Electronics and Computer Science

JISC

Archiving research data and research publications.

Dr Leslie Carr, Intelligence, Agents Multimedia, University of Southampton

Dr Simon Coles, School of Chemistry, University of Southampton

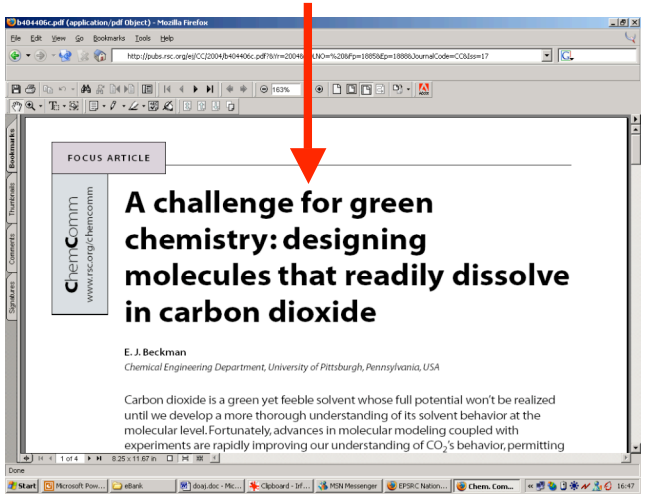
Dr Liz Lyon, UKOLN, University of Bath

Overview

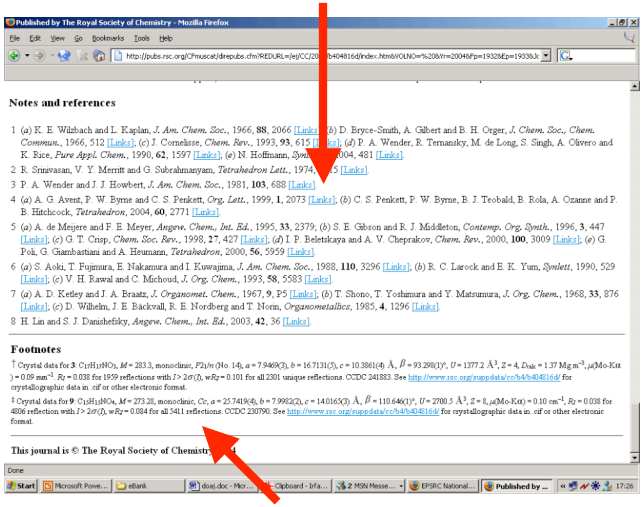
- In an Open Access environment
  - scientific outputs are openly available
  - described by appropriate metadata
  - in Institutional Repositories
  - harvestable by OAI protocols
- Scientists can use the same infrastructure
  - (here eprints.org software and an existing scientific portal service)
  - to provide maximal open access
  - to all their data, as well as their published articles
    - raw data, intermediate calculations, final results
    - in a searchable, accessible form
- BUT this is subject to ongoing investigation.

Current chemistry publishing protocols

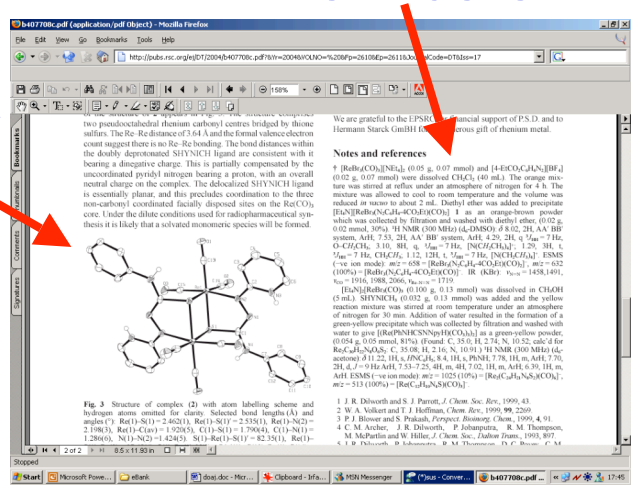
Ideas and interpretations



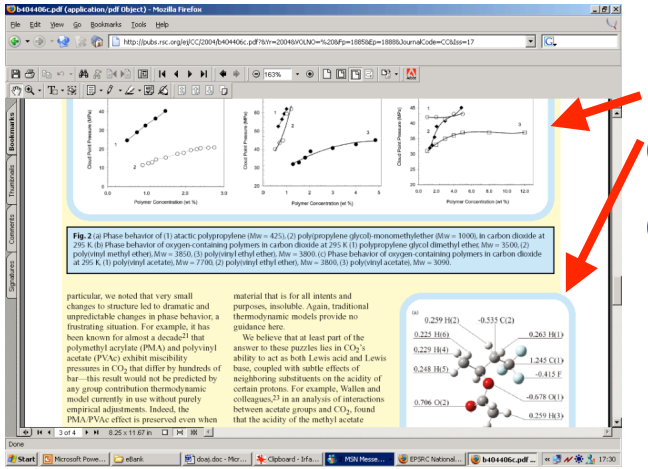
Hooks into the literature

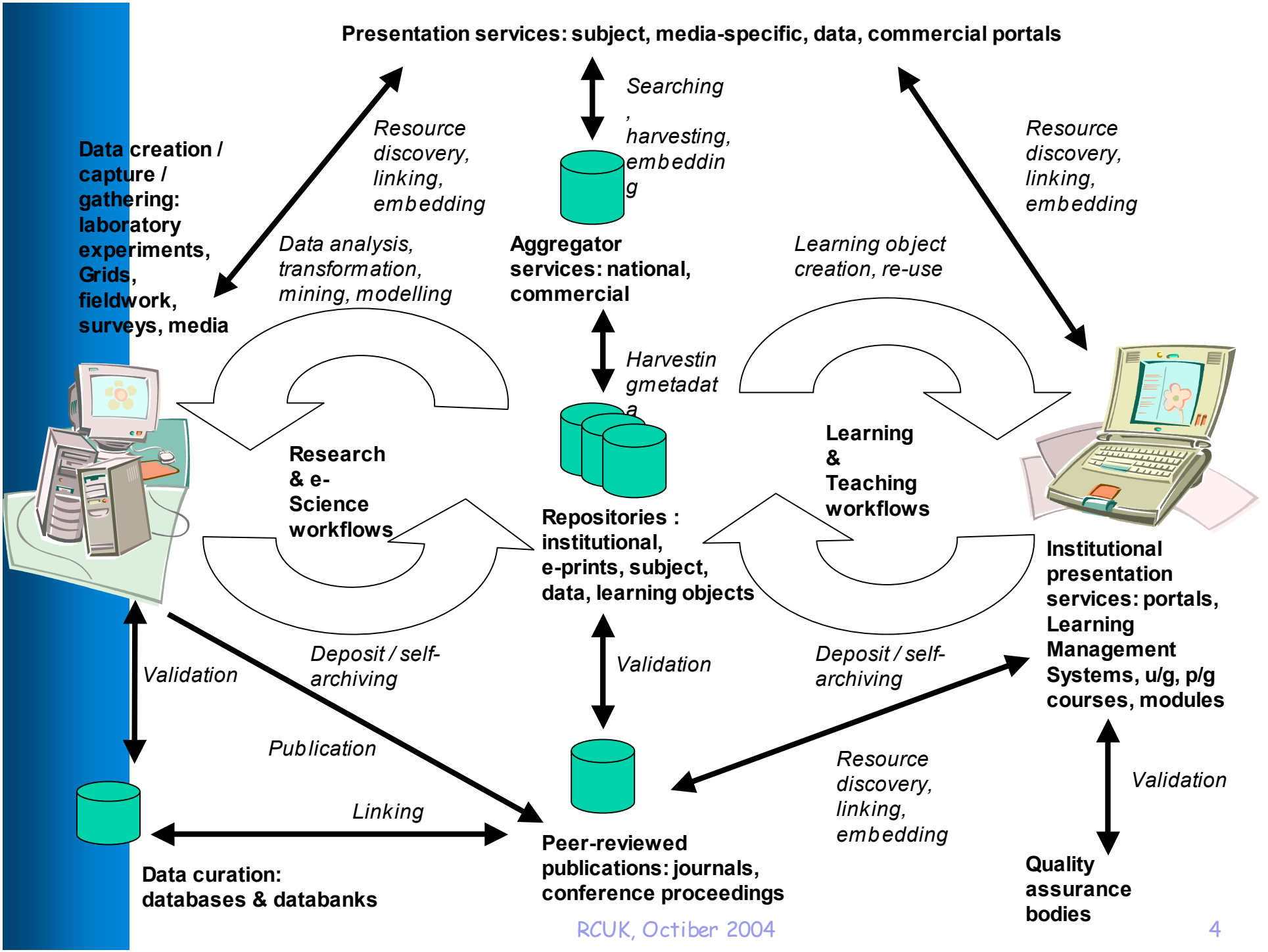


Raw data!

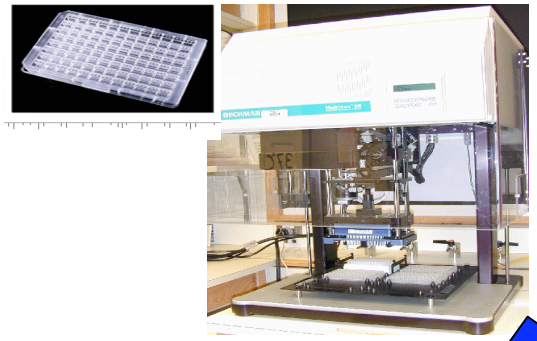


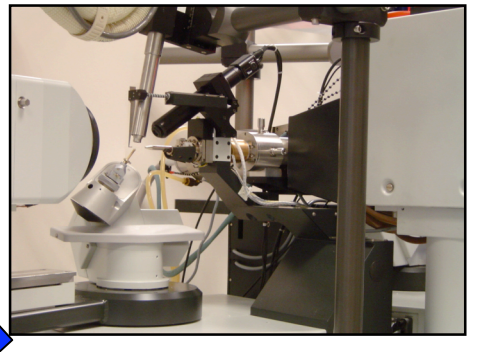
Results & derived data




Presentation services: subject, media-specific, data, commercial portals
Searching, harvesting, embeddin g
Resource discovery, linking, embedding
Resource discovery, linking, embedding
Data creation / capture / gathering: laboratory experiments, Grids, fieldwork, surveys, media
Data analysis, transformation, mining, modelling
Aggregator services: national, commercial
Learning object creation, re-use
Harvestin gmetadat a
Research & e-Science workflows
Repositories : institutional, e-prints, subject, data, learning objects
Learning & Teaching workflows
Institutional presentation services: portals, Learning Management Systems, u/g, p/g courses, modules
Validation
Deposit / self-archiving
Validation
Deposit / self-archiving
Publication
Linking
Resource discovery, linking, embedding
Validation
Data curation: databases & databanks
Peer-reviewed publications: journals, conference proceedings
Quality assurance bodies

The data deluge

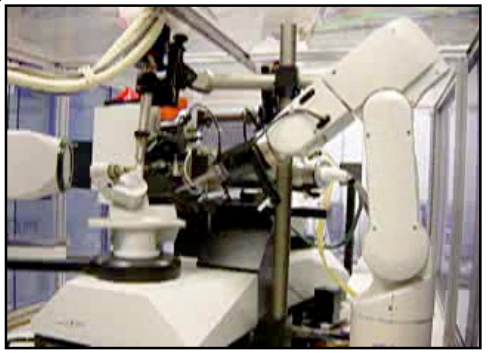




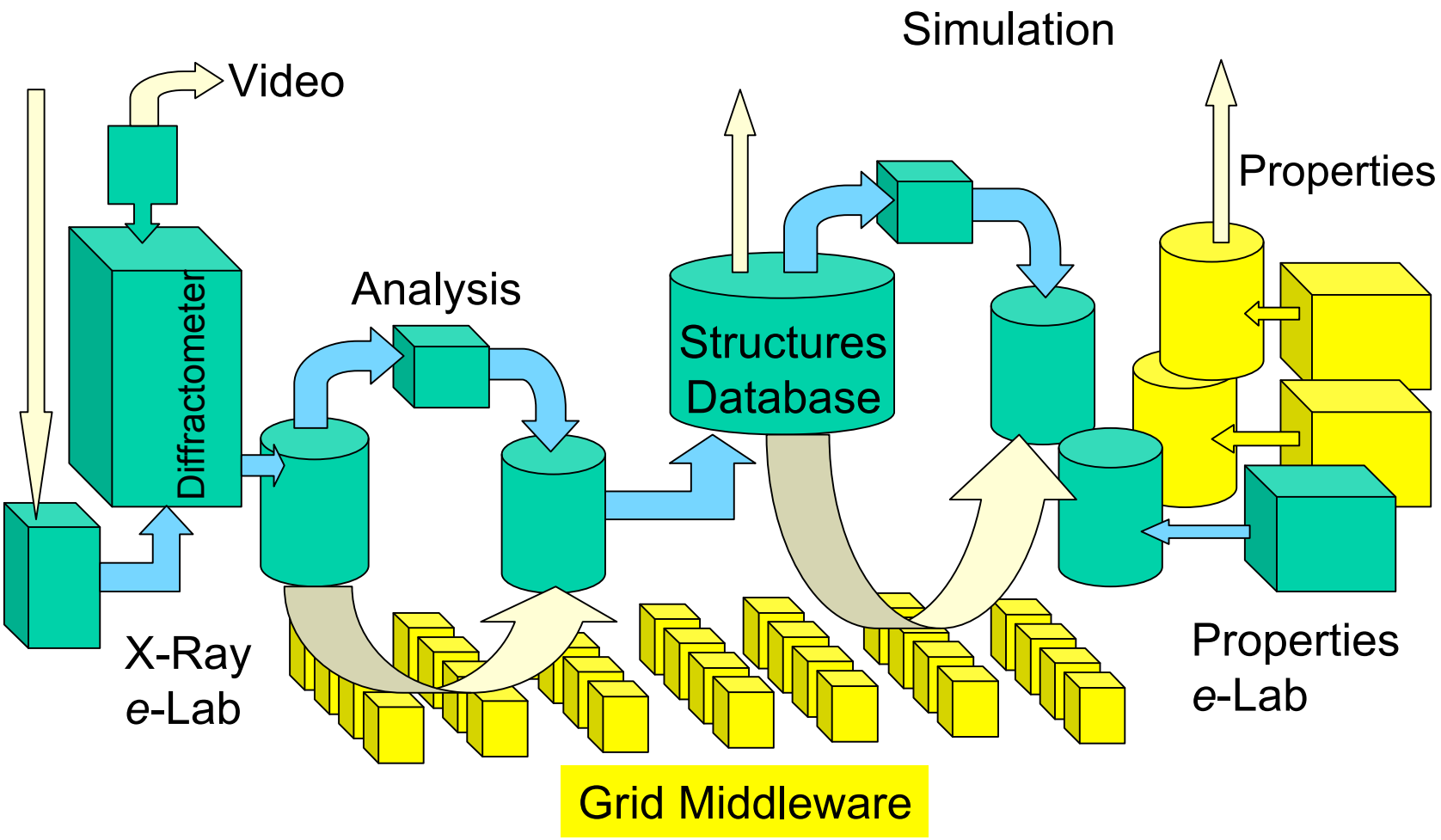
Data Overload!

EPSRC National Crystallography Service

How do we disseminate?

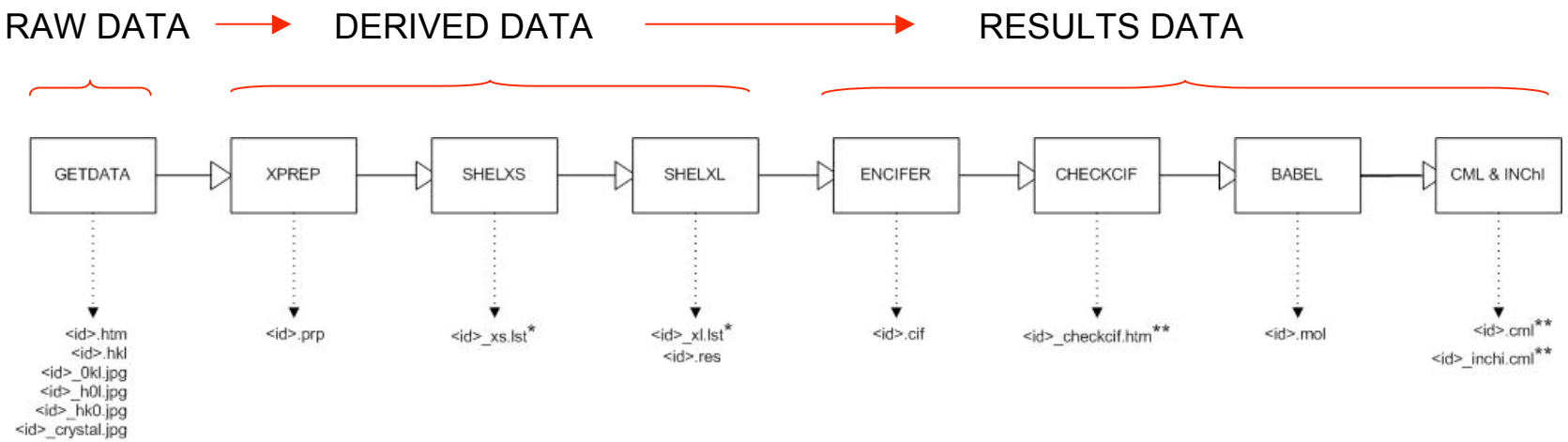


CombeChem: An EPSRC pilot project

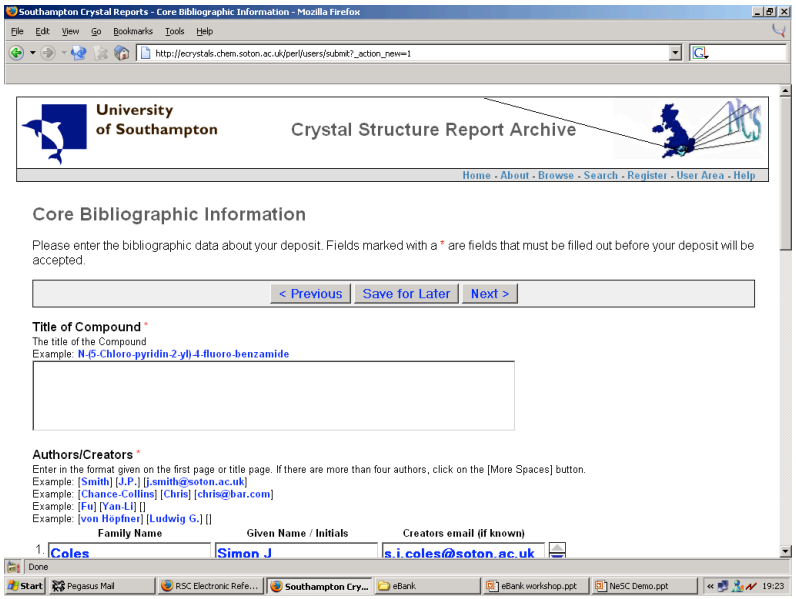


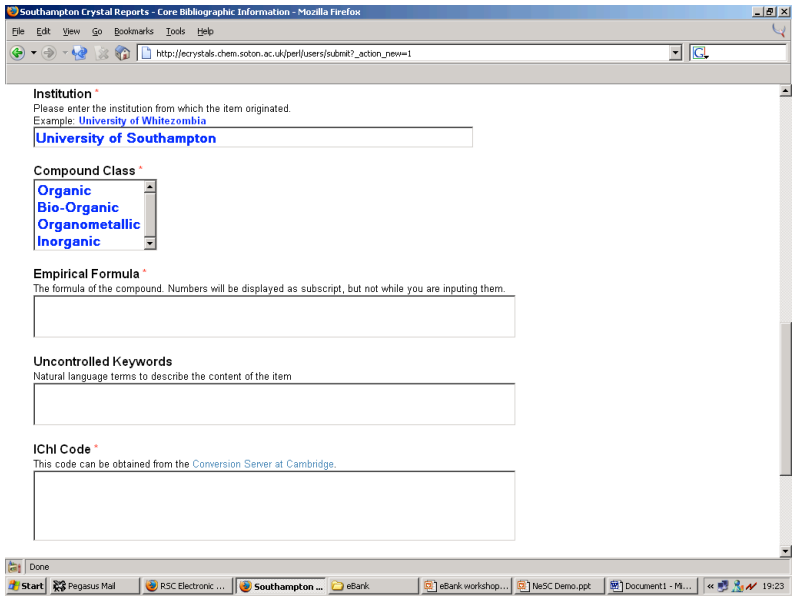
Crystallography workflow

- **Initialisation**: mount new sample on diffractometer & set up data collection
- **Collection**: collect data
- **Processing**: process and correct images
- **Solution**: solve structures
- **Refinement**: refine structure
- **CIF**: produce CIF (Crystallographic Information File format)
- **Report**: generate Crystal Structure Report



Deposition into the archive





An Archive entry

EBank Southampton - 2-(N-Ferrocenylmethylcarbamoyl)-5-(methoxycarbonyl)-3,4-diphenylpyrrole - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://eprints.ebank.ecs.soton.ac.uk/archive/00000025/

University of Southampton

Crystal Structure Report Archive

Home - About - Browse - Search - Register - User Area - Help

2-(N-Ferrocenylmethylcarbamoyl)-5-(methoxycarbonyl)-3,4-diphenylpyrrole

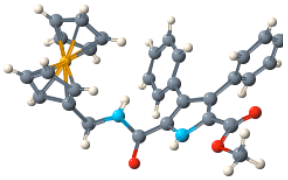
Susanne L. Huth, Michael B. Hursthouse, Simon J. Coles, Mark E. Light, Peter N. Horton, Phil A. Gale, G. Denuault and C. N. Warriner.

University of Southampton

$C_{30}H_{26}FeN_2O_3$

CCDC Code: XUZSUG
IChI Code: C30H26FeN2O3,1H3-36-20(35)18-16(14-10H-6H-3H-7H-11H-14)15(13-8H-4H-2H-5H-9H-13)17(33H-18)19(34)32H-12H2-30-28H-26H-27H-29H(30)31(26,27,28,30)21H-22H(31)24H(31)25H(31)23H(21)31
Compound Class: Organometallic
Keywords: Supramolecular Chemistry
Creation Date: 31 May 2002
Deposited By: Susanne L. Huth
Deposited On: 10 May 2004

Data collection parameters



Applet jmol started

Start | Pegasus Mail | WinGX v1.6... | ISSC-XIII - M... | EBank Sout... | Windows Mes... | eBank ISSC a... | ECM22EPrint...

ecrystals.chem.soton.ac.uk

Data collection parameters

Chemical formula	C30 H26 Fe N2 O3
Crystallisation Solvent	
Crystal morphology	
Crystal system	Orthorhombic
Space group symbol	Pbca
Cell length a	6.0816(4)
Cell length b	24.8503(16)
Cell length c	31.120(3)
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	
R Factor (Obs)	0.0573
R Factor (All)	0.1185
Weighted R Factor (Obs)	0.1046
Weighted R Factor (All)	0.1243

Available Files

Final Result

02sot064.CIF	19k
02sot064.cml	8k
02sot064_checkcif.html	14k

Refinement

02sot064.RES	9k

Solution

02sot064.PRP	5k

Processing

02SOT064.HTM	6k
02sot064.HKL	338k

Other Files

02sot064.DOC	113k
02sot064.LST	49k

All the way back to the underlying data…

Available Files

Data collection parameters

Chemical formula	C30 H26 Fe N2 O3
Crystallisation Solvent	
Crystal morphology	
Crystal system	Orthorhombic
Space group symbol	Pbca
Cell length a	6.0816(4)
Cell length b	24.8503(16)
Cell length c	31.120(3)
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	
R Factor (Obs)	0.0573
R Factor (All)	0.1185
Weighted R Factor (Obs)	0.1046
Weighted R Factor (All)	0.1243

Final Result

02sot064.CIF	19k
02sot064.cml	8k
02sot064_checkcif.html	14k

Refinement

02sot064.RES	9k

Solution

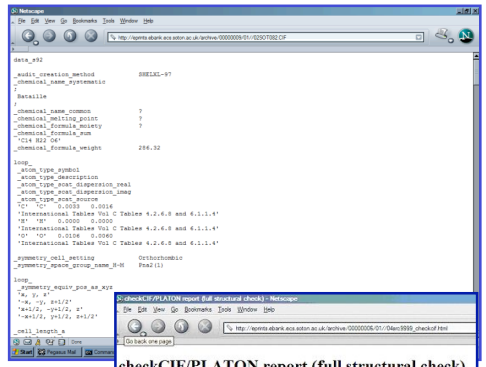
02sot064.PRP	5k

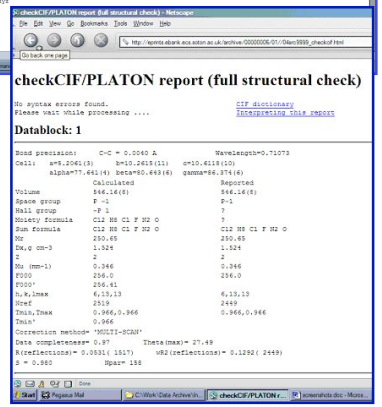
Processing

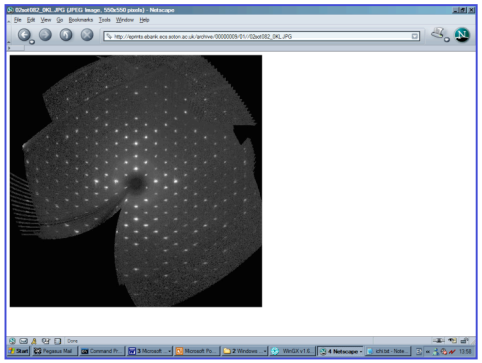
02SOT064.HTM	6k
02sot064.HKL	338k

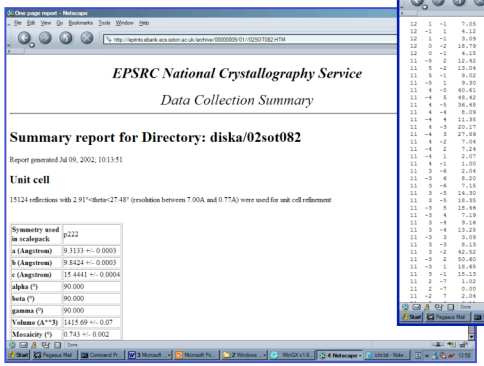
Other Files

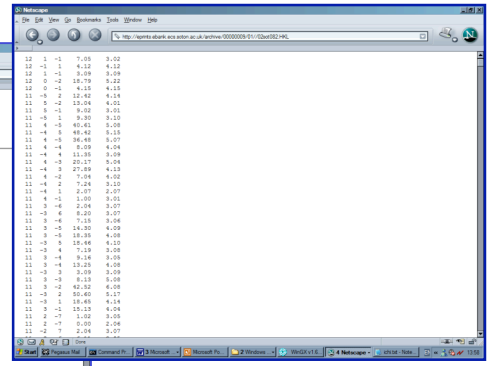
02sot064.DOC	113k
02sot064.LST	49k



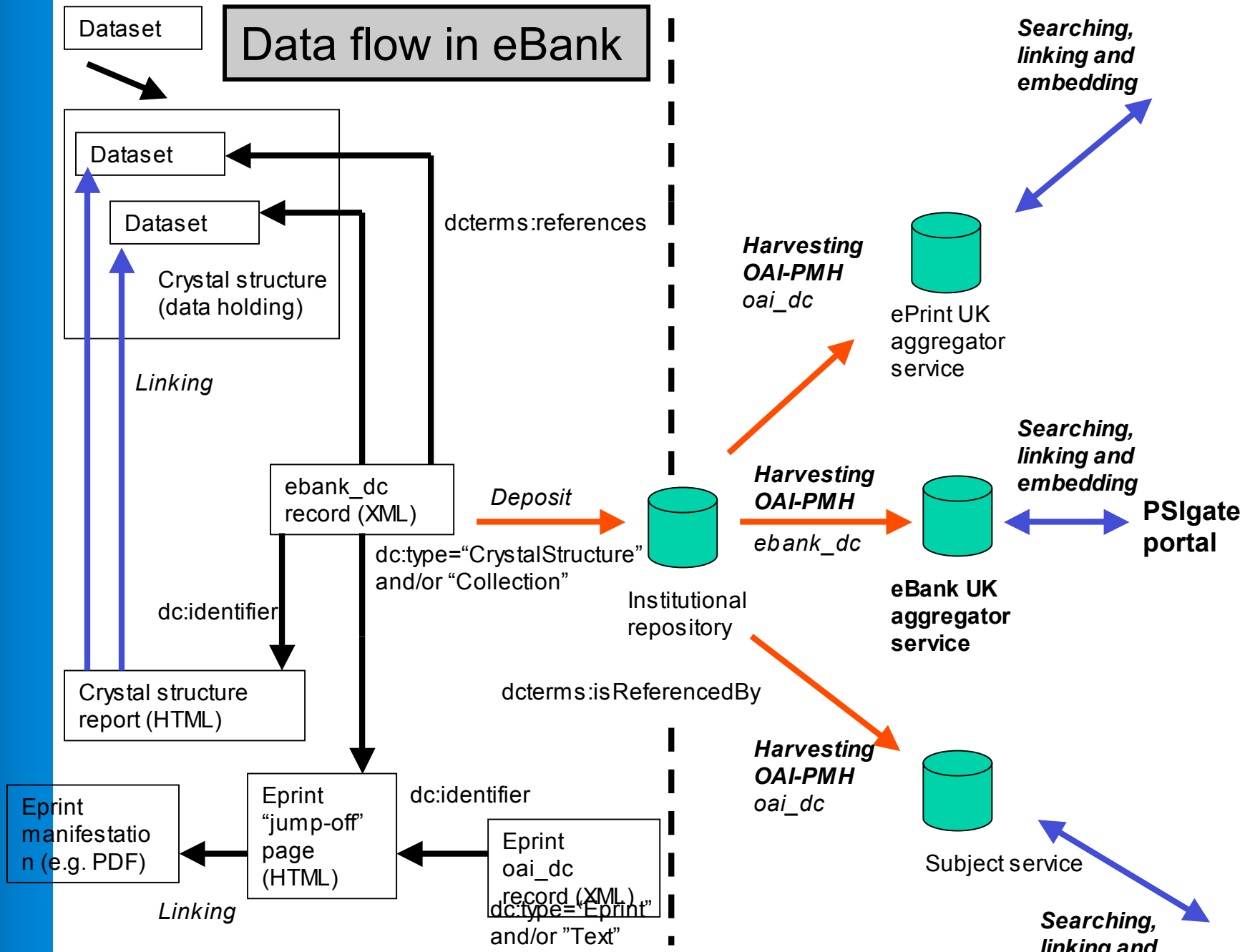






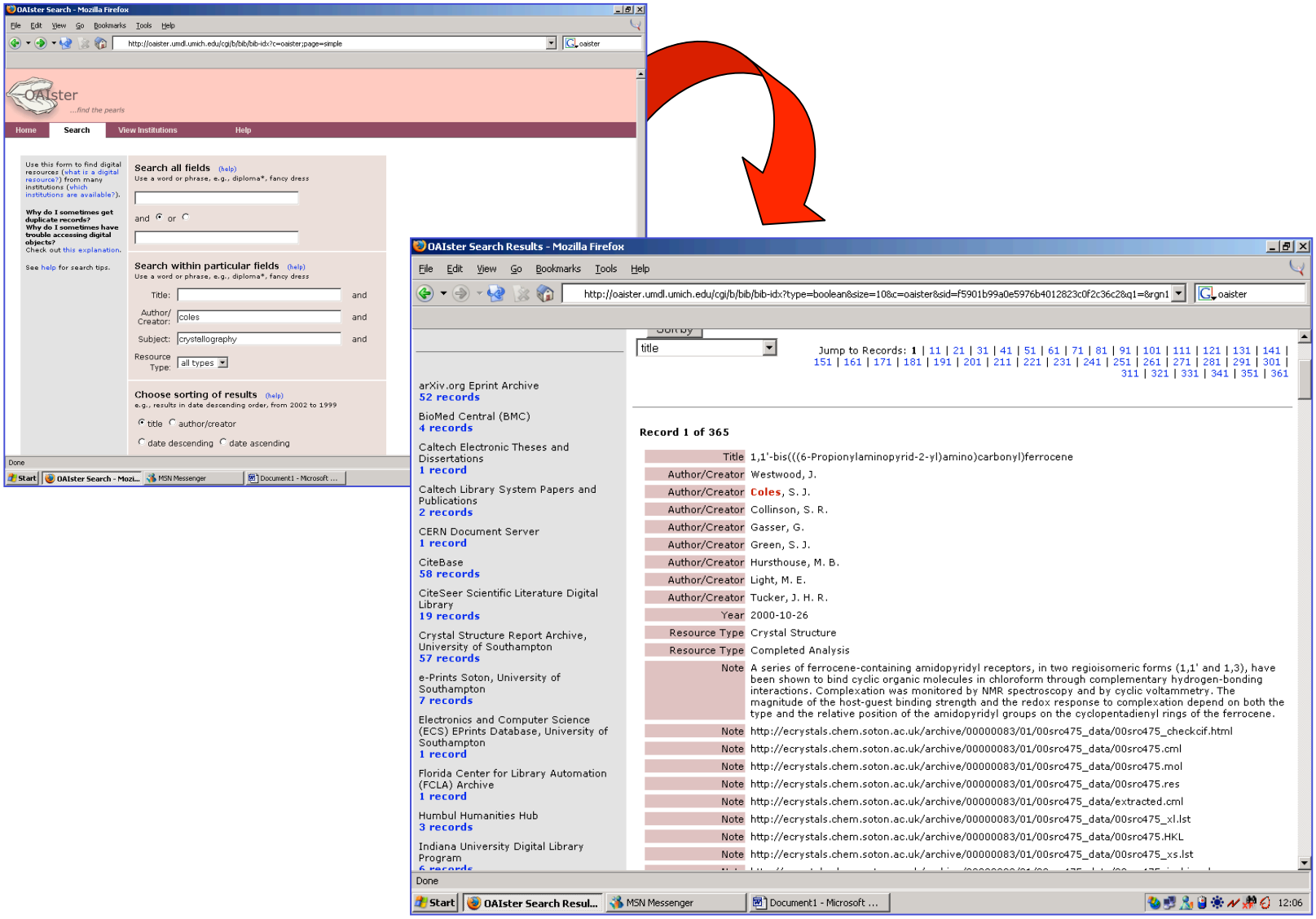


Data flow in eBank

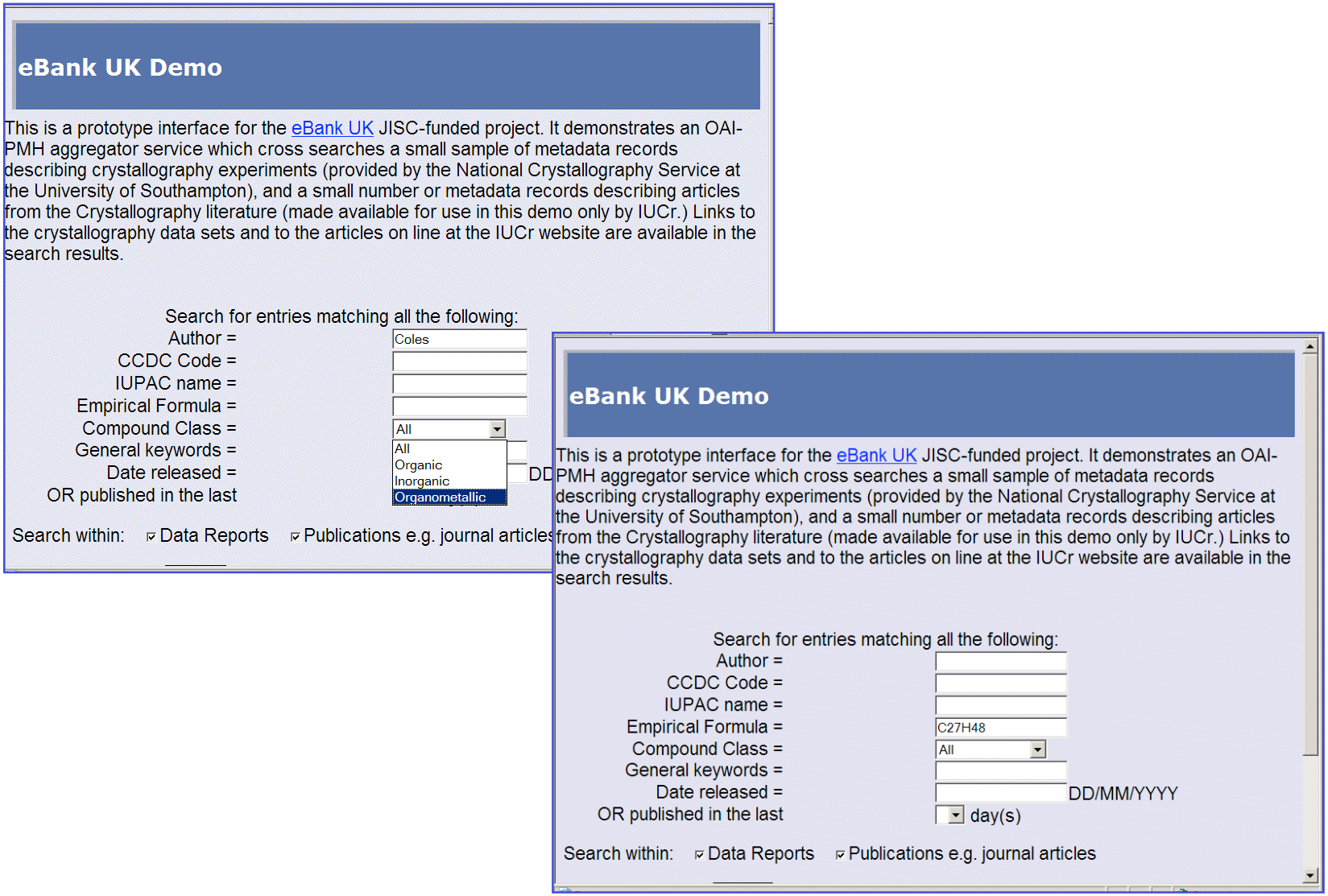


Model input Andy Powell, UKOLN.

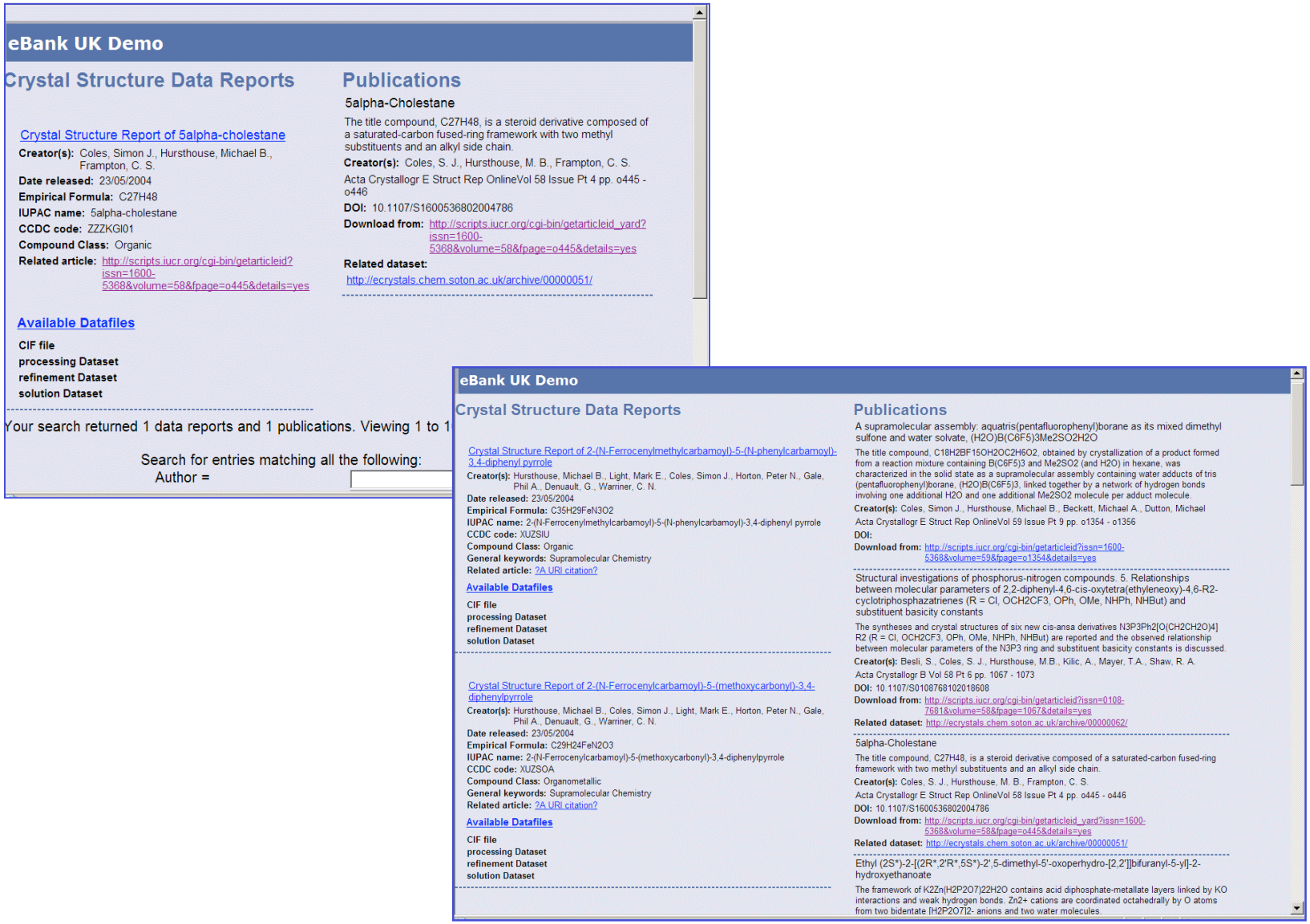
Harvesting: OAIster



Linking and aggregating: Search & discover



Linking and aggregating: Hit browsing



And finally…
eBank embedded in a science portal