UNIVERSITY OF Southampton
School of Chemistry

eCrystals

R4L Repository for the Laboratory

New Web-based Approaches to Communicating and Teaching in Chemistry & Crystallography

Simon Coles
EPSRC National Crystallography Service
School of Chemistry
University of Southampton

@malaria

blogs@ChemTools
The thoughts of Chemists

my experiment beta

“The internet wasn't created for mockery! It was created so scientists from different universities could share datasets....”

Simpson, H. *The Simpsons* (2005), Eds. Groening, M., Brooks, J.L. & Simon, S., Series 16, Episode 8, Original air date (US) 06-Feb-2005.

http://www.tvtome.com/tvtome/servlet/GuidePageServlet/showid-146/epid-346864/

Social Networks for Chemists

Google generation: new behaviour and approach

Sharing Rich Media

- Video + Paper = Pubcast

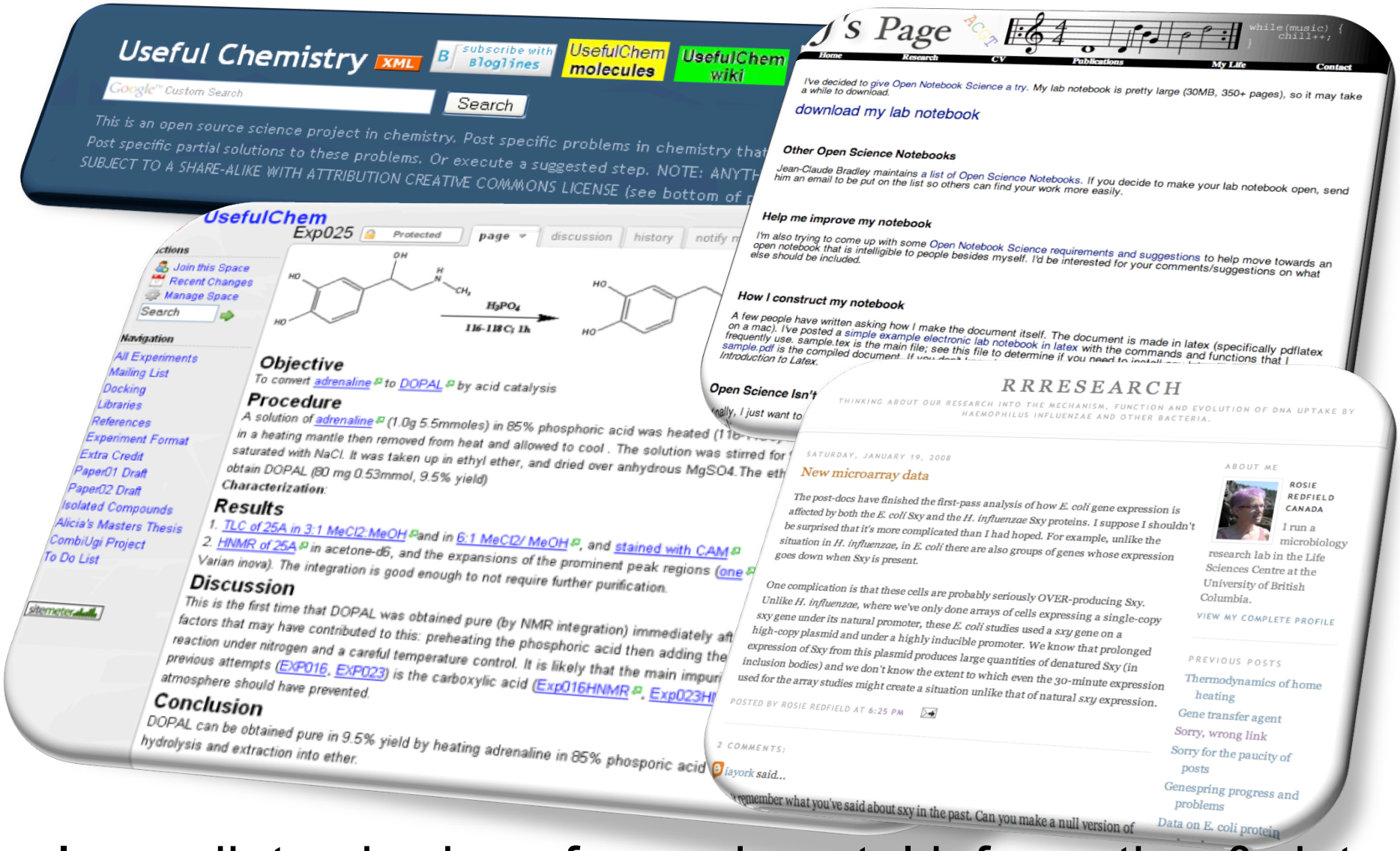
New Approaches to ‘Sharing Experiments’

- Specialised domain-oriented innovations

Formation of Open Communities

- New approaches surfacing and growing FAST

Open Notebook Science



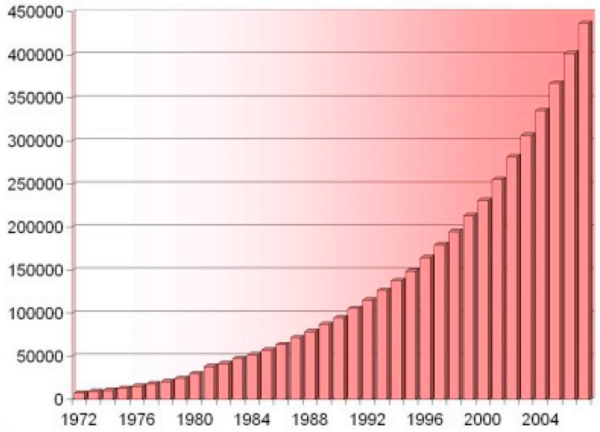
- Immediate sharing of experimental information & data

New Information Exchange Environments

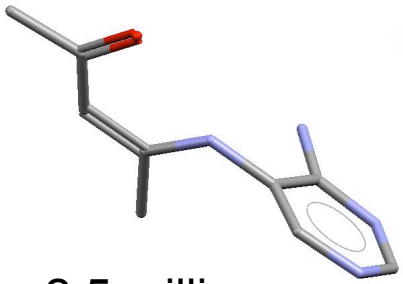
- Immersive alternative to conventional browsing & interaction

Data Deluge

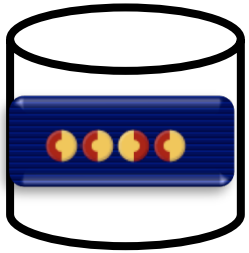
• 40 years ago a PhD student would determine about 3 crystal structures for their thesis – this can now be easily achieved in a day



35 million



2.5 million



0.5 million

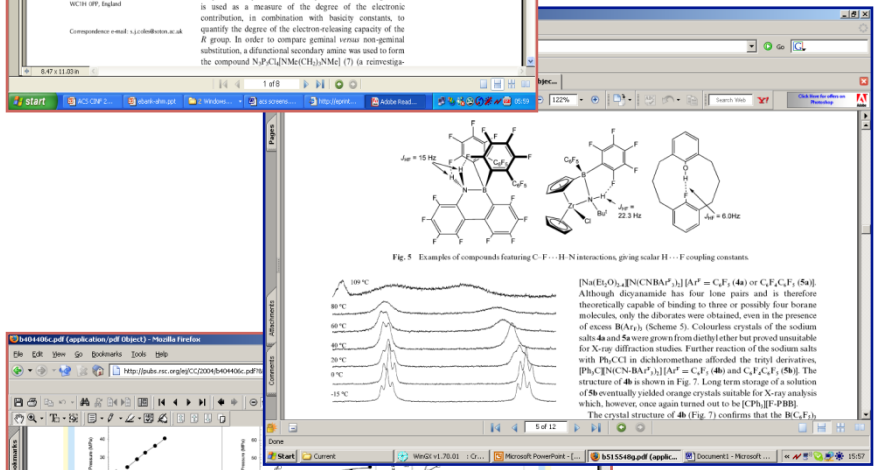
'Few thousand'

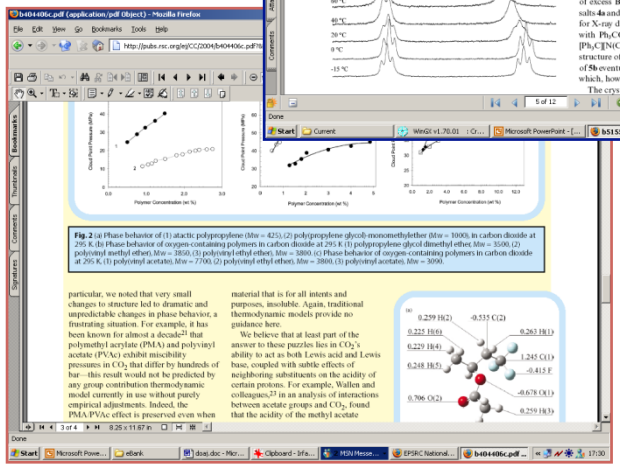
The primary cause is the current data publication process, which is tied to journal articles and peer review

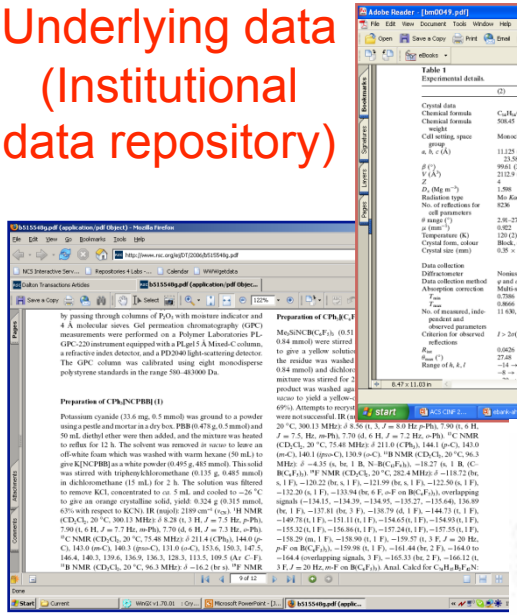
The Solution

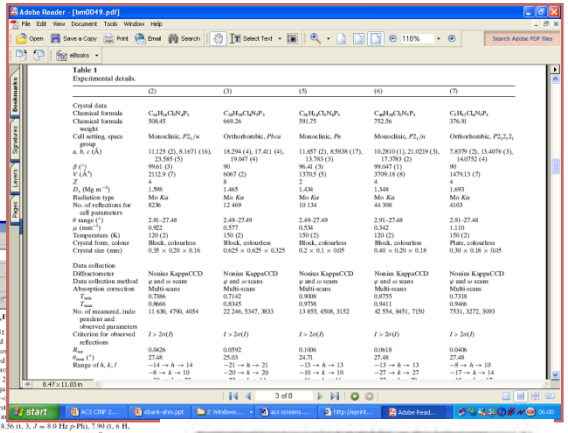
Intellect & Interpretation (Journal article, report, etc)

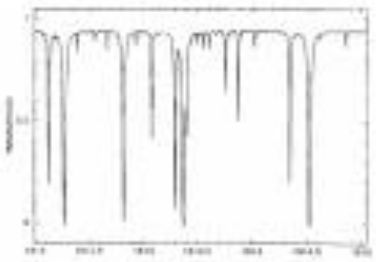
Underlying data (Institutional data repository)

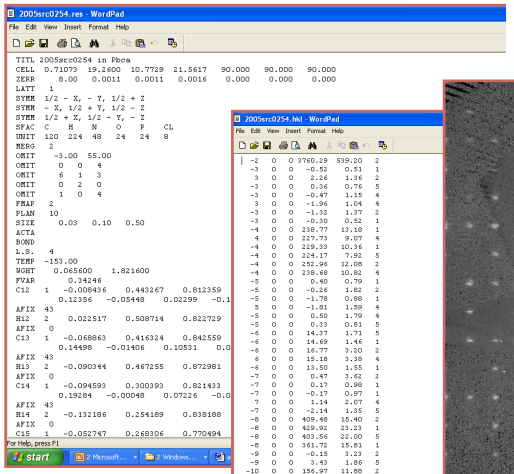


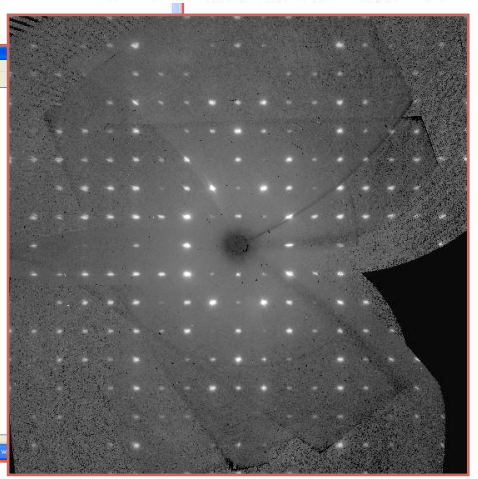












The eCrystals Data Repository

eCrystals — University of Southampton

Home | About | Browse by Year | Browse by People

Logged in as Dr Richard A Stephenson | Manage deposits | Profile | Saved searches | Review | Admin | Logout | Search

2,2-trimethylenedioxy-4,4,6,6-tetrachlorocyclotriphosphazene

Sample Originator: A. Kilic[a], M. Odlyha[b], A. Uslu[a], David B. Davies[b] and R.A. Shaw[b].

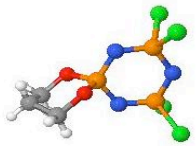
Data Collection: Mark E. Light[c], Simon J. Coles[c] and Susanne. L. Huth[c]

Structure Determination: Simon J. Coles[c], Michael B. Hursthouse and J.S. Rutherford.

Gebze Institute of Technology[a]
Birkbeck College[b]
University of Southampton[c]

$C_3H_6Cl_4N_3O_2P_3$

InChI=1/C3H12Cl4N3O2P3/c4-13(5)8-14(6,7)10-15(9-13)11-2-1-3-12-15/h8-10,13-15H,1-3H2



Jmol

Identification Number:	10.3737/ecrystals.chem.soton.ac.uk/300
Controlled Keywords:	cyclophosphazene, phase transition, variable temperature
Date Created:	28 March 2007
Deposited On:	*21 Apr 2008 15:56*
Deposited By:	Dr Simon J Coles

Data collection parameters

Chemical formula	C3 H6 Cl4 N3 O2 P3
Crystallisation Solvent	
Crystal morphology	Rod
Crystal system	Orthorhombic
Space group symbol	Pna2(1)
Cell length a	13.4804(14)
Cell length b	10.6442(9)
Cell length c	8.8479(7)
Cell angle alpha	90.00
Cell angle beta	90.00
Cell angle gamma	90.00
Data collection temperature	274(2)

Refinement results

Solution figure of merit	0.0569
R Factor (Obs)	0.0334
R Factor (All)	0.0380
Weighted R Factor (Obs)	0.0871
Weighted R Factor (All)	0.0905

Citation: Kilic, A. and Odlyha, M. and Uslu, A. and Davies, David B. and Shaw, R.A. and Light, Mark E. and Coles, Simon J. and Huth, Susanne. L. and Hursthouse, Michael B. and Rutherford, J.S. (2007) University of Southampton, Crystal Structure Report Archive. (doi:10.3737/ecrystals.chem.soton.ac.uk/300)
Export as: EndNote BibTeX ASCII Citation

Available Files

Final Result

2005sjc0007.cif	11k
2005sjc0007.cml	4k
2005sjc0007.fcf	138k

Validation

2005sjc0007_checkcif.htm	9k

Refinement

2005sjc0007.res	5k
2005sjc0007_xl.lst	29k

Solution

2005sjc0007.prp	5k
2005sjc0007_xs.lst	44k

Processing

2005sjc0007.hkl	532k
2005sjc0007.htm	11k
2005sjc0007_0kl.jpg	91k
2005sjc0007_h0l.jpg	87k
2005sjc0007_hk0.jpg	79k

Data Collection

2005sjc0007_crystal.jpg	17k

Other Files

2005sjc0007.doc	186k
2005sjc0007.inchi	1k
2005sjc0007.ins	4k
2005sjc0007.mol	2k
2005sjc0007.p4p	1k
2005sjc0007_ellipsoid.gif	21k

Repository Staff Only: item control page

- Quick & simple to deposit
- Software tools
- Laboratory archive
- Community involvement
- ‘Embargo’ facility
- Structured foundations
- Discoverable & harvestable

UKOLN

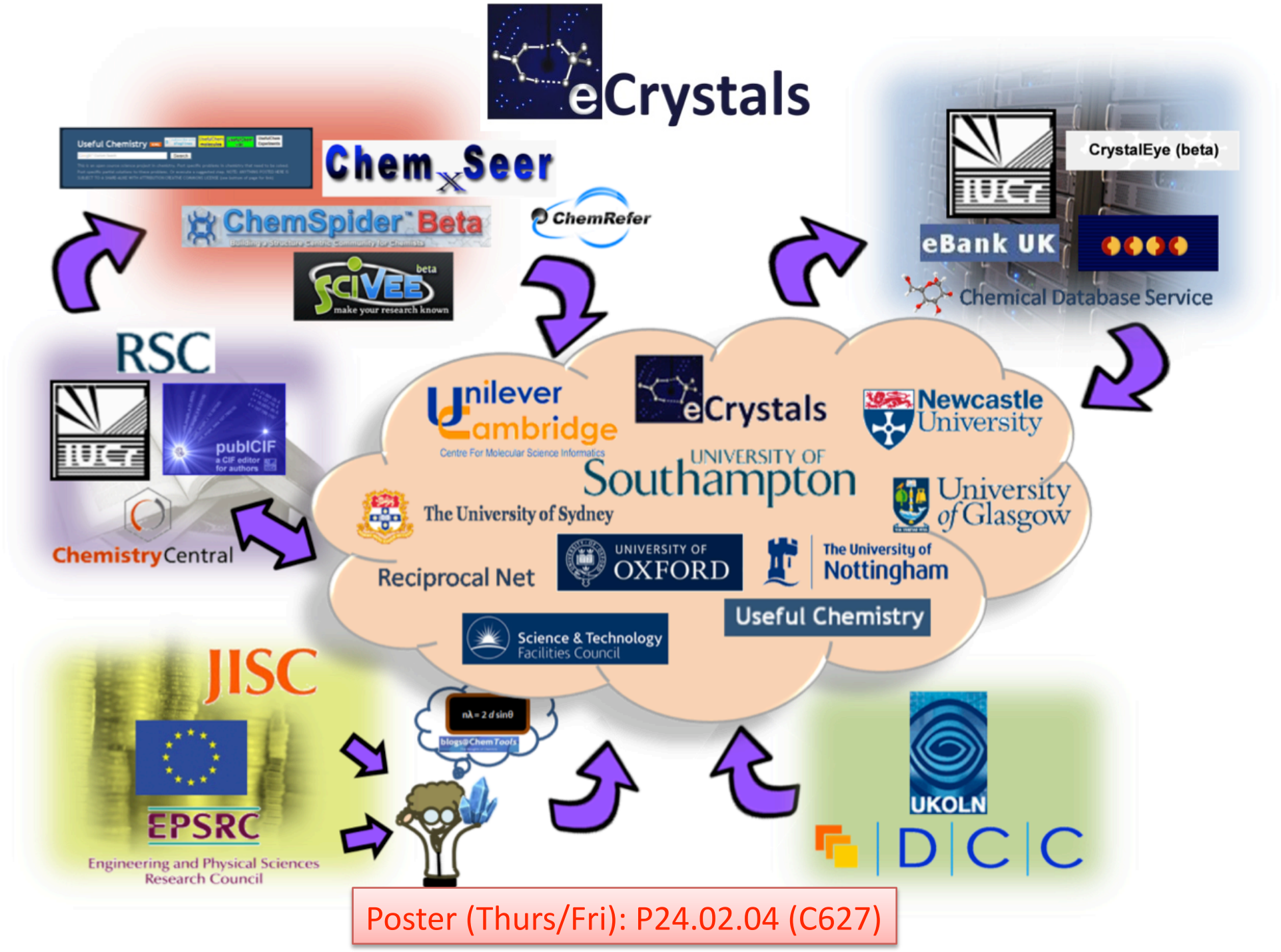
DCC

eprints

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

A Thorough Approach to Dissemination

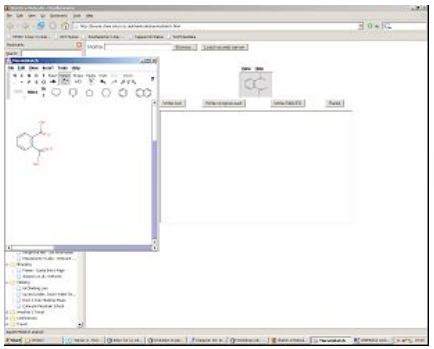
- Using simple Dublin Core protocol (OAI-PMH)
    - Crystal structure
    - Title (Systematic IUPAC Name)
    - Authors
    - Affiliation
    - Creation Date
- Additional chemical information through Qualified Dublin Core
    - Empirical formula
    - International Chemical Identifier (InChI)
    - Compound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



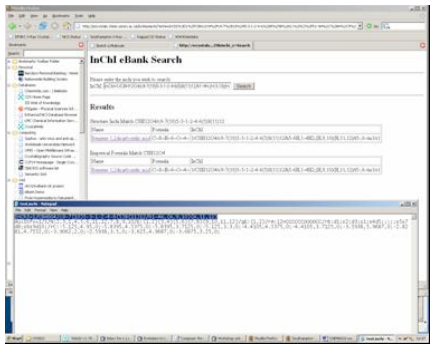
Poster (Thurs/Fri): P24.02.04 (C627)

Simple Integration into Teaching

- Component of MChem course…using REAL data



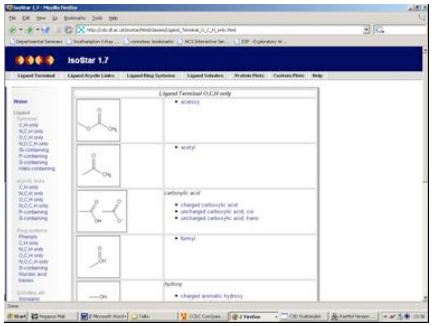
Devise



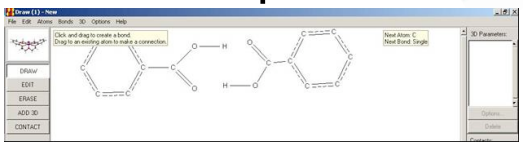
Search

Discover

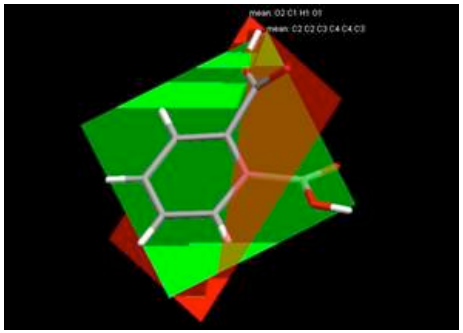
Manipulate

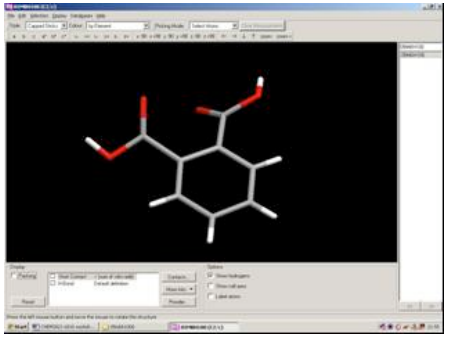


Compare



Analyse





Open Science Experiment

eCrystals — University of Southampton

Home | About | Browse by Year | Browse by People

Login | Create Account — Search

N-(4-methylphenyl)acetamide

Sample Originator: Terry L. Threlfall[a].

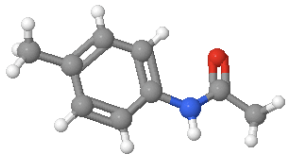
Data Collection: Susanne L. Huth[a]

Structure Determination: Susanne L. Huth[a], Michael B. Hursthouse[a] and Simon J. Coles[a].

University of Southampton[a]

$C_9H_{11}N_1O_1$

InChI=1/C9H11NO/c1-7-3-5-9(6-4-7)10-8(2)11/h3-6H,1-2H3,(H,10,11)



Jmol

Identification Number:	10.3737/ecrystals.chem.soton.ac.uk/544
Controlled Keywords:	para-substituted acetanilide
Date Created:	21 June 2008
Deposited On:	01 Aug 2008 17:29
Deposited By:	Ms L.S. Huth

Data collection parameters

Chemical formula	C9 H11 N O

Available Files

Final Result

2008lsh008.cif	11k
2008lsh008.cml	3k
2008lsh008.fcf	89k

Validation

2008lsh008_checkcif.htm	7k

eCrystals — University of Southampton

Home | About | Browse by Year | Browse by People

Login | Create Account — Search

N-(4-methoxyphenyl)acetamide

Sample Originator: Terry L. Threlfall[a].

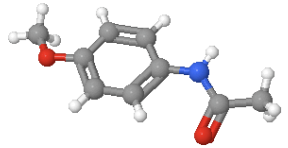
Data Collection: Susanne L. Huth[a]

Structure Determination: Susanne L. Huth[a], Michael B. Hursthouse[a] and Simon J. Coles[a].

University of Southampton[a]

$C_9H_{11}N_1O_2$

InChI=1/C9H11NO2/c1-7(11)10-8-3-5-9(12-2)6-4-8/h3-6H,1-2H3,(H,10,11)



Jmol

Identification Number:	10.3737/ecrystals.chem.soton.ac.uk/543
Controlled Keywords:	para-substituted acetanilide
Date Created:	21 June 2008
Deposited On:	01 Aug 2008 17:24
Deposited By:	Ms L.S. Huth

Data collection parameters

Chemical formula	C9 H11 N O2

Available Files

Final Result

2008lsh013.cif	11k
2008lsh013.cml	3k
2008lsh013.fcf	95k

Validation

2008lsh013_checkcif.htm	7k

eCrystals — University of Southampton

Home | About | Browse by Year | Browse by People

Login | Create Account — Search

N-(4-ethoxyphenyl)acetamide

Sample Originator: Terry L. Threlfall[a].

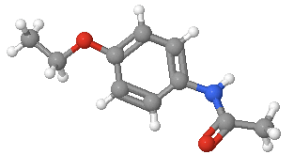
Data Collection: Susanne L. Huth[a]

Structure Determination: Susanne L. Huth[a], Michael B. Hursthouse[a] and Simon J. Coles[a].

University of Southampton[a]

$C_{10}H_{13}N_1O_2$

InChI=1/C10H13NO2/c1-3-13-10-6-4-9(5-7-10)11-8(2)12/h4-7H,3H2,1-2H3,(H,11,12)



Jmol

Identification Number:	10.3737/ecrystals.chem.soton.ac.uk/547
Controlled Keywords:	para-substituted acetanilide
Date Created:	07 July 2008
Deposited On:	01 Aug 2008 17:45
Deposited By:	Ms L.S. Huth

Data collection parameters

Chemical formula	C10 H13 N O2

Available Files

Final Result

2008lsh032.cif	11k
2008lsh032.cml	4k
2008lsh032.fcf	104k

Validation

2008lsh032_checkcif.htm	7k

Refinement

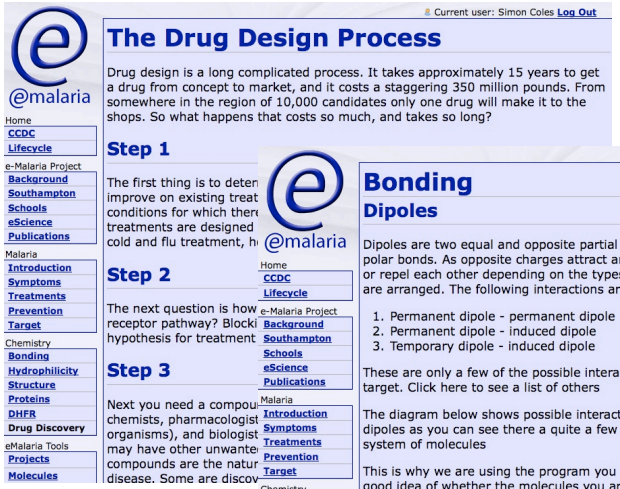
http://ecrystals.chem.soton.ac.uk/544/

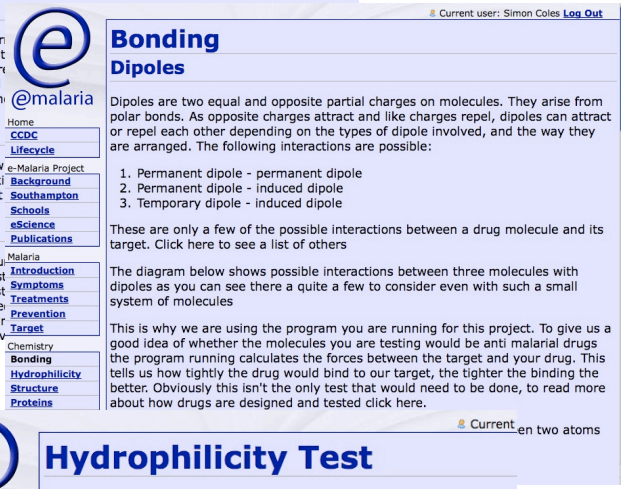
http://ecrystals.chem.soton.ac.uk/543/

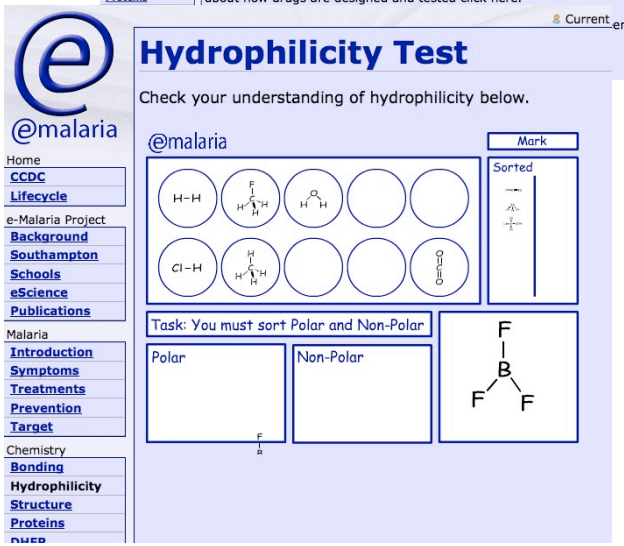
http://ecrystals.chem.soton.ac.uk/547/

Susanne Huth Talk (Sat): MS.98.3 (C163)

eMalaria: Drug Discovery by School Children

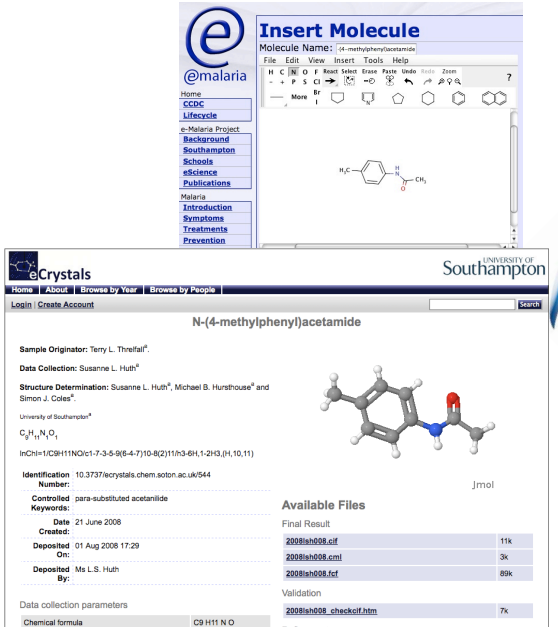


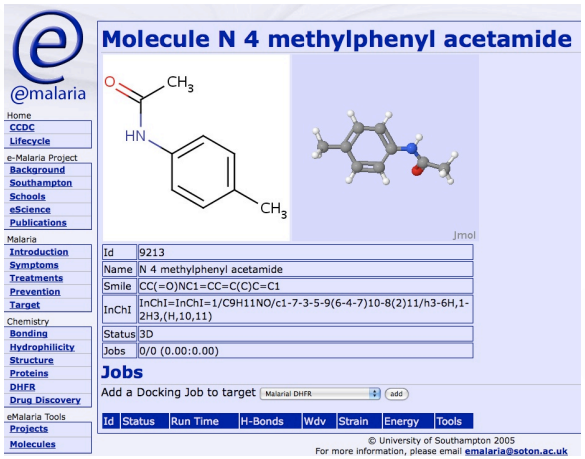


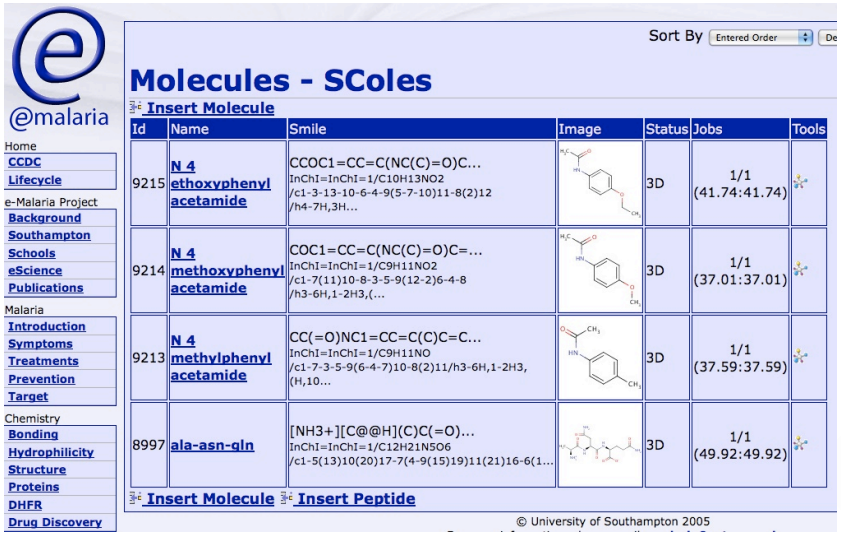


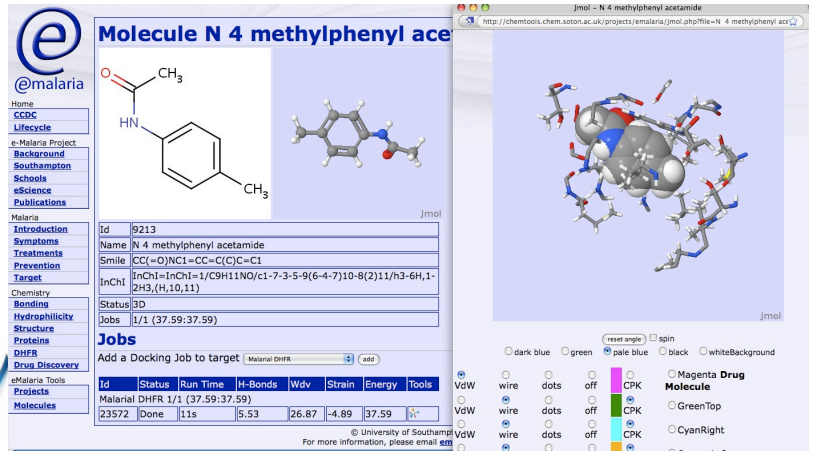
http://chemtools.chem.soton.ac.uk/projects/emalaria/

Open Science Experiment

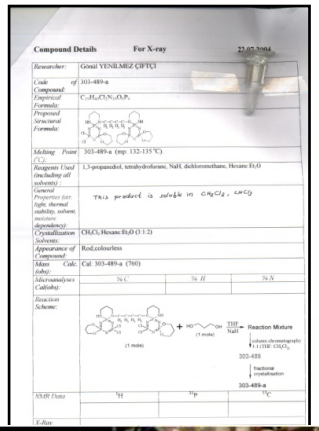


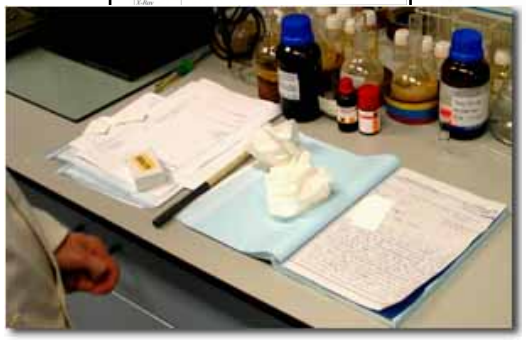






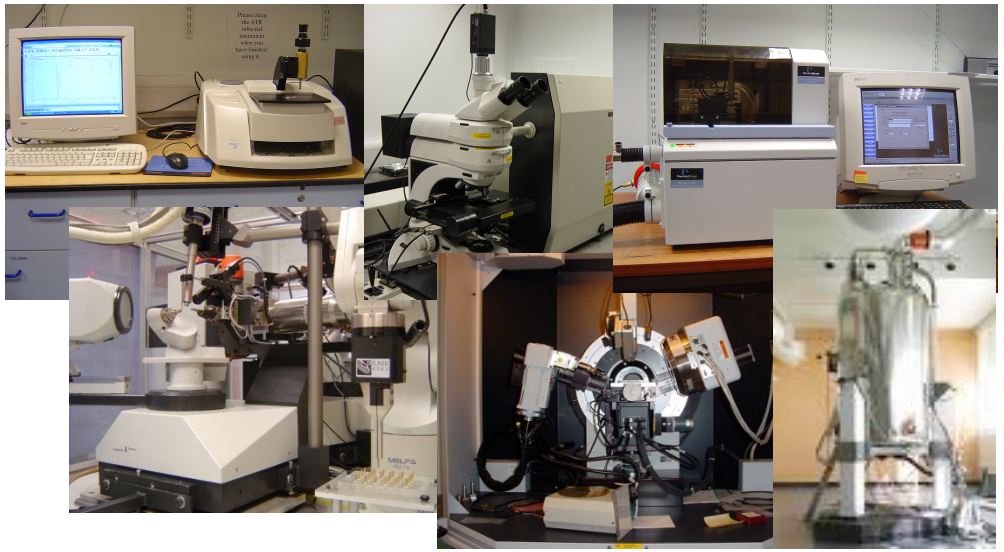
General Chemistry Issues: Data Generation

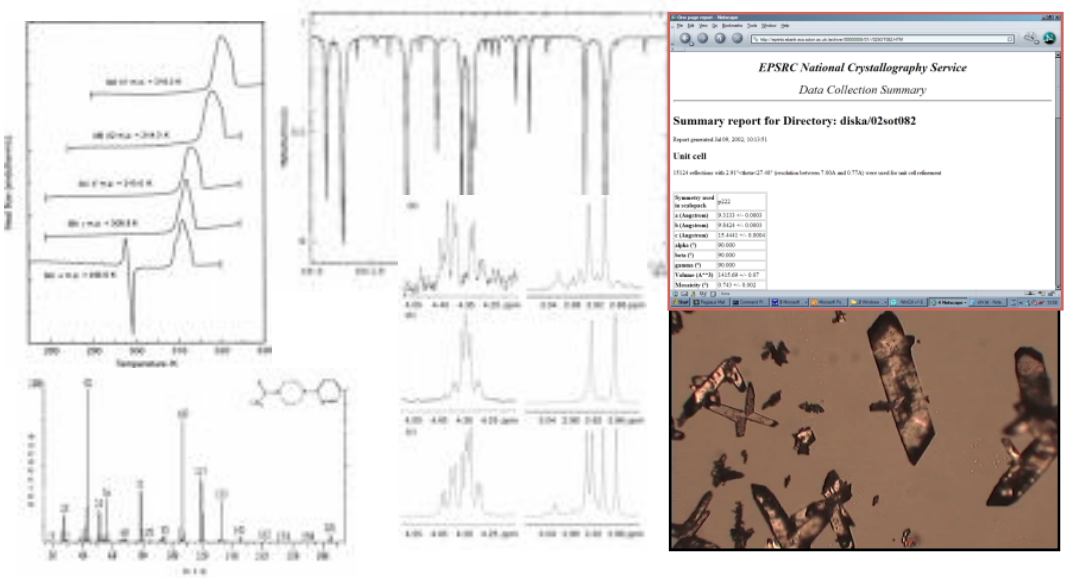




Synthesis

Characterisation





Shortfalls in 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"

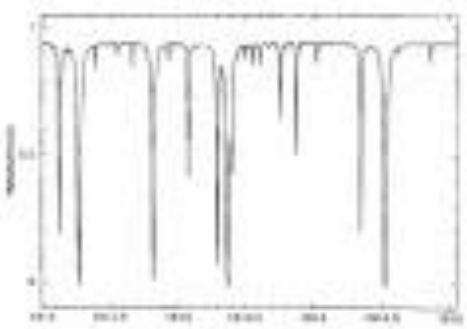
"Lost in some research assistant's computer, the data are often irretrievable or an undecipherable string of digits"

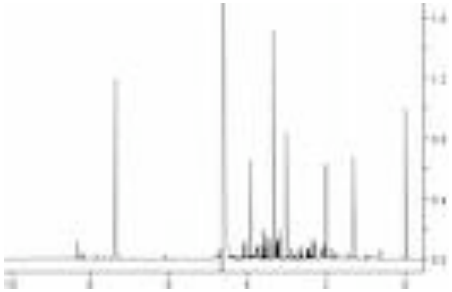
"To vet experiments, correct errors, or find new breakthroughs, scientists desperately need better ways to store and retrieve research data"

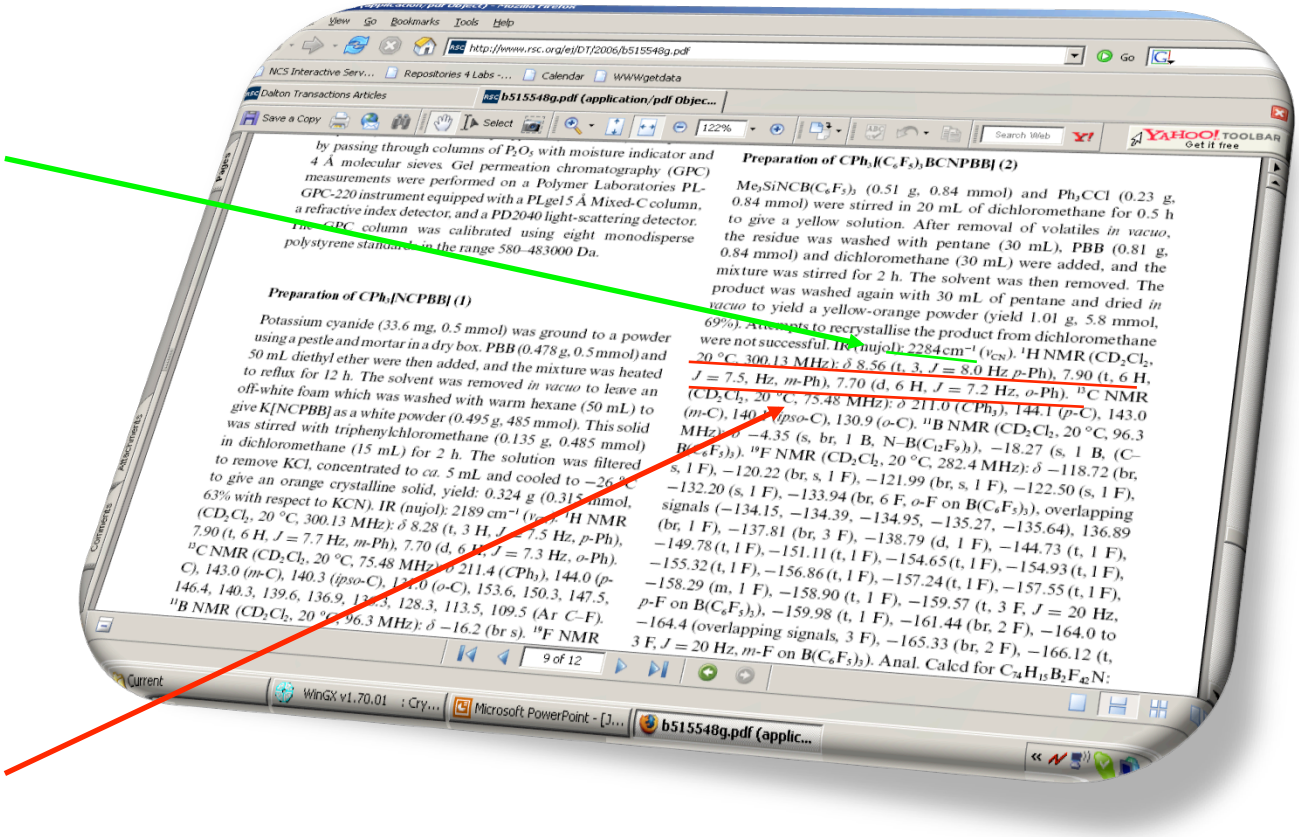
"Data from Big Science is … easier to handle, understand and archive. Small Science is horribly heterogeneous and far more vast. In time Small Science will generate 2-3 times more data than Big Science."

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

Analysis Data Publication & Information Loss



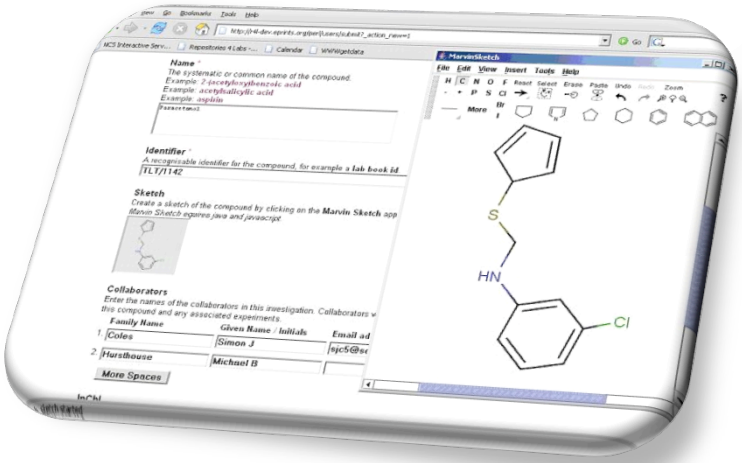




Spectroscopic analysis is often performed to ensure a reaction is proceeding according to plan – as a result <5% are published (via a process with heavy information loss)

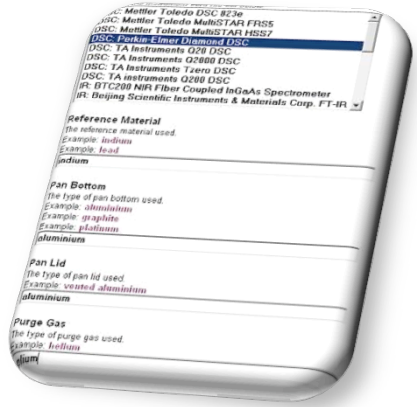
A General Chemistry Laboratory Repository

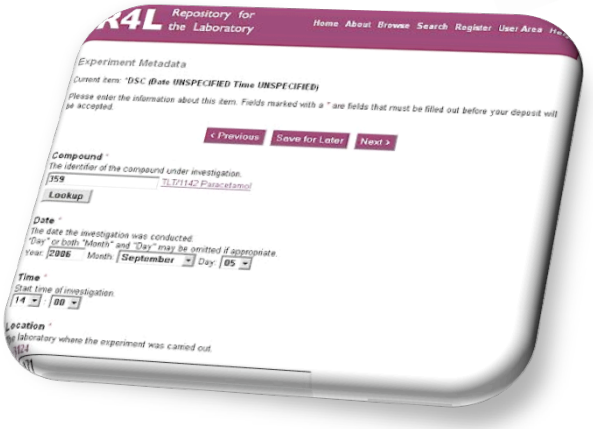
Create new compound (parent record)



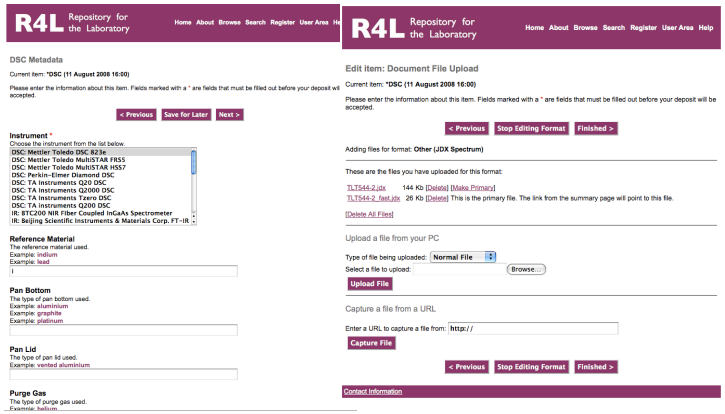
Add new experiment type

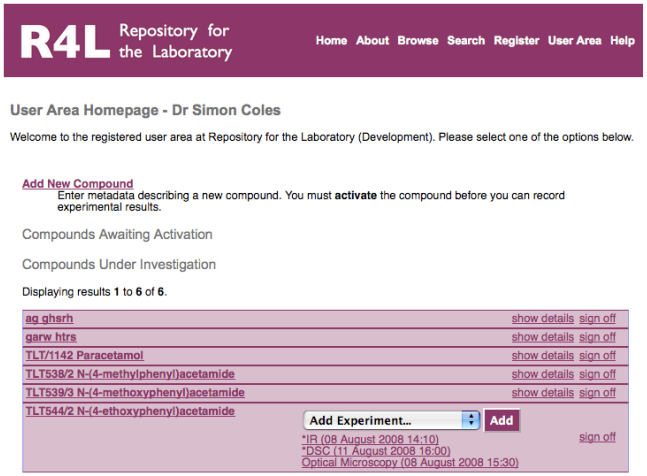
Add metadata and upload data files

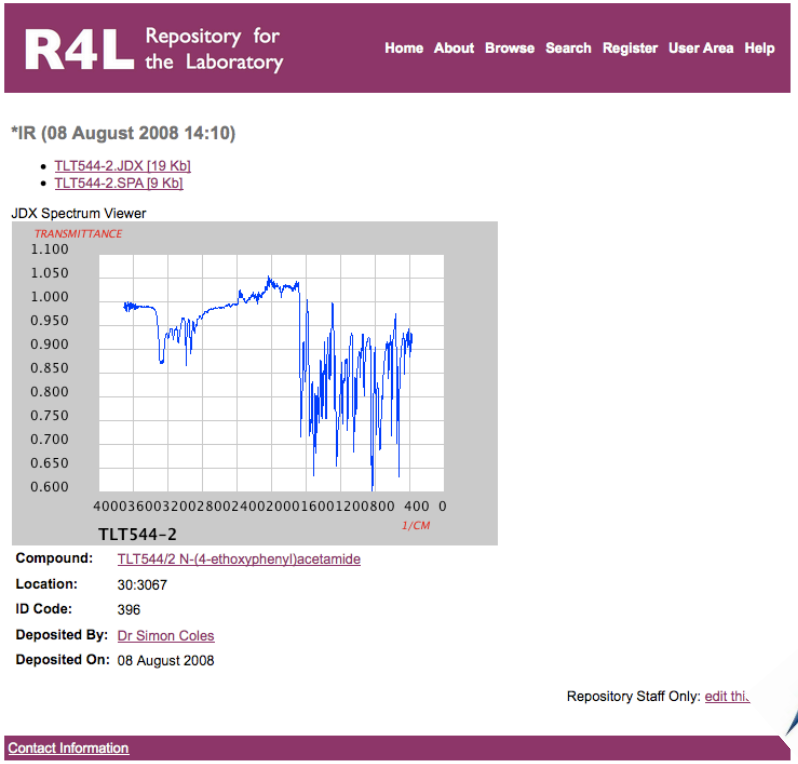


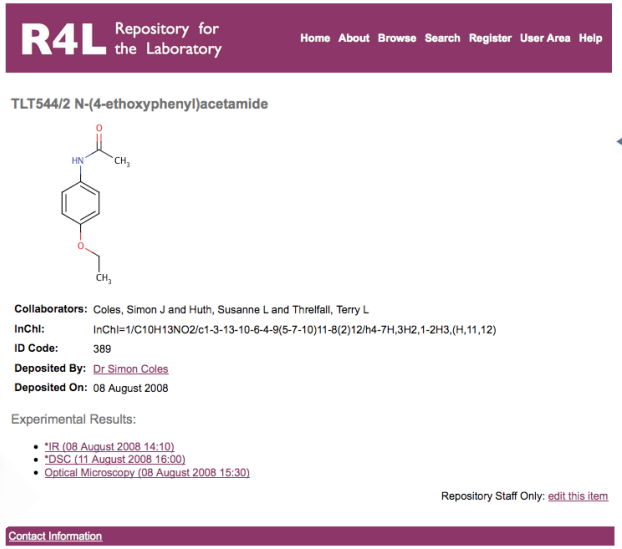


Open Science Experiment



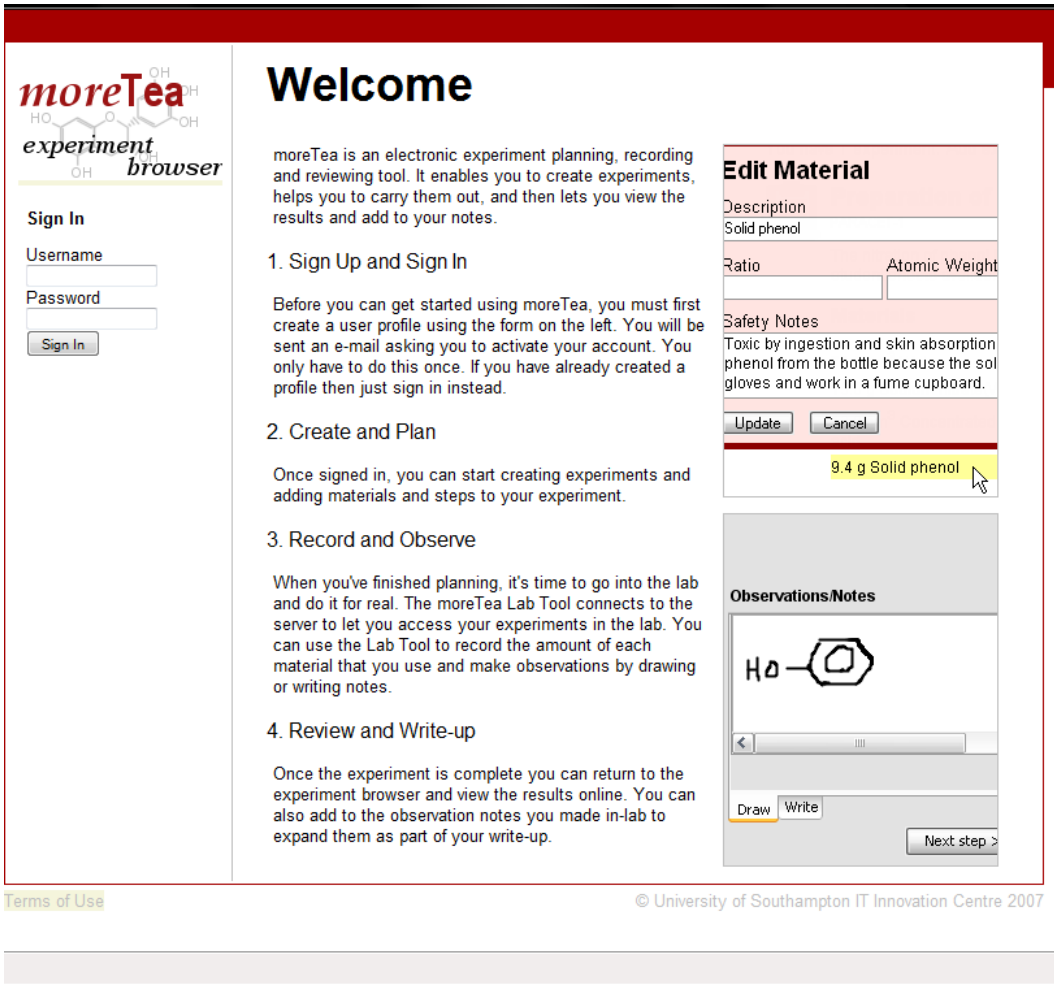


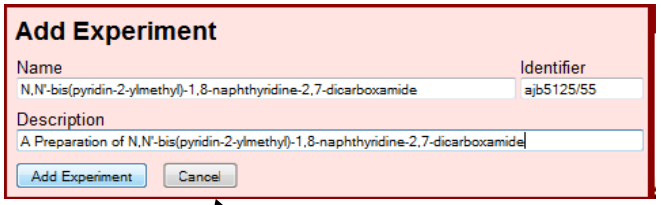


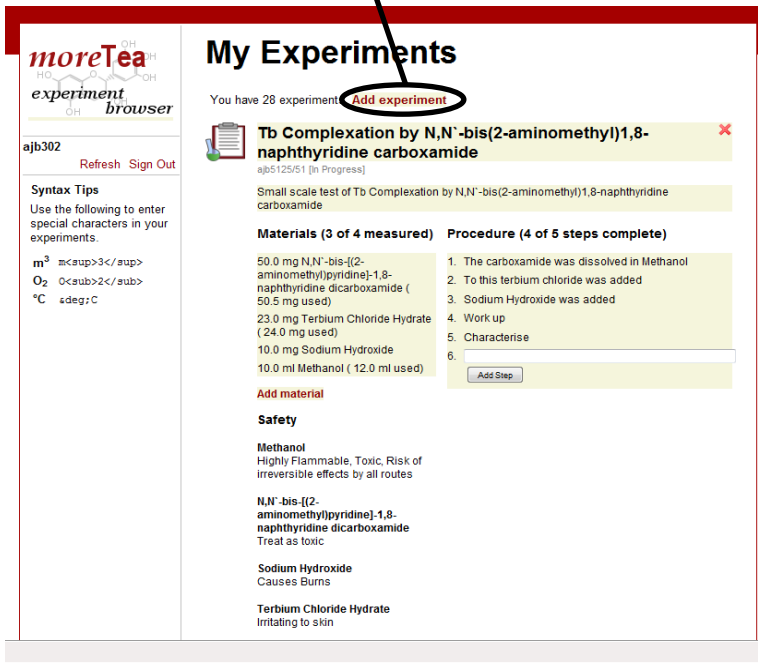


http://r4l.eprints.org

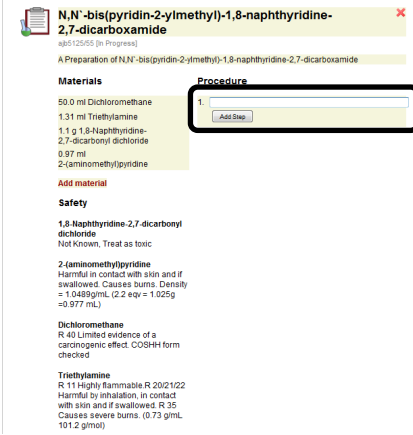
A New Kind of Electronic Lab Notebook

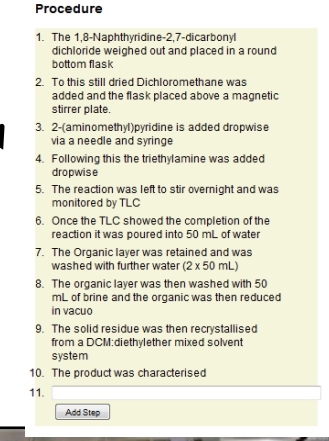


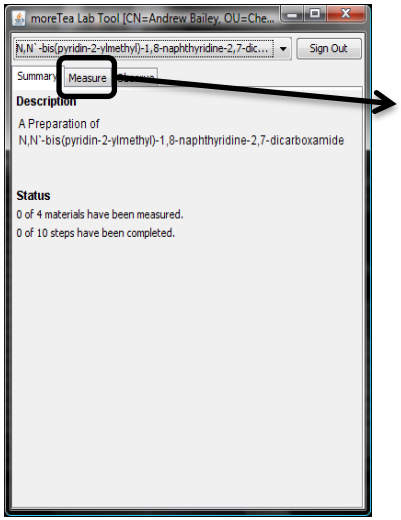


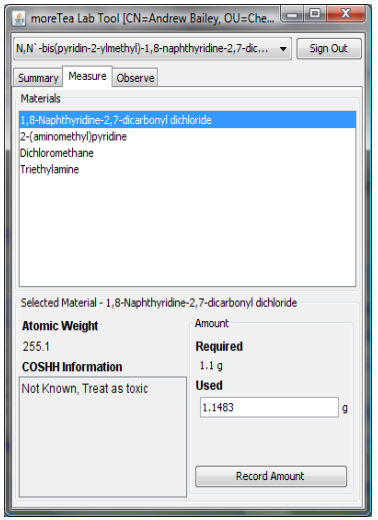


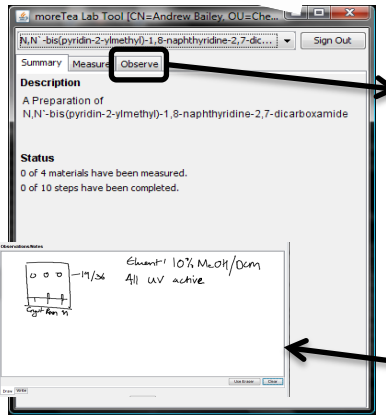
A New Kind of Electronic Lab Notebook

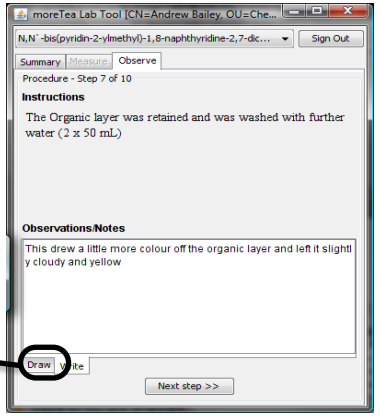








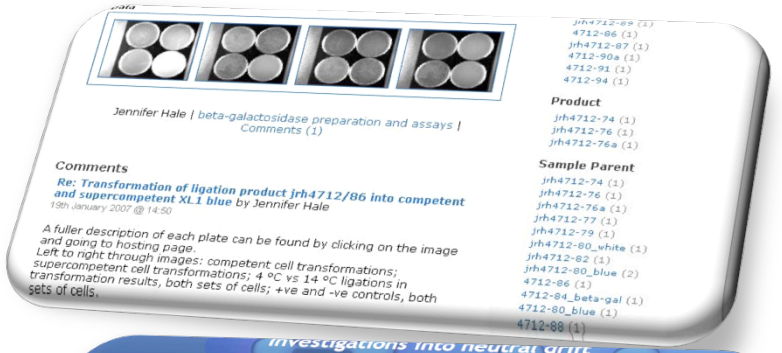


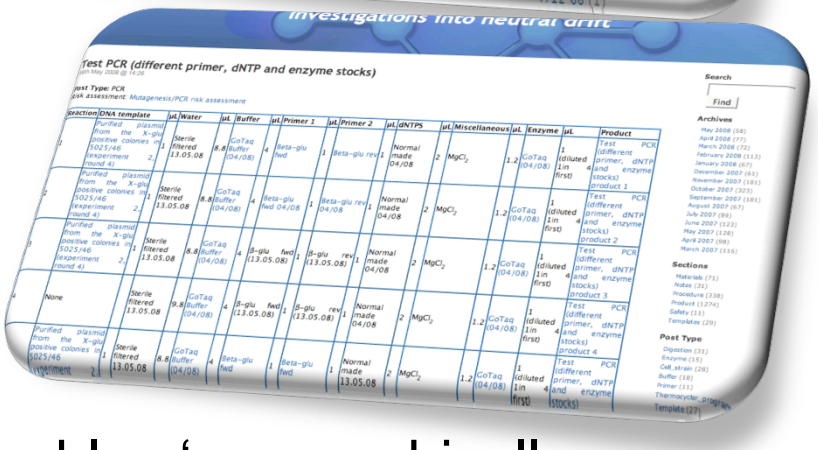


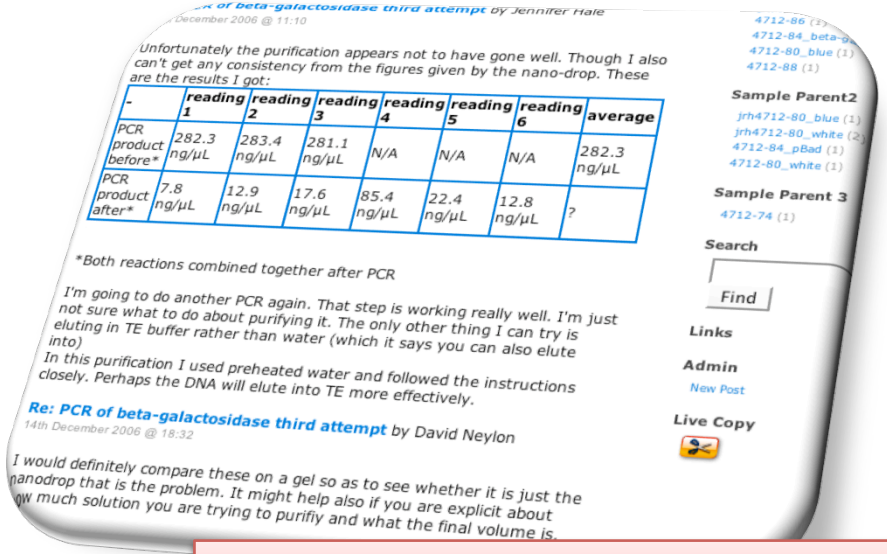
Analysis & Discussion: Blogging Experiments

A repository can…

- Allow one to *put*, *store* and *get*
- Provide *search* and *browse* functionality
- NOT provide the *presentation* and *discussion* functions essential to working up a scientific study



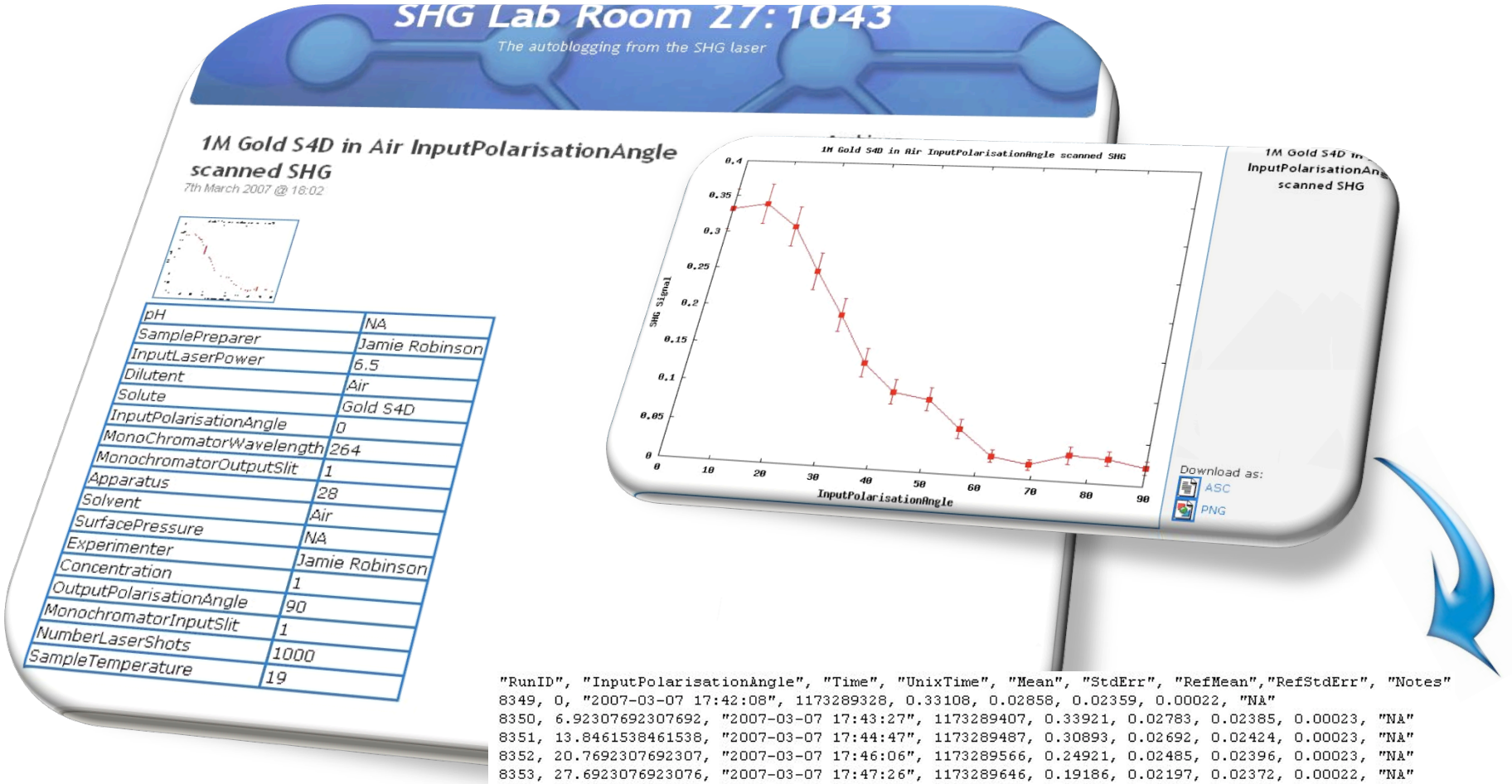




- Enables 'geographically distributed collaborative research'
- Can be open or private
- A useful approach for sharing 'failed' experiments?

http://chemtools.chem.soton.ac.uk/projects/blog/

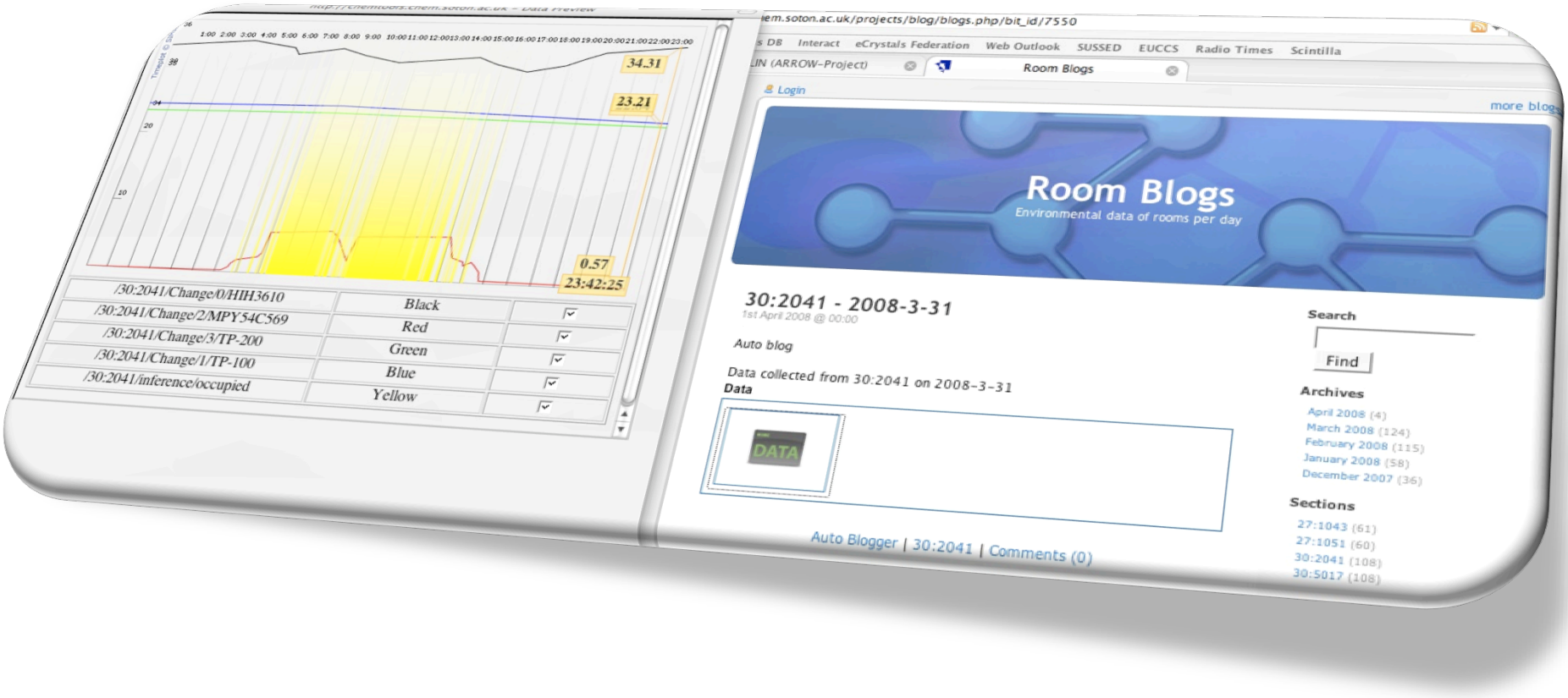
Automatic Blogging by Machines




```
"RunID", "InputPolarisationAngle", "Time", "UnixTime", "Mean", "StdErr", "RefMean","RefStdErr", "Notes"
8349, 0, "2007-03-07 17:42:08", 1173289328, 0.33108, 0.02858, 0.02359, 0.00022, "NA"
8350, 6.92307692307692, "2007-03-07 17:43:27", 1173289407, 0.33921, 0.02783, 0.02385, 0.00023, "NA"
8351, 13.8461538461538, "2007-03-07 17:44:47", 1173289487, 0.30893, 0.02692, 0.02424, 0.00023, "NA"
8352, 20.7692307692307, "2007-03-07 17:46:06", 1173289566, 0.24921, 0.02485, 0.02396, 0.00023, "NA"
8353, 27.6923076923076, "2007-03-07 17:47:26", 1173289646, 0.19186, 0.02197, 0.02372, 0.00022, "NA"
8354, 34.6153846153845, "2007-03-07 17:48:45", 1173289725, 0.13147, 0.01842, 0.02377, 0.00022, "NA"
8355, 41.5384615384614, "2007-03-07 17:50:05", 1173289805, 0.09684, 0.01536, 0.02407, 0.00023, "NA"
8356, 48.4615384615383, "2007-03-07 17:51:24", 1173289884, 0.08984, 0.01470, 0.02406, 0.00023, "NA"
8357, 55.3846153846152, "2007-03-07 17:52:43", 1173289963, 0.05578, 0.01183, 0.02456, 0.00023, "NA"
8358, 62.3076923076921, "2007-03-07 17:54:03", 1173290043, 0.02520, 0.00770, 0.02401, 0.00023, "NA"
8359, 69.230769230769, "2007-03-07 17:55:22", 1173290122, 0.01750, 0.00661, 0.02438, 0.00023, "NA"
8360, 76.1538461538459, "2007-03-07 17:56:41", 1173290201, 0.03136, 0.00998, 0.02397, 0.00023, "NA"
8361, 83.0769230769228, "2007-03-07 17:58:01", 1173290281, 0.03012, 0.00846, 0.02393, 0.00023, "NA"
8362, 89.9999999999997, "2007-03-07 17:59:20", 1173290360, 0.02174, 0.00789, 0.02383, 0.00023, "NA"
```

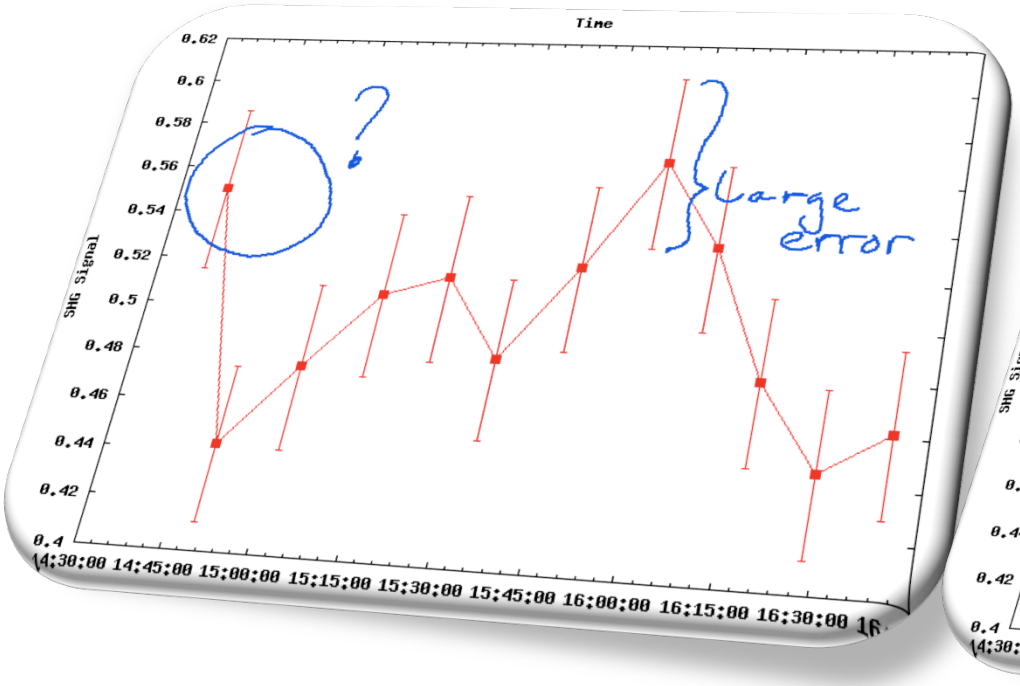
Automatic Blogging by Sensors

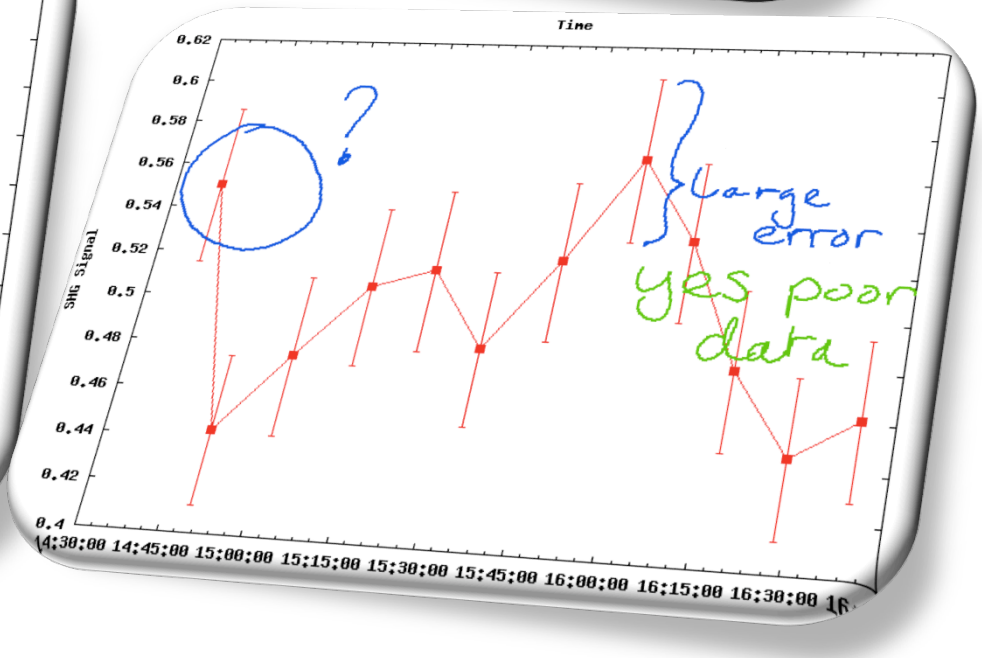
- Continuous log of ‘environmental’ conditions in a laboratory
- Instant detection of erroneous events
- Correlate with inconsistencies in datasets



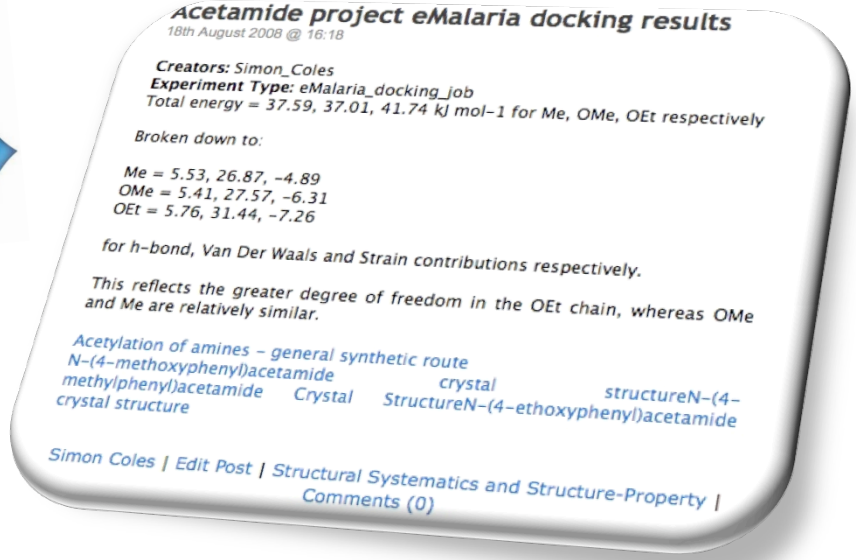
Comments and Collaborative Tools

- Annotation tools allow comments and foster collaboration and / or communication
- Need for more advanced Blog tools / technology around data

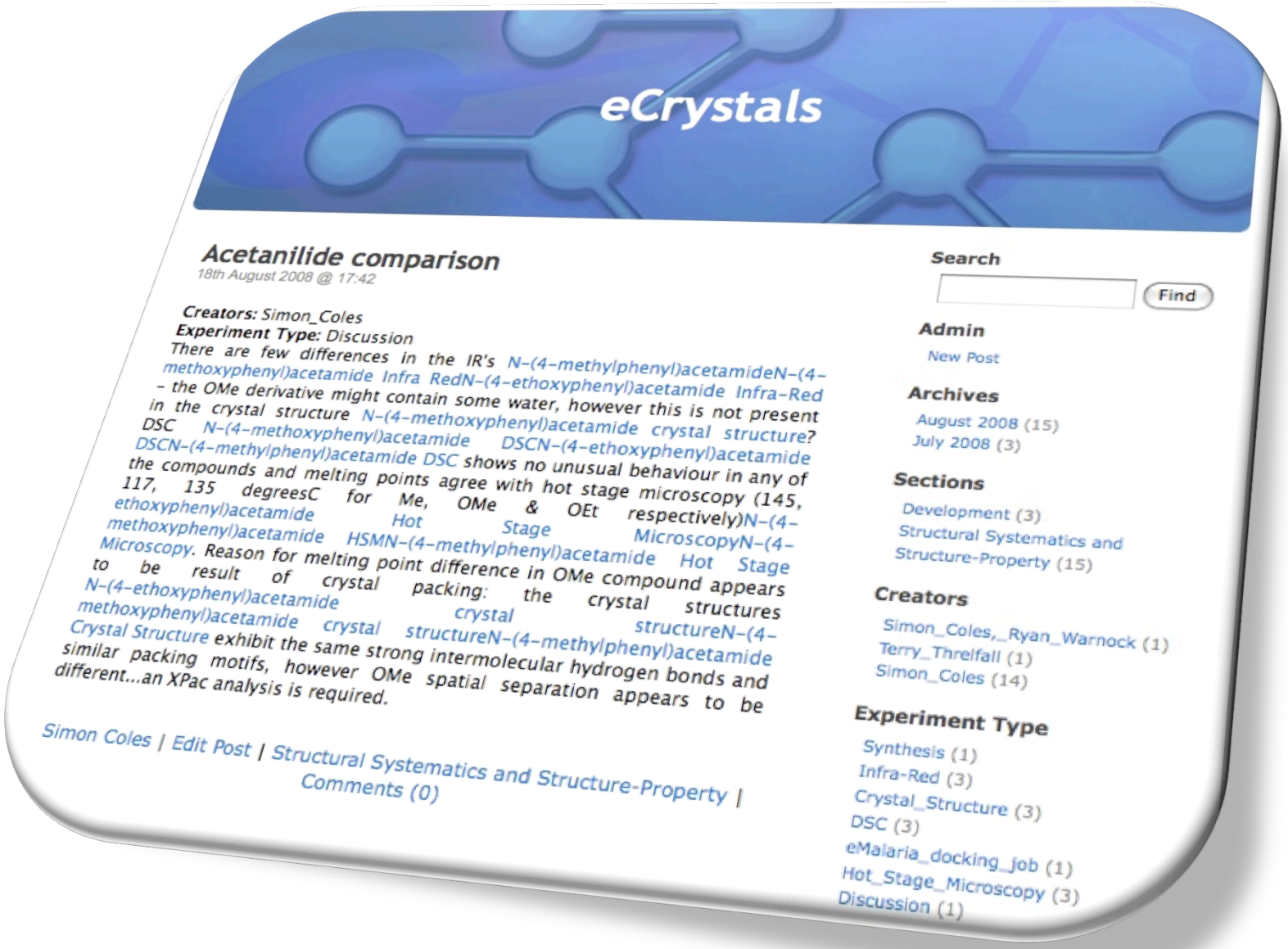




Open Science Experiment



Open Science Experiment



Packaging, Describing and Sharing

- New moves in Digital Libraries community to enable distributed repositories to exchange content
- OAI-ORE (Open Archives Initiative – Object Reuse and Exchange)
- http://www.openarchives.org/ore/
- Describes an aggregation of objects in an exchangeable format
- Microsoft funded eChemistry testbed project

Los Alamos NATIONAL LABORATORY

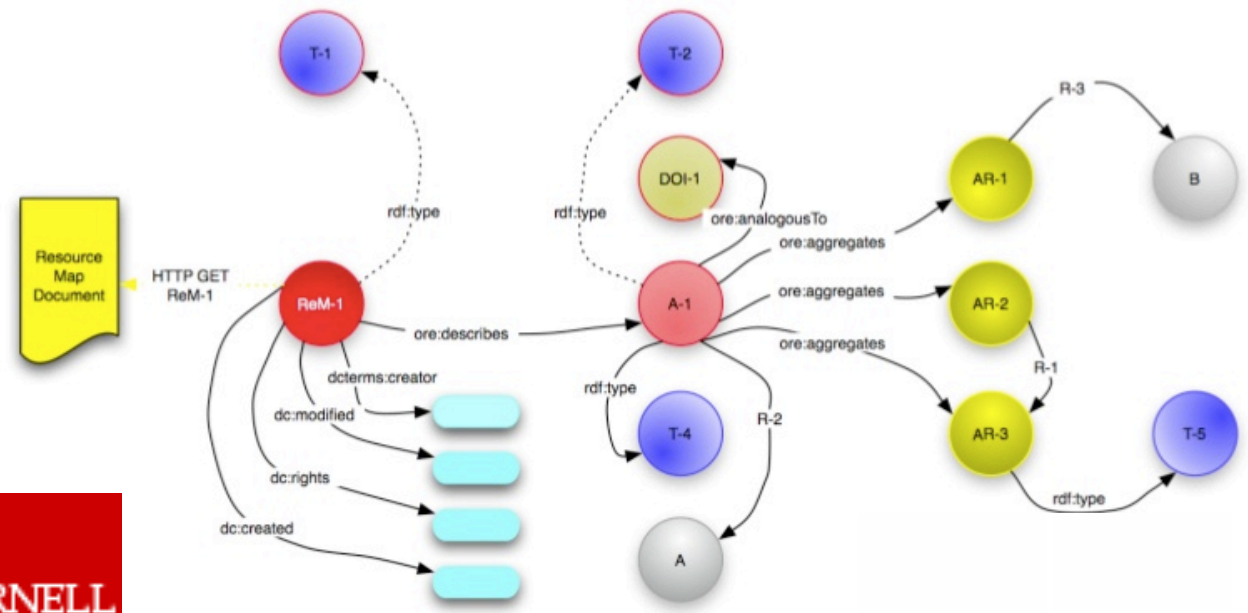
Indiana University School of informatics

ChemXSeer

PubChem

Unilever Cambridge Centre For Molecular Science Informatics

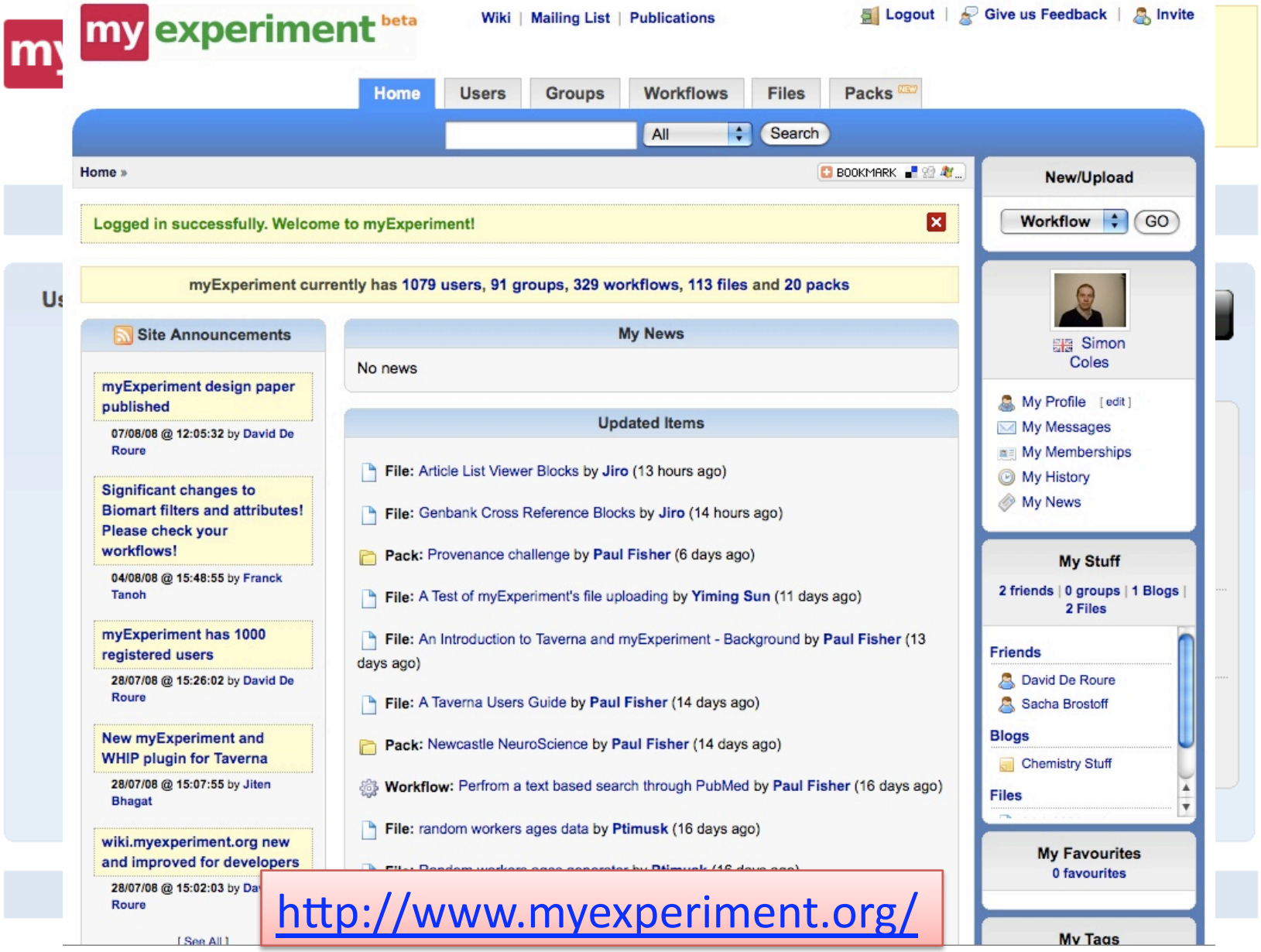
CORNELL



Microsoft

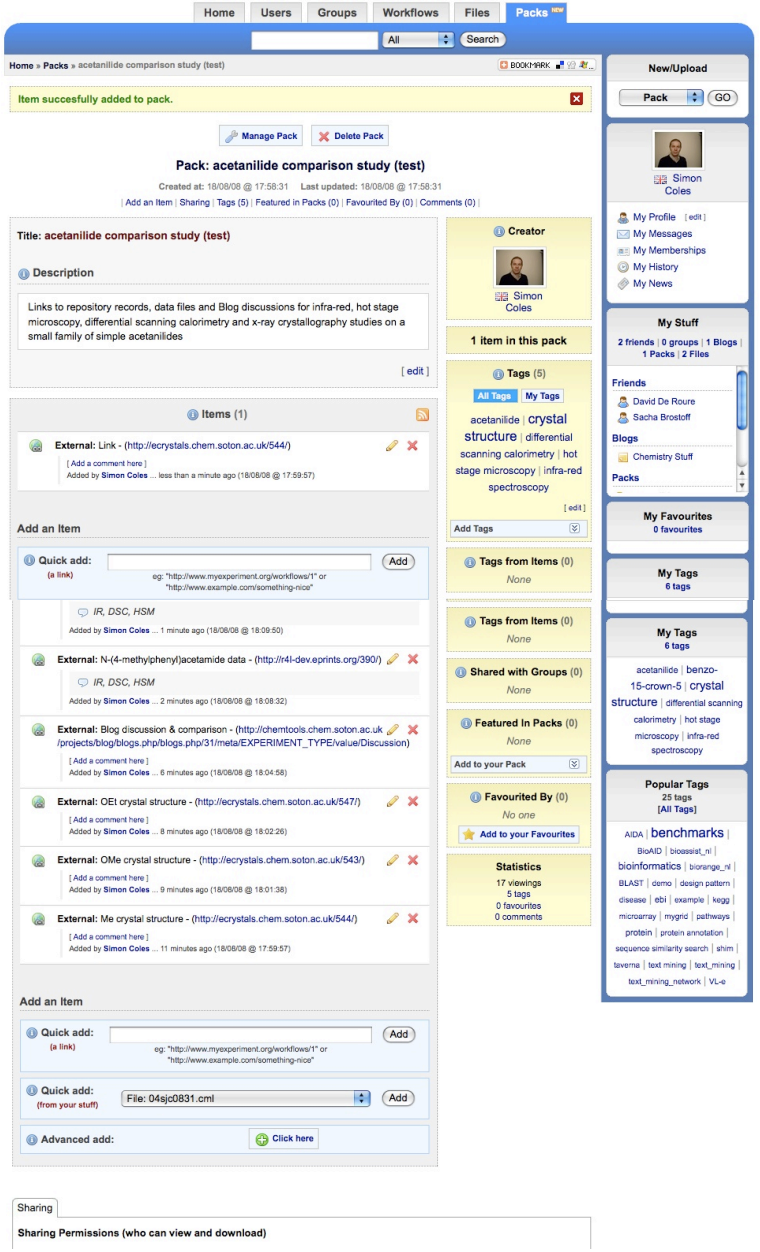
Towards a New Model for Chemical Information Exchange

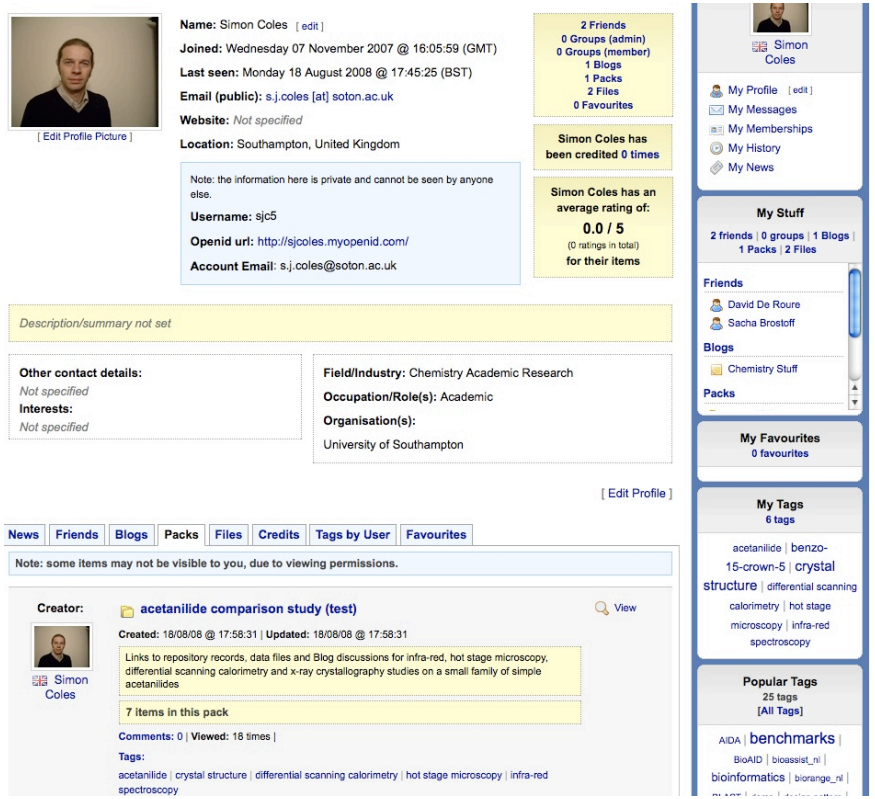
‘Sharing Experiments’



http://www.myexperiment.org/

Open Science Experiment





A solid foundation for Open/Self-Publishing and Communication of Chemistry Data???

Thanks to:

• Mike Hursthouse, Jeremy Frey, Cameron Neylon, Andrew Milsted, Richard Stephenson, Jamie Robinson, Steven Wilson, Andrew Bailey, Mark Borkum

• Dave DeRoure, Les Carr, monica schraefel, Chris Gutteridge, Tim Myles-Board, Arouna Woukei, Dave Tarrant, Stuart Middleton

• Liz Lyon, Rachel Heery, Monica Duke, Michael Day, Traugott Koch, Manjula Patel, Pete Cliff