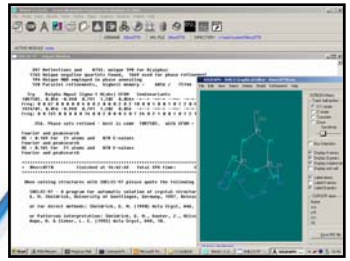


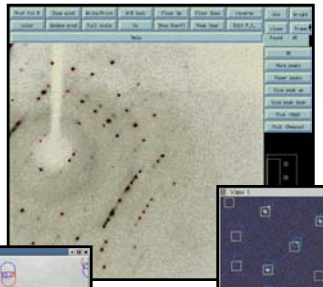
Publication

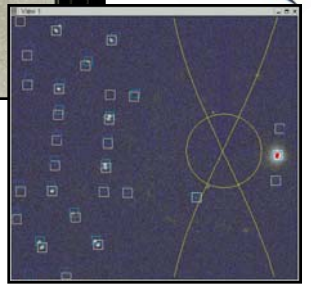




Structure Solution

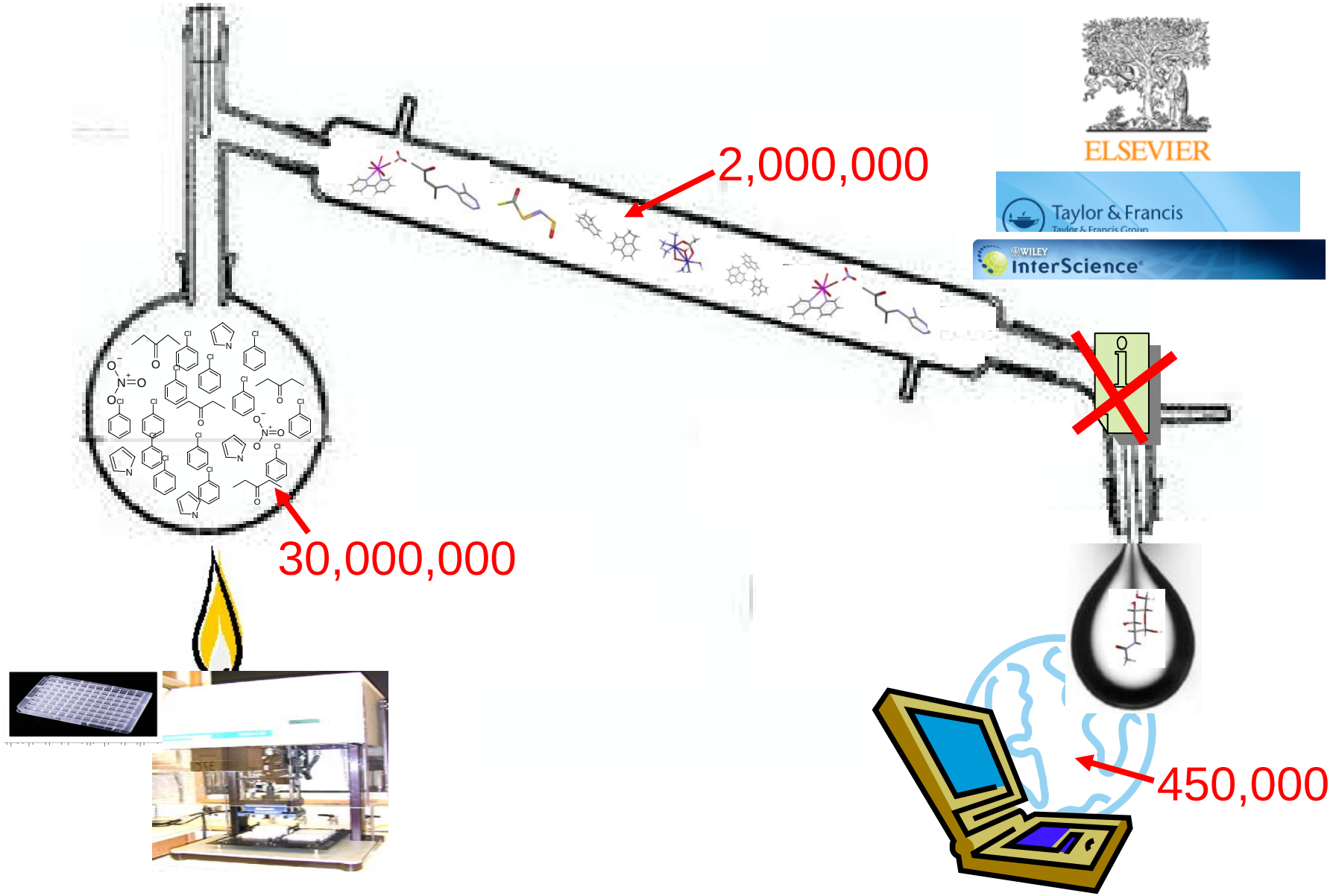






Data Processing

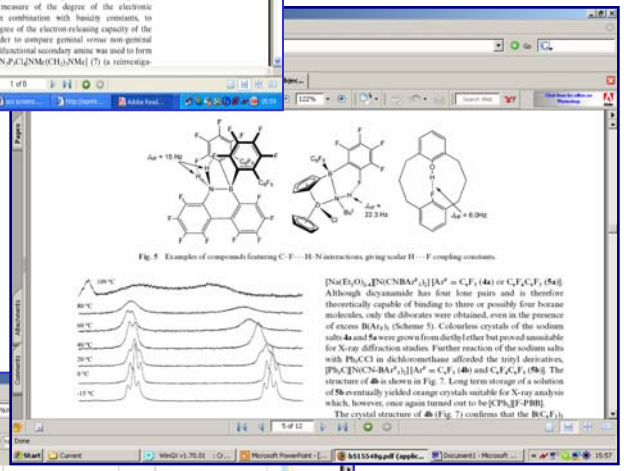
Current Situation - Data Deluge

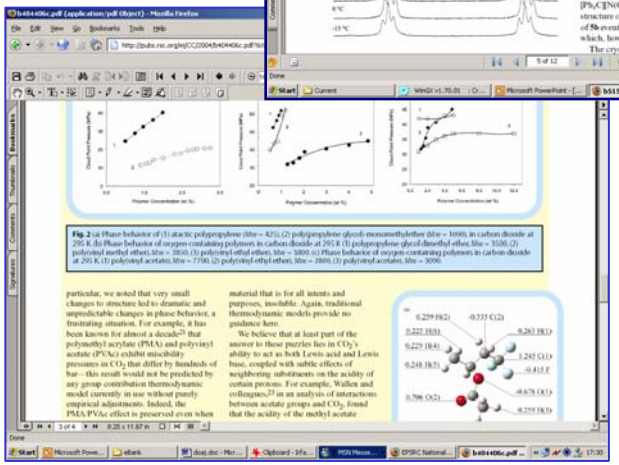


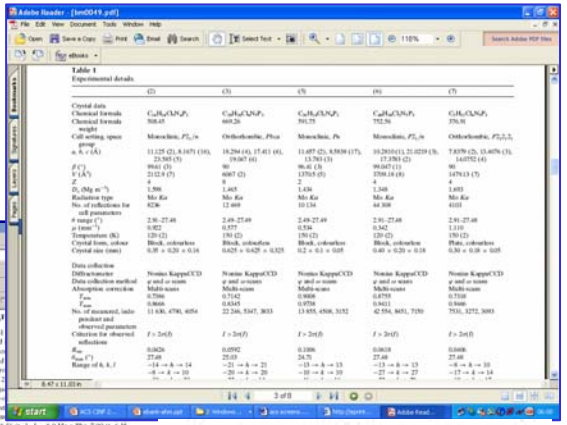
Separating Data from Interpretations

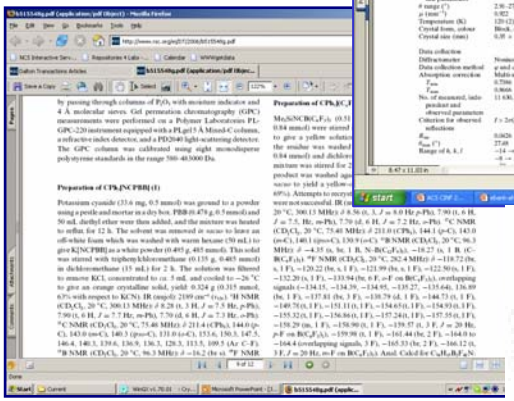
Intellect & Interpretation (Journal article, report, etc)

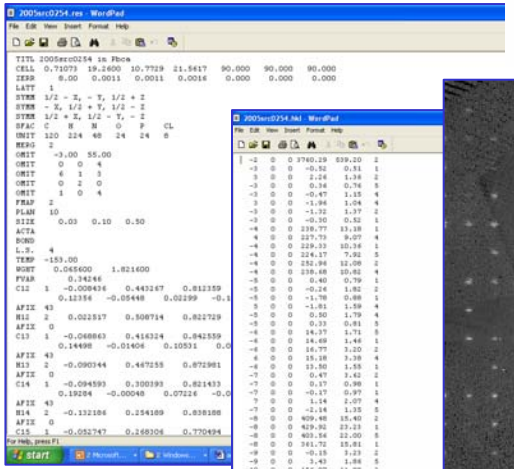
Underlying data (Institutional data repository)

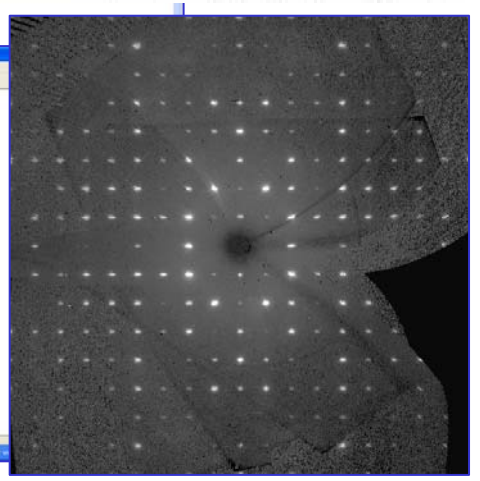


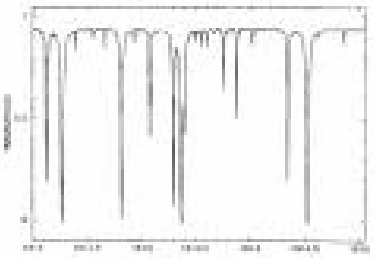




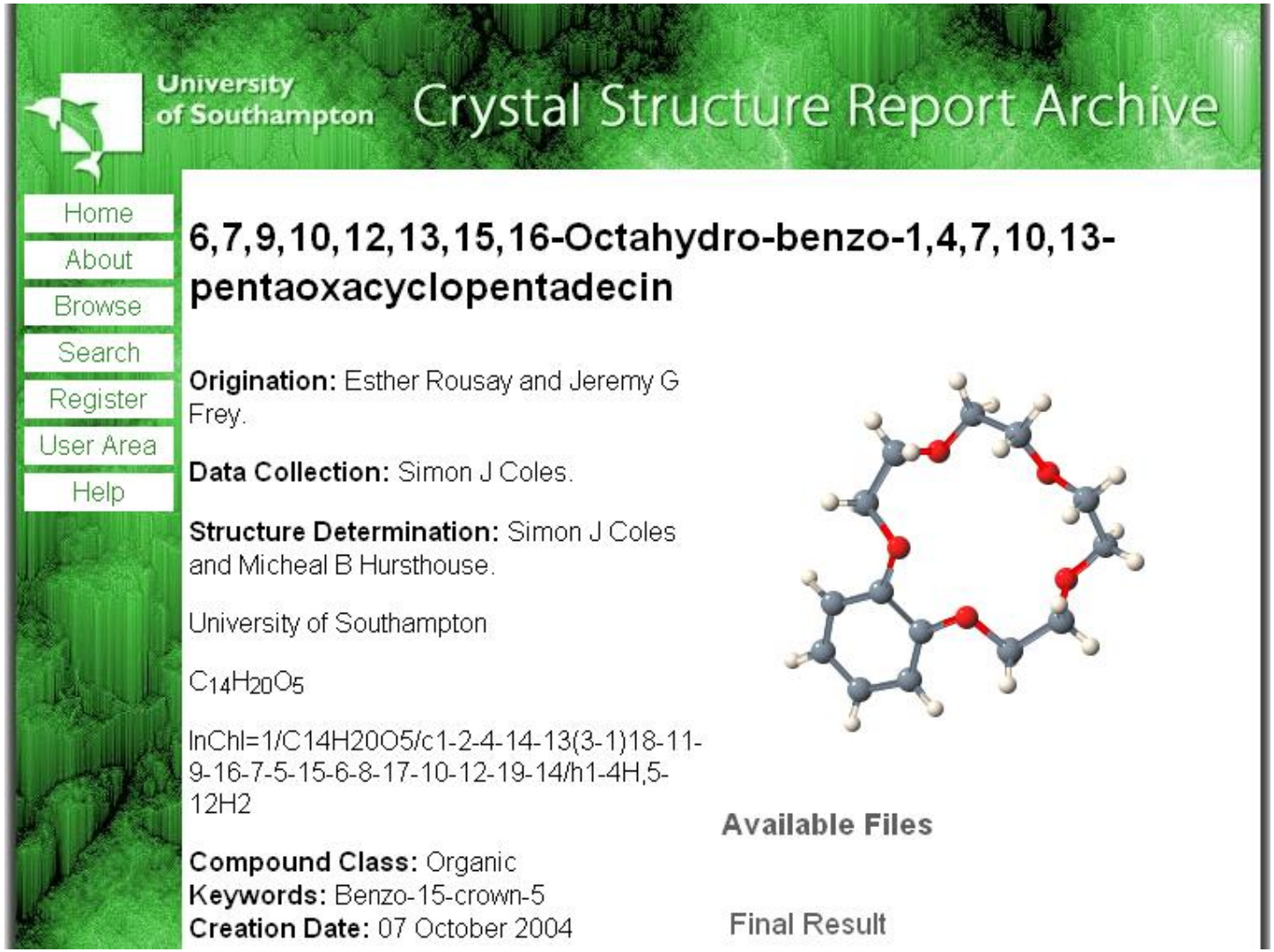








Laboratory Data Management and Archive



The eCrystals Public Data Archive

University of Southampton Crystal Structure Report Archive

Home
About
Browse
Search
Register
User Area
Help

6,7,9,10,12,13,15,16-Octahydro-benzo-1,4,7,10,13-pentaoxacyclopentadecin

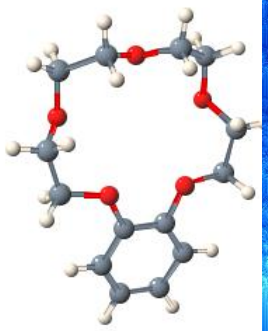
Simon J Coles, Michael B Hursthouse, Jeremy G Frey and Esther Rousay.

University of Southampton

$C_{14}H_{20}O_5$

InChI=1/C14H20O5/c1-2-4-14-13(3-1)18-11-9-16-7-5-15-6-8-17-10-12-19-14/h1-4H,5-12H2

DOI: 10.594/ecrystals.chem.soton.ac.uk/145
Compound Class: Organic
Keywords: crown ethers
Creation Date: 07 October 2004
Deposited By: A.N. Admin
Deposited On: 20 February 2006



Available Files

Depositor Comments

Structure already known, but accurately redetermined for a local research project.

Data collection parameters

Chemical formula	C14 H20 O5
Crystallisation Solvent	
Crystal morphology	Plate
Crystal system	Orthorhombic
Space group symbol	Pbca
Cell length a	16.4963(18)
Cell length b	8.325(3)
Cell length c	20.061(6)
Cell angle alpha	90.00
Cell angle beta	90.00
Cell angle gamma	90.00
Data collection temperature	120(2)

Refinement results

Solution figure of merit	0.0409
R Factor (Obs)	0.0487
R Factor (All)	0.0977
Weighted R Factor (Obs)	0.1008
Weighted R Factor (All)	0.1192

Citation: Coles, S.J., Hursthouse, M.B., Frey, J.G. and Rousay, E. (2004), Southampton, UK, University of Southampton, Crystal Structure Report Archive. (doi:10.1594/ecrystals.chem.soton.ac.uk/145)

Final Result

04sjc0831.cif	13k
04sjc0831.cml	6k

Validation

04sjc0831_checkcif.htm	7k

Refinement

04sjc0831.res	6k
04sjc0831_xl.lst	34k

Solution

04sjc0831.prp	6k
04sjc0831_xs.lst	39k

Processing

04sjc0831.hkl	702k
04sjc0831.htm	10k
04sjc0831_0kl.jpg	57k
04sjc0831_h0l.jpg	85k
04sjc0831_hk0.jpg	88k

Data Collection

04sjc0831_crystal.jpg	17k

Other Files

04sjc0831.doc	78k
04sjc0831.fcf.txt	155k

http://ecrystals.chem.soton.ac.uk

eCrystals 'Global Federation' Model

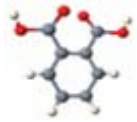
intute Google ChemRefer™

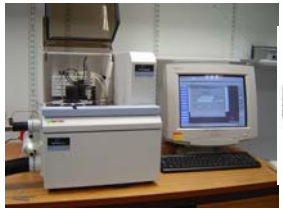
Data discovery, linking, citation

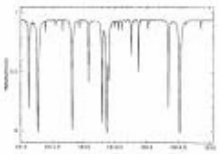
Presentation services / portals

Data discovery, linking, citation

Data creation & capture in "Smart lab"







Chemical Database Service

Aggregator services

IUCr

eBank UK

Chemistry Central

Publishers: peer-review journals, conference proceedings, etc

IUCr

RSC

Search, harvest

Search, harvest

blogs@ChemTools

Deposit

Data analysis

Publication

Laboratory repository

Institutional data repositories

Validation

Search, harvest

Subject Repository

IUCr

Deposit

R4L Repository for the Laboratory

Deposit , Validation

Deposit

DCC

CCLRC

SPECTRa

University of Southampton Crystal Structure Report Archive

Curation

Preservation Deposit

Institution Library & Information Services

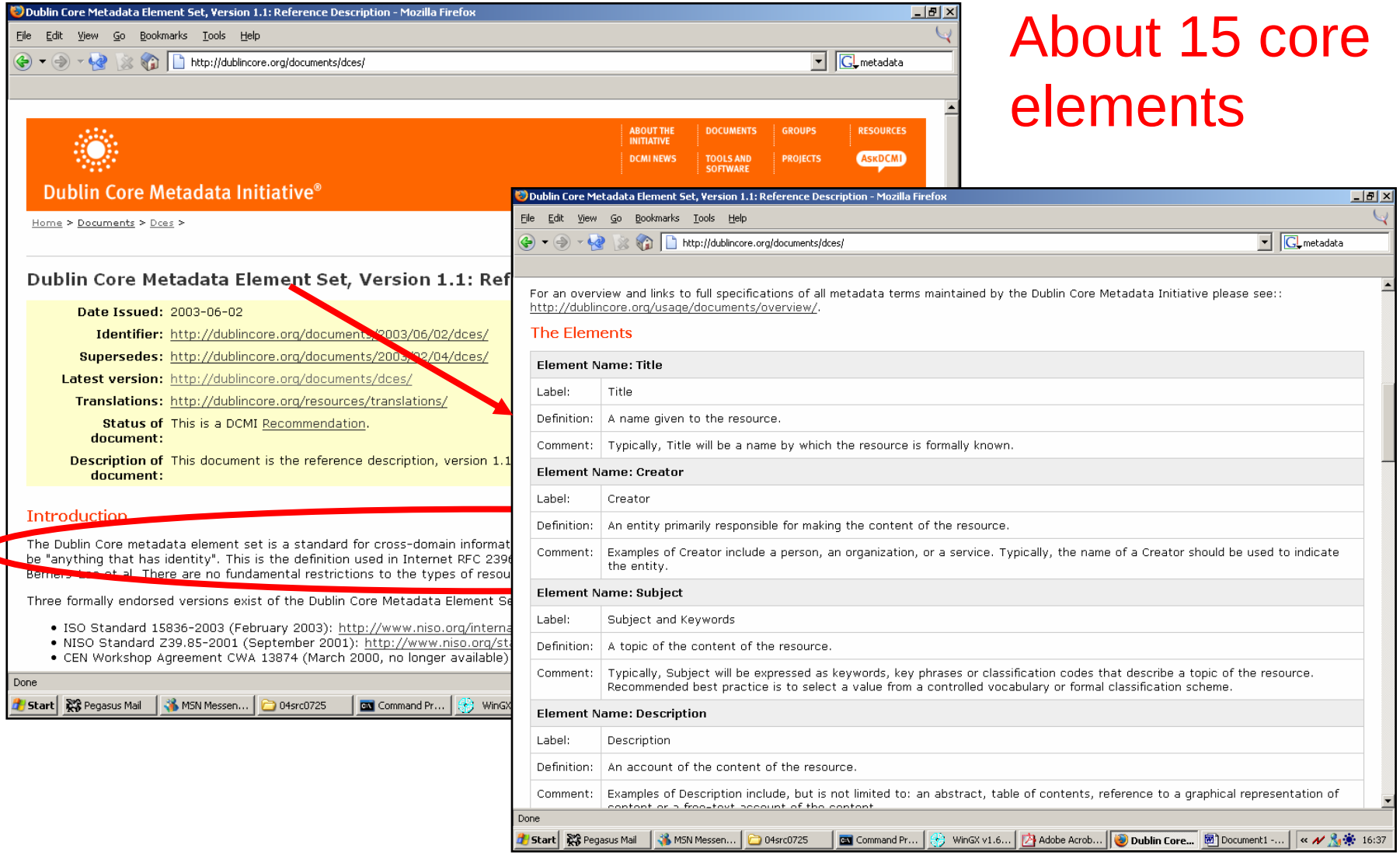
CCLRC

Reciprocal Net

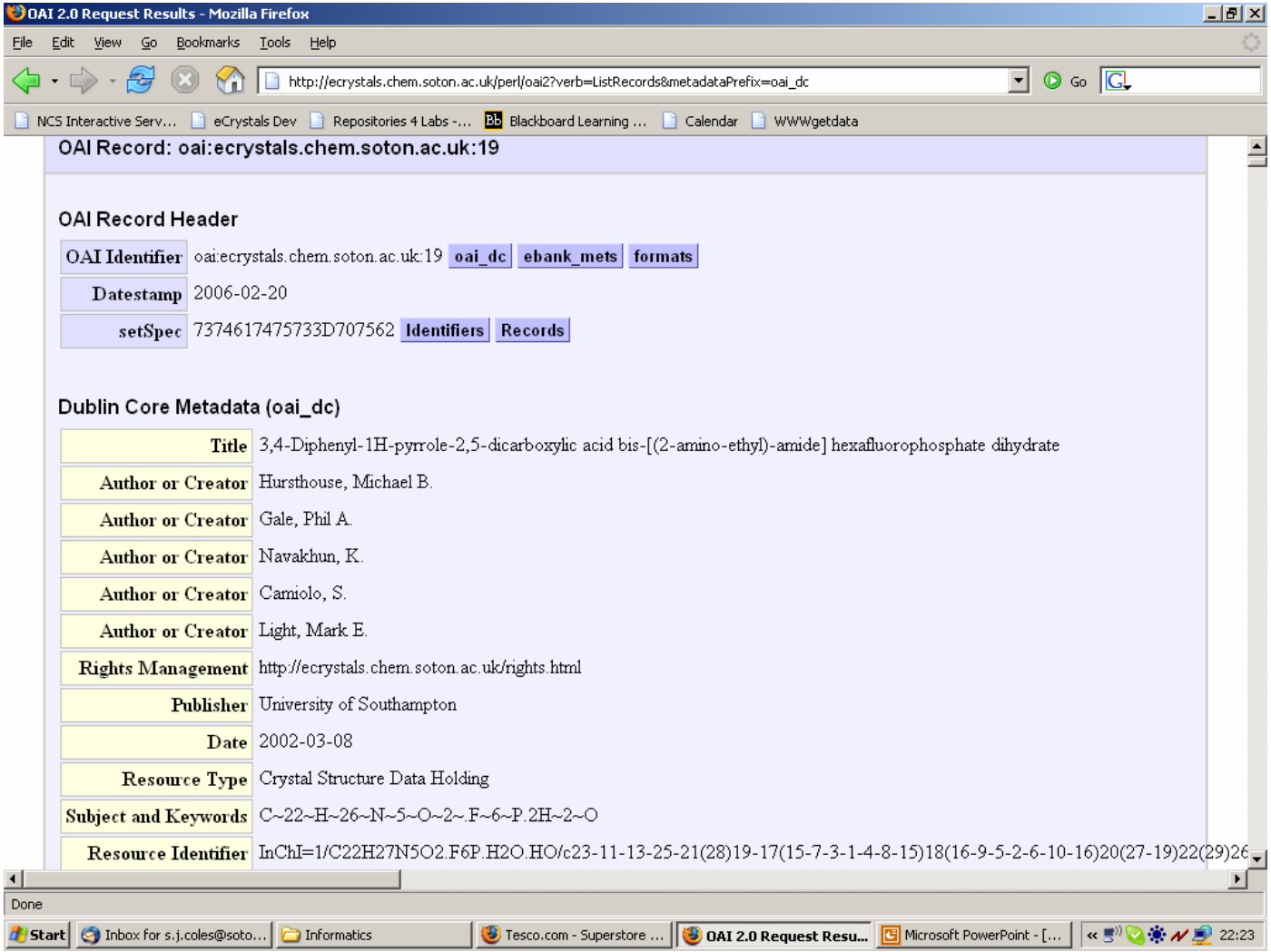
The University of Sydney

Metadata standards: Dublin Core

About 15 core elements



Metadata Publication



ecrystals.chem.soton.ac.uk/perl/oai2

Metadata Publication

- Using simple Dublin Core
    - Crystal structure
    - Title (Systematic IUPAC Name)
    - Authors
    - Affiliation
    - Creation Date
- Additional chemical information through Qualified Dublin Core
    - Empirical formula
    - International Chemical Identifier (InChI)
    - Compound Class & Keywords
- Specifies which ‘datasets’ are present in an entry
- DOI *http://dx.doi.org/10.1594/ecrystals.chem.soton.ac.uk/145*
- Rights & Citation *http://ecrystals.chem.soton.ac.uk/rights.html*
- Application Profile *http://www.ukoln.ac.uk/projects/ebank-uk/schemas/*

The New World? Describing and Exchanging Experimental Processes

We are currently seeking further funding and developing collaborations to build myExperiment.

myexperiment beta

Home About Get Involved! Source Code

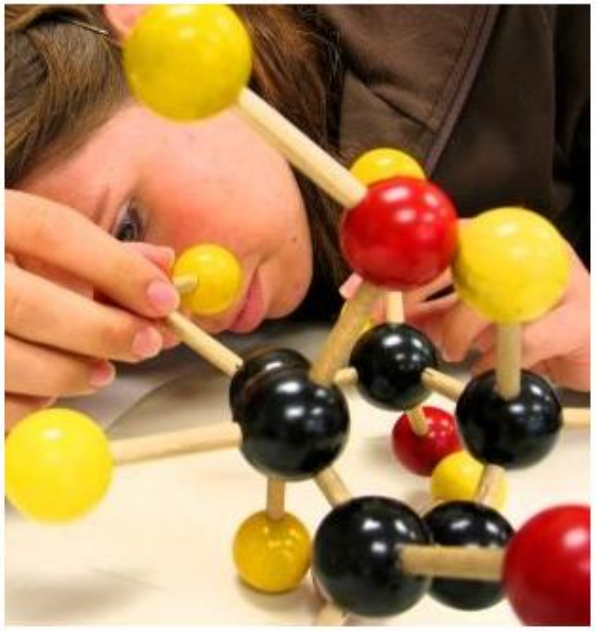
myExperiment makes it really easy for people to share experiments and discuss them. You will be able to swap workflows and publications as well as documents, photos and videos, all from any web browser.

We're not ready to launch yet, but if you want to know when we're ready, you can join our mailing list

THE BUZZ

"Their kids may have got there first but scientists will soon have their very own version of MySpace, where they will be able to share preliminary results, ideas and research tools." — New Scientist Tech, October 2006.

Currently, a lightweight repository of workflows and the Taverna BioService Finder are available.



SEE ALSO

- About myExperiment
- Case Studies

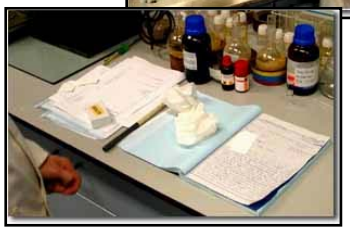
GET INVOLVED!

myExperiment is a free software project. You can help us by testing code, writing patches and writing

BLOG

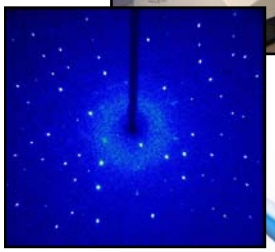
The Need for Experimental Metadata

Synthesis

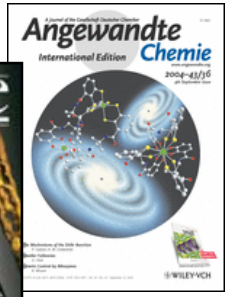


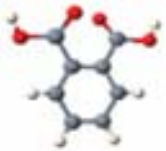
Preparation

Data Collection

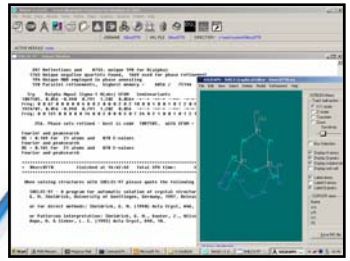


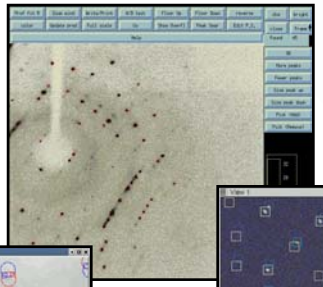
Publication

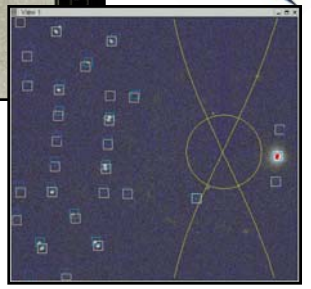




Structure Solution







Data Processing

Why do we need metadata for crystallographic raw data?

1. Preservation and curation.
   Description of file types, software (versions).
2. Collaboration and service provision.
   What level of metadata is necessary (if any)?
3. Discovery.
   Is eCrystals level sufficient?
4. Validation and reuse.
   Description of synthesis, sample preparation, data collection process, file types, processing software and methodology.

Changing Times!

- eCrystals Federation requirements capture (report Jun 2007)
- Robust software & transition to ePrints3 (Summer 2007)
- Capability for ‘site configuration’ as defined by laboratory working practices (Current)
- Roll out to other laboratories (Summer 2007)
- Collaboration with Sun to link eCrystals to terabyte store for archival and access to raw data (Summer 2007)
- Improve linkage between repository and:
    - desktop /
    - data management systems
    - workflow enactors