

## Administration, raw diffraction data, structure factors and coordinates at the UK's National Crystallography Service Management of people, process and data

Prof. Simon Coles (s.j.coles@ncs.ac.uk)

Director, UK National Crystallography Service

# Service UK National Crystallography Service The NCS

- The most powerful and highest throughput chemical crystallography laboratory
- National Research Facility, Est 1981
- Expert research staff
- Synchrotron component
- Around 100 users from UK academia
  - Chemists, crystallographers, materials scientists
  - Advanced techniques

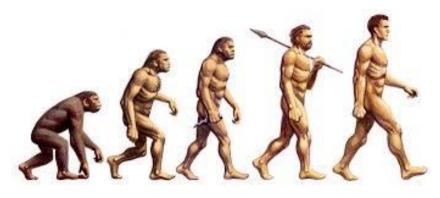






## **Crystallography Service** The NCS and Data Management

- Several eras...
  - 2001-2008 CombeChem e-Science
  - 2003-2009 JISC eBank, eCrystals, R4L, I2S2, ...
  - 2008-2010 Microsoft OREChem
  - 2010-2012 HEFCE Smart Research Framework (LabTrove)
  - 2013-2015 Jisc CREAM
  - 2016- Its what we do...



### © **NCS** | UK National Crystallography Service Mandates: it all started with... Research Councils UK principles

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few
  restrictions as possible in a timely and responsible manner that does not harm intellectual property.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best
  practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly
   <u>available to enable other researchers to understand the research and re-use potential</u> of the data. Published results should always include
   information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process
  is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all
  stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

## Crystallography Service

### Interpreted by my research council as...

EPSRC Engineering and Phys Research Council	sical Sciences						<b>D</b> Search
*	FUNDIN	G	RESEARCH	INNOVATION	SKILLS	NEWS, EVENTS AND PUBLICATIONS	ABOUT US
About us Service standards	<pre>and policies &gt;</pre>	Home / Ab	bout us / Service standards and po	icies / EPSRC policy framework on re	search data		See also Access to research publications e <sup>2</sup>
EPSRC policy framework on research data		EPSRC policy framework on research data This policy framework sets out EPSRC's expectations & concerning the management and provision of access to EPSRC-funded research data.				Related links Freedom of Information Act 2000 [ICO.] C	
Principles Scope and benefits		EPSRC recognises that a range of institutional policies and practices can satisfy these expectations, and encourages research organisations to develop specific approaches which, while aligned with EPSRC's expectations, are appropriate to their own structures and cultures. The expectations arise from seven core principles c <sup>2</sup> which align with the core RCUK principles on data sharing. Two of the principles are of particular importance: firstly, that publicly funded research data should generally be made as widely and freely available as possible in a timely				Freedom of Information (Scotland) Act 2002 [GOV.UK] Freedom of Information and research data: Questions and answers *	
Exploitation of research results and collaborative research		and responsible manner; and, secondly, that the research process should not be damaged by the inappropriate release of such data. The framework was endorsed by the EPSRC Council in March 2011 and implemented from 01 May 2011. It was developed with the benefit of advice from university administrators, from academics, and from research collaborators based in industry.					
Impact, timescales	and support	The policy	reflects the principal UK lega	al provisions intended to assure	public access to publicly	held information, the most relevant of which to	
Responsibility for costs		EPSRC-funded research data are contained in the Freedom of Information Act (2000) and the Freedom of Information (Scotland) Act (2002) a (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004 and the Environmental Information (Scotland) Act (2002) a (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004) and the Environmental Information (Scotland) Act (2002) a (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004 and the Environmental Information (Scotland) Act (2002) a (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004 and the Environmental Information (Scotland) Act (2002) a (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004 and the Environmental Information (Scotland) Act (2002) a (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004 and the Environmental Information (Scotland) Act (2002) a (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004 and the Environmental Information (Scotland) Act (2002) a (other relevant legislation includes the Environmental Information (Scotland) Act (2002) a (other relevant legislation (Scotland) Ac					
Expectations		Information (Scotland) Regulations 2004). These Acts allow any person to ask any public authority (including Universities) for any information they believe to be held by that authority, and require the authority to respond in writing stating whether or not they hold the information sought and, if so, to supply that information unless certain exemptions apply. The exemptions, which may be absolute or qualified, generally relate to considerations such as national security, law enforcement, commercial interests or data protection; all of these may be relevant to research dat and a range of guidance is available to help universities understand their obligations (See for example advice published by Joint Information Service Committee c <sup>2</sup> .) Note: the exemptions in Scotland differ in certain important respects from those in the remainder of the UK.					

### **Calculational Crystallography Service** Resulting in 'data principles' Spot the difference...

- i. EPSRC-funded research data is a public good produced in the public interest and should be made freely and openly available with as few restrictions as possible in a timely and responsible manner.
- ii. EPSRC recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process (including the collaborative research process) is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these constraints are considered at all stages in the research process.
- iii. Sharing research data is an important contributor to the impact of publicly funded research. To recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- iv. EPSRC-funded researchers should be entitled to a limited period of privileged access to the data they collect to allow them to work on and publish their results. The length of this period will depend on the scientific discipline and the nature of the research.
- v. Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice and should exist for all data. Data with acknowledged long term value should be preserved and remain accessible and useable for future research.
- vi. Sufficient metadata should be recorded and made openly available to enable other researchers to understand the potential for further research and re-use of the data. Published results should always include information on how to access the supporting data.
- vii. It is appropriate to use public funds to support the preservation and management of publicly-funded research data. To maximise the scientific benefit which can be gained from limited budgets, the mechanisms for managing and providing access to research data should be both efficient and cost-effective in the use of such funds.

## Solution Crystallography Service However they expect that...

i. Research organisations will promote internal awareness of these principles and expectations and ensure that their researchers and research students have a general awareness of the regulatory environment and of the available exemptions which may be used, should the need arise, to justify the withholding of research data

Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed.

- iii. Each research organisation will have specific policies and associated processes to maintain effective internal awareness of their publicly-funded research data holdings and of requests by third parties to access such data; all of their researchers or research students funded by EPSRC will be required to comply with research organisation policies in this area or, in exceptional circumstances, to provide justification of why this is not possible.
- iv. Publicly-funded research data that is not generated in digital format will be stored in a manner to facilitate it being shared in the event of a valid request for access to the data being received (this expectation could be satisfied by implementing a policy to convert and store such data in digital format in a timely manner).
- v. Research organisations will ensure that appropriately structured metadata describing the research data they hold is published (normally within 12 months of the data being generated) and made freely accessible on the internet; in each case the metadata must be sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it. Where the research data referred to in the metadata is a digital object it is expected that the metadata will include use of a robust digital object identifier (Forexample as available through the DataCite organisation 0).
- vi. Where access to the data is restricted the published metadata should also give the reason and summarise the conditions which must be satisfied for access to be granted. For example 'commercially confidential' data, in which a business organisation has a legitimate interest, might be made available to others subject to a suitable legally enforceable non-disclosure agreement.
- vii. Research organisations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10 years from the date that any researcher 'privileged access' period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party; all reasonable steps will be taken to ensure that publicly-funded data is not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK.
- viii. Research organisations will ensure that effective data curation is provided throughout the full data lifecycle, with 'data curation' and 'data lifecycle' being as defined by the Digital Curation Centre. The full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated within the research organisation, and where research data is subject to restricted access the research organisation will implement and manage appropriate security controls; research organisations will particularly ensure that the quality assurance of their data curation processes is a specifically assigned responsibility.
- ix. Research organisations will ensure adequate resources are provided to support the curation of publicly-funded research data; these resources will be allocated from within their existing public funding streams, whether received from Research Councils as direct or indirect support for specific projects or from higher education funding councils as block grants.



### NCS data management approach

What we need to do to ensure our users are compliant when they use 'our' data...

	Metal-Organic Fireworks: MOFs as integrated structural so for pyrotechnic materials Ith January 20xx, Jhh January 20xx 9/x0xx00000x			
Iniversity of Southampton EPrints User Are	a			
Logged in as Dr Simon Coles   Manage deposits   Mar	nage shelves   Search by Identity   REF outputs selection (IRMS)   ResearchFish   My Statistics   Profil	le   Saved searches   Logout		
/iew Item: Metal-Organic Fireworks! MOF	s as Structural Scaffolds for Pyrotechnic Materials			
Item has been deposited.				
Your item will not appear on the public website	until it has been checked by an editor.			
	This item is in review. It will not appear in the repository until it has been checked by an editor.			
Preview	Details Actions	History		
Coles, Simon , Blair, Lisa and Vrcelj, Ranko (UNSPEC	FIED) Metal-Organic Fireworks! MOFs as Structural Scaffolds for Pyrotechnic Materials. [dataset] Download			
	PDF (Supporting information for Metal Organic Fireworks) - Supplemental Material Restricted to Registered users only until 31 August 2015. Available under License Data: Open Database License (ODbL) (Attribution-Share Alike). Download (10Mb)			
	Archive (ZIP) (Supporting Data for Metal Organic Fireworks) - Data Restricted to Registered users only until 31 August 2015. Available under License Data: Open Database License (ODbL) (Attribution-Share Alike). Download (2447Kb)			
	Description/Abstract			
A new approach to formulating pyrotechnic materials is presented whereby constituent ingredients are bound together in a solid-state lattice in the form of a Metal-Organic Framework. This reduces the batch inconsistencies arising from the traditional approach of combining powders by ensuring the key ingredients are 'mixed' in stoichiometric quantities and are in intimate contact. Further benefits for the application of these types of material are increased safety levels as well as simpler logistics, storage and manufacture. A systematic series of new frameworks comprising fuel and oxidiser agents (group 1 and 2 metal nodes and terephthalic acid derivatives as linkers) has been synthesised and structurally characterised. These new materials have been assessed for pyrotechnic effect by calorimetry and burn tests. Results indicate that these materials exhibit the desired properties of a pyrotechnic mate-rial and that the effect can be correlated to the di-mensionality of the structure.				

Item Type: Dataset

Divisions: Faculty of Natural and Environmental Sciences > Chemistry > Characterisation and Analytics

### DOI: 10 1039/v0vv00000v

### Crystallography Service Useful in the Core Fields



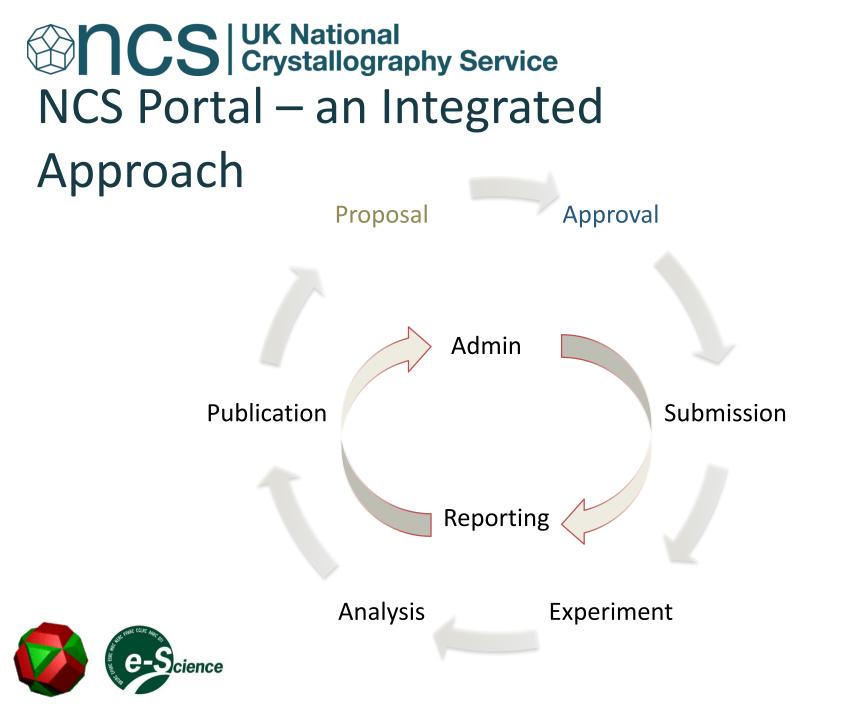
Field	Value
Metadata Created	27 January 2016, 13:48 (UTC+00:00)
creator	
creator_identifier	
creator_identifier_scheme	
creator_affiliation	
title_type	
subject	
subject_scheme	
subject_scheme_identifier	
keywords	
geo_location	
geo_location_point	
geo_location_box	
geo_location_place	
format	
unique_resource_identifier	
publisher	
publisher_identifier	
project	
funder_name	
funder_id	
date	
date_type	
rights	
license_name	
license_uri	
access	
resource_type	
resource_type_additional	
contact	
contact_email	
language	
related_identifier	
related_identifier_type	
relation_type	
image_thumbnail	
metadata_created	
metadata_license	
dditional Info	
Field	Value
Maintainer	University of Southampton



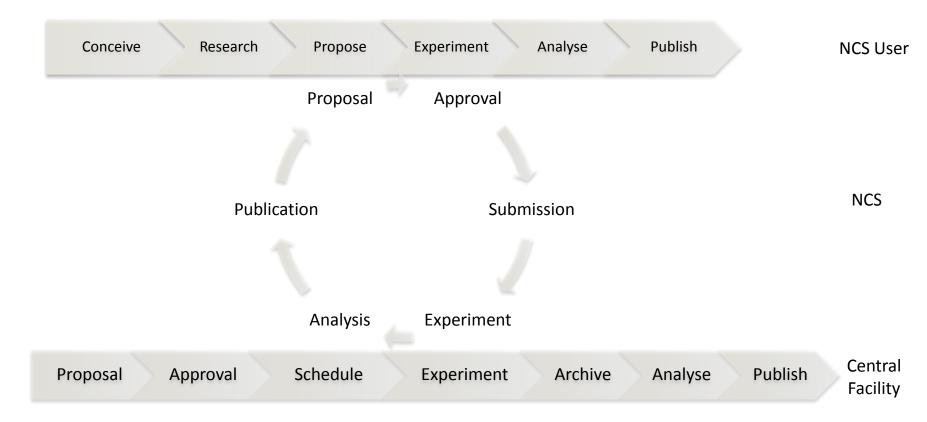
### as Structural als

sented whereby constituent form of a Metal-Organic from the traditional approach of ixed' in stoichiometric quantities tion of these types of material ge and manufacture. A widiser agents (group 1 and 2 s been synthesised and assessed for pyrotechnic effect naterials exhibit the desired be correlated to the

520inf	← Explore -
	← Explore -
13:48 (UTC+	00:00)



# **Operation Across Organisations**





## **Crystallography Service** Sample Information Management

### System

<Screen>

<?xml version="1.0"?>
<fileType Type = "Sample">
<fulleType Type = "Sample">
<less.Athay"/>
<ProjectName = "Russ.Athay"/>
<ProjectName Name = "P1" />
<SampleName Name = "P5IMS11" />
<ImageDirectory Directory = "C:\Images\Rus
<ImageTemplate Template = "EP5IMS11"/>
<ProjectInfo>

<DatabaseID ID = "AX345"/>
<UserID ID = "AX2456"/>
<UserID ID = "CAMSM33"/>
<Priority Priority = "Medium"/>
<SubmissionForm Form = "Standard OnLine"
<UserForentias Properties = "Multiple C
<BulkMorphology Morphology = "Mostly rod
<SolutionGoal Goal = "Publishable Struct
<UserFormula Formula = "GOSN3H13"/>
<LabelingScheme Scheme = "Standard"/>
<Sensitivity Sensitivity = "Light Sensit
<Safety Safety = "Caustic medium"/>
<DateProjectSubmission Date = "2012.01.06"/>
</ProjectInfo>

<SampleInfo>

<CrystalNumber Number = "4"/> <Crystallographer Name = "Graham"/> <PreparationScheme Scheme = "Chip of lar <FragmentCut Cut = "Yes"/> <CrystalColor Color = "Colorless"/> <CrystalSize X = "0.20" Y = "0.20" Z = " <CrystalMorphology Morphology = "Prism"/ <Solvent Solvent = "PEG200"/> <ManipulationMedium Medium = "Stuff"/> <MountType Type = "Fiber"/> <SampleMountID ID = "NCS2765"/> <VisualQuality Quality = "Unknown"/> <OpticalProperties Properties = "Transpa <VisuallyTwinned Twinned = "No"/> <DateDatabaseExtraction Date = "2012.01. <DateMounted Date = "2012.01.07"/> <DateScreened Date = "2012.01.07"/> <DateCollected Date = "2012.01.08"/> <DateProcessed Date = "2012.01.09"/> <DateSolved Date = "2012.01.09"/> <DateDatabaseDeposit Date = "2012.01.10" </SampleInfo>

<Screen>

<ScreenProgram Program = "CrystalClear3.</pre>

<ScreenProgram Program = "CrystalClear3.1a4"/> <Environment TempC = "128.00" Pressure = "29.89" Humidity <ColdStreamTemp TempC = "-148.40"/> <Instrument Instrument = "AFC12 Saturn 724+"/> <Wavelength Wavelength = "0.7070000"/> <Element Element = "Moly"/> <Solve> <ScreenScans> <Scan ScanNumber = "1" Axis = "Omega"> <ScanRange Start = "-60.0" End = "-59.0" Width = "1.0 <GonioAxes ChiOrKappa = "54.0" PhiOrOmega = "0.0"/> <DetectorSettings Distance = "48.8" Exposure = "10.0"</pre> <DetectorMode Binning = "2" Dezinger = "Yes"/> </Scan> </ScreenScans> <Rank Rank = "135"/> <IndexProgram Program = "CrystalClear3.1a4"/> <IndexAlgorithm Algorithm = "d\*TREK"/> <CSDMatches> <CSDMatch Match = "Someone in Cambridge did this years <CSDMatch Match = "St Andrews did this in 1994"/> </CSDMatches> <State Type = "Screen"> <Cell A = "5.123" B = "13.987" C = "14.764" Alpha = "90 <CellESD A = "0.000000" B = "0.000000" C = "0.000000" A <CrystalSystem System = "Orthorhombic"/> <SpaceGroupName Name = "P 2 2 2"/> <LaueClass Class = "m m m"/> <Lattice Lattice = "P"/> <Centricity Centricity = "acentric"/> <SpaceGroupNumber Number = "16"/> </Solve> <Rmerge Rmerge = "4.30"/> <History> <Completeness Completeness = "98.70"/> <Redundancy Redundancy = "4.30"/> <TotalReflections Reflections = "34987"/> <UniqueReflections Reflections = "27654"/> <RejectedReflections Reflections = "2456"/> <MaxResolution Resolution = "0.83"/> </State> </Screen> <Collect> <StrategyProgram Program = "CrystalClear 3.1a4"/> <CollectSchedule Schedule = "Default\_Collect"/> <PredictedCompleteness Completeness = "99.80"/> <PredictedRedundancy Redundancy = "4.40"/> <CollectProgram Program = "CrystalClear 3.1a4"/> <Environment TempC = "128.00" Pressure = "29.89" Humidity

<EmpiricalTmin Tmin = "0.04"/> <EmpiricalTmax Tmax = "0.66"/> <Twins> <Twin Number = "1" TwinLaw = "Slide 1/2"/> <Twin Number = "2" TwinLaw = "Twist 1/3"/> </Twins> </Process> <SolutionProgram Program = "Olex2/Shelx"/> <SolutionMethod Method = "Direct"/> <Restraints> <Restraint Parameter = "Param1" Value = "0.0000" Tolerance = "0.1000"/> <Restraint Parameter = "Param2" Value = "1.0000" Tolerance = "0.2000"/> </Restraints> <Constraints> <Constraint Parameter = "Param3" Value = "0.0000" Tolerance = "0.1000"/> <Constraint Parameter = "Param4" Value = "-1.0000" Tolerance = "0.0010"/> </Constraints> <Structure Type = "Solve"> <R1 Percentage = "4.30"/> <Atoms> <Atom> <AtomCoord Label="01" Symbol="0" x="0.000000" y="0.100000" z="0.200000" Occupancy="1.000000"/> <AtomAniso U11="0.000000" U22="0.010000" U33="0.020000" U23="0.000000" U13="0.003000" U12="0.006000"/> </Atom> <Atom> <AtomCoord Label="N1" Symbol="N" x="0.000000" y="0.200000" z="0.400000" Occupancy="1.000000"/> <AtomAniso U11="0.000000" U22="0.020000" U33="0.040000" U23="0.000000" U13="0.006000" U12="0.012000"/> </Atom> </Atoms> </Structure> <Step Date = "12.01.11" Type = "Screen"></Step> <Step Date = "12.01.11" Type = "Note">The sample was abandoned here becaue the reflections were very weak and it could not be indexed</Step> </History> <Results> <ValidationProgram Program = "CheckCif"/> <ValidationErrors> <ValidationError Error = "Rmerge too high"/> <ValidationError Error = "Resolution too low"/> </ValidationErrors> </Results> <EmailAddress Email = "some\_user@localhost"/> <Status Status = "Ready"/> <LastError Error = "None"/> Riaaku <Rank Rank = "135"/>

# Solution (individual)

### Crystallography Service

Current User <u>Service Account</u> (info@ncs.ac.uk)

### Application for an Allocation

Period: 1st Ma	y 2013 - 31st	October 2013
----------------	---------------	--------------

Personal Details	3	
Name:	Dr Mike Coogan	
Email Address:	m.coogan@lancaster.ac.uk	
Department	School of Chemistry	
Institution:	Lancaster University	
Address:	B6 Faraday Building Lancaster University Lancaster LA1 4YB	
Europhia a	United Kingdom	Request for this
Funding		
Funding Source:	Not Funded	Outline of Scie
Local Facilities		We have develop
Local Facilities:	No	are useful cell im
Reason(s) why ad	ditional facilities are required:	hosts for smaller
	his is a new department just starting up and while we intend to buy a diffracton next year at the earliest.	encapsulation. A

### Last Allocation

	FSA	DSO
	Full Structure	Data Collection
Your last allocation was:	5	5
Your usage over the last period was:	0	2

### Could you indicate the percentage of these:

Have not/cannot be worked up

Have been fully refined

Have been written up into a report or thesis

Have been published (please give full details of

paper(s) in the next section)

### Request for this Allocation Outline of Scientific Program

We have developed a range of complexes based around luminescent transition metal fragments which are useful cell imaging applications. Some of these are macromolecular structures which can act as hosts for smaller molecules or ions and the luminescence is in many cases modulated upon encapsulation. As well as expanding the range of metals which we can use in cell imaging to the early TMs (Zr, Hf) as well as the more traditional late TMs we are looking at new host macromolecules: hetero-cycle appended calixarenes which form interesting hydrogen bonded structures as well as complexes with a range of transition metals; also complexes based around expanded structures based on polypyridyls in dimeric, trimeric and larger assemblies. All these complexes show interesting photophysical properties, e.g. acting as sensors, acting as imaging agents. The balance between crystalinity and solubility / lipophilicity (essential for crossing cell membranes) is difficult to build in or predict (especially in the macromolecules) with such macromolecules and in many cases it is difficult to obtain high quality single crystals of these species, and thus as well as access for data collections only, we have requested full structure solution for difficult cases where the expertise of the service will be essential.

5

5

### Publications

### Additional Information

Please accept my apologies for the lack of samples outputs for the last period- this is a result of the development of the new chemistry department at Lancaster having been more complex that originally anticipated, so no lab-work has been possible until this week as major refurbishments were undertaken to bring the facilities to standard.

Attached Files

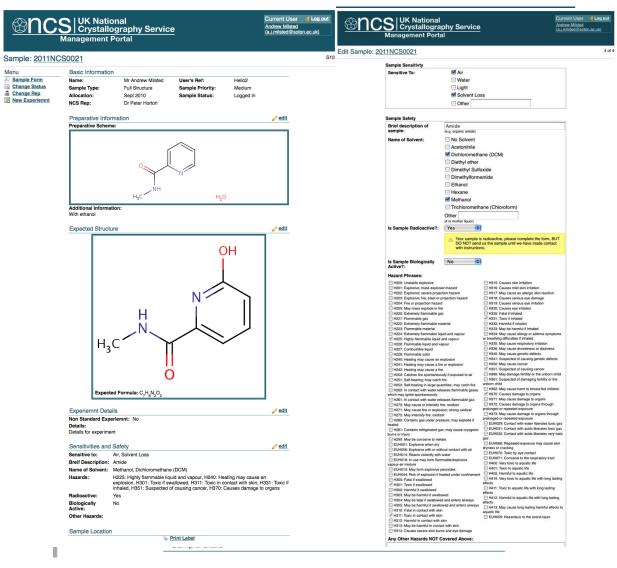
## **Crystallography Service** Application

### (service crystallographer)

Last Allocation

	FSA	DSO
	Full Structure	Data Collection Only
Your last allocation was:	20	15
Your usage over the last period was:	6	0
Could you indicate the percentage of these:		
Have not/cannot be worked up		70
Have been fully refined		15
Have been written up into a report or thesis		5
Have been published (please give full details of paper(s) in the next section)		
Request for this Allocation	20	15
Supported Researchers		
1. Dr Bruno Linclau		
Funding Source: Other Funding		
Other Grants:		
Publications:		
Research Keywords:		
Carbohydrate chemistry, organofluorine chemistry, to	otal synthesis	

# Sample submission



## OCS UK National Crystallography Service

## Management Portal

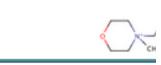
### Sample: 2012NCS0650

돈.	Samp	le Fo	orm
----	------	-------	-----

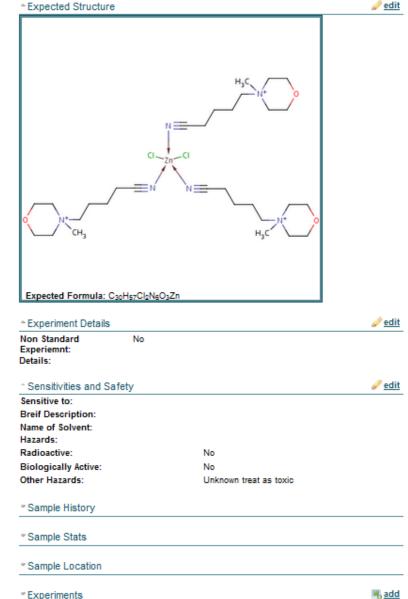
- Change Status
- Change Rep
- 🔏 New Sample (User)

Basic Information	1
Name:	Dr Peter N
Sample Type:	Data Colle
Allocation:	May 2012
NCS Rep:	Dr Graham

* Preparativ	ve Information
Preparative	Scheme:



Additional Information: crystallized from melt, no solver



# Crystallography Service

Current User al Log out Graham Tizzard

(q.j.tizzard@ncs.ac.uk)

S2369

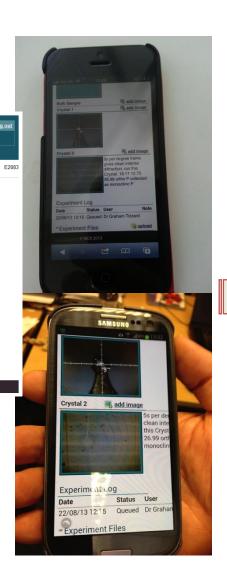
Sample: <u>2013NCS0783</u>

Management Portal

	Menu	Basic Informatio	n				🥔 <u>edit</u>	
	Sample Form	Name:	Mrs Sarah Mils	ted User's	Ref:	TEST 1		
	Change Status	Sample Type:	Full Structure	Samp	le Priority:	Medium		
	Change Rep	Allocation:	November 2013		le Status:	Data Collect	ing	
	Kew Sample (User)	NCS Rep:	Dr Graham Tizz			-	5	
		Preparative Inf	ormation				🥔 <u>edit</u>	
		Expected Struct	cture				🥔 <u>edit</u>	
		₹ Evporimont Do					🥔 <u>edit</u>	
Chan	ge Sample Status		×					
New	Failed	-	У				<u>v cun</u>	
Statu	s: Draft							
	Submitted							
Note	Logged In (Queued)		44					
	Data Collecting							
	Data Re-Collecting							
	Refered to DLS	Save Can	cel					
	Processing							
	Completed Failed	Experiments					遇 <u>add</u>	
	Withdrawn							
	Referred to User							
	Paused							
			2013NCS0768	17-CAM-KFC-4	Full Structure	Medium	Logged In (Queued) Logged In (Queued)	11/12/13
			2013NCS0769	23-CAM-KFC-5	Full Structure	Medium	Logged In (Queued)	11/12/13

# Service UK National Crystallography Service

	S UK National Crystallogra Management Porta		2	Current Use <u>Graham Tizza</u> ( <u>g.j.tizzard@n</u>	d
Sample: 2013NC	S0516 :: Experimen	t: Examinatio	n		
This Experiment	Basic Information				
Change Status	Experiment Started By:	Dr Graham Tizza	rd Experiment Started:	21/08/13 15:33	
Experiments	Experimental Report			a new set	
G First Examination	Packaging			喝 add image	
Section Section	Bulk Sample			喝 add image	
Section Examination	Crystal 1			喝 add image	
	Experiment Log				
	Date	Status	User	Note	
	21/08/13 15:33	Queued	Dr Graham Tizzard		
	Experiment Files			(%) upload	





ebsite © NCS 2013

### **Crystallography Service User View**

CS UK National Crystallography Service **Management Portal** 

Current User a Log out Mark Light (light@soton.ac.uk)

Н	0	n	n	е	

Menu	Your Samples				
Kew Sample	Sample Code	Your Reference	Туре	Priority	Status
Messages	2013NCS0775	2013sot0091	Full Structure	High	Completed
🧬 User Admin	2013NCS0776	2013sot0094	Full Structure	High	Logged In (Queued)
Applications	2013NCS0777	2013sot0093	Full Structure	Medium	Logged In (Queued)
Stats	2013NCS0778	2013sot0092	Full Structure	Medium	Logged In (Queued)
Stats Admin	2013NCS0779	2013sot0090	Full Structure	Medium	Logged In (Queued)
	2013NCS0780	2013sot0089	Full Structure	Medium	Logged In (Queued)
Search	2013NCS0801	2013SOT0095	Full Structure	Medium	Logged In (Queued)

### Current Allocation

Period: November 2013 01/11/13 - 30/04/14					
Sample Type:	4/14				
Full Structure	13/20				
Data Collection Only	15/15				
Sample Priority:					
High	8/10				
Medium	10/15				
Low	10/10				
Reserved 0/0					
(Remaining/Total A	warded)				

### Website © NCS 2013

### S UK National Crystallography Service $(\mathfrak{A})$ Management Portal

### Sample: 2013NCS0775

Menu	Basic Informat	ion						
Sample Form	Name:	Dr Mark Light	User's Ref:	2013sot0091				
	Sample Type:	Full Structure	Sample Priority:	High				
	Allocation:	November 2013	Sample Status:	Completed				
	NCS Rep:	Dr Graham Tizzard	Next Step:	ExaminationEx				
	Preparative Information							
	Expected Str	ucture						

Experiment Details

Sensitivities and Safety

м	UK National Crystallogra	al	Current User 🧃 Log or Mark Light (light@soton.ac.uk)
lit Sample: <u>S240</u>	<u>)0</u>		1
C	reparative Scheme Diagram: Elle Edit View Insert Ato	m Bond Structure Tools Help	
	· · · · · · · · · · · · · · · · · · ·		
	UK National	1	C Br T Current User
M it Sample: <u>S240</u>	anagement Porta	al	<u>(light@soton.ac.uk)</u> 4
Sa	ample Sensitivty		
S	Sensitive To:	Air Water Light Solvent Loss	
	mple Safety		
Sa			
B	Brief description of ample:	(e.g. organic amide)	
B	Brief description of	No Solvent Acetonitrile Cichloromethane (DCM) Diethyl ether Dimethyl Sulfoxide Dimethylformamide	
B	Brief description of ample:	No Solvent Cactonitrile Cacton	
e s N	Brief description of ample:	No Solvent Acetonitrile Cichloromethane (DCM) Diethyl ether Dimethyl Sulfoxide Dimethyl Sulfoxide Ethanol Hexane Methanol Grichloromethane (Chloroform) Other (rin mother lauor) No	
e s N is	Srief description of ample: lame of Solvent:	No Solvent Cacetonitrile Cichioromethane (DCM) Direthyl ether Dimethyl Sulfoxide Dimethyl Sulfoxide Ethanol Hexane Methanol Trichioromethane (Chloroform) Other (In mother liquor)	
B S N Is Is A H H	Brief description of ample: Lame of Solvent: s Sample Radioactive?: s Sample Biologically	No Solvent Acetonitrile Cichloromethane (DCM) Diethyl ether Dimethyl Sulfoxide Dimethyl Sulfoxide Ethanol Hexane Methanol Grichloromethane (Chloroform) Other (rin mother lauor) No	59

H320: Causes eye irritation

H330: Eatal if inha

H204: Fire or projection hazard

H205: M

\_\_\_\_Y

## **Crystallography Service** Data Availability

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

Citation: Onyeabo, Romanus O and Edwards, Mark and Spencer, John and Tizzard, Graham J and Coles, Simon J (2010) University of Southampton, Crystal Structure Report Archive. (doi:10.5258/ecrystals /1505)

Export as: oreChem EndNote BibTeX ASCII Citation



Helping you to find, access, and reuse data



Southampton Crystals ogin | Create Account 6,7,9,10,12,13,15,16-Octahydro-benzo-1,4,7,10,13-pentaoxacyclopentadecin Sample Originator: Esther Rousay<sup>a</sup> and Jeremy G. Frey<sup>4</sup> Data Collection: Simon J. Colesi Structure Determination: Simon J. Coles<sup>8</sup> and Michael B. Hursthouse University of Southerpolog C14H20O5 InChi=1/C14H20O5/c1-2-4-14-13(3-1)18-11-9-16-7-5-15-8-8-17-10-12-19-14/h1-4H.5-12H2 Identification 10.5258/ecrystals/145 Imol Number Controlled crown others, crown Available Files Keywords: Date 07 October 2004 Final Result Created: 04sjc0831.cif 13k Deposited 21 Jan 2008 15:29 On: 04sjc0831.cml 6k Deposited Dr Simon J Coles 04sjc0831.fcf.txt 155k By: Validation Depositor Comments 04sjc0831\_checkcif.htm 7k Structure already known, but accurately redetermined for a local research project. Refinement Data collection parameters 04sjc0831.res 6k Chemical formula C14 H20 O5 04sjc0831\_xL Crystal morphology Plate Solution Crystal system Orthorhombi 04sjc0831.prp Space group symbol Pbca 04sjc0831\_xs Cell length a 16.4963(18 Processing Cell length b 8.325(3) 04sjc0831.hkl Cell length c 20.061/6 04sjc0831.htn Cell angle alpha 90.00 04sjc0831\_0k Cell angle beta 90.00 04sjc0831\_h0 Cell angle gamma 90.00 04sjc0831\_hk Data collection temperature 120(2) Data Collection Refinement results 04sjc0831\_cry Solution figure of merit 0.0409 R Factor (Obs) 0.0487 Other Files R Eactor (AI) 0.0977 04sjc0831.do Weighted R Factor (Obs) 0.1008 04sjc0831.ins Weighted R Factor (All) 0.1192 04sjc0831.mo 04sjc0831.p4 Citation: Rousay, Esther and Frey, Jeremy G. and Coles, Simon J. and 04sjc0831.pcf

Hursthouse, Michael B. (2004) University of Southampton, Crystal Structure Report Archive. (doi:10.5258/ecrystals/145) Export as: oreChem EndNote BibTeX ASCII Citation

04sjc0831\_ell

eprints 🛛 🔛

lat		34k
2		6k
lst		39k
!		702k
<u>n</u>		10k
1.jpg		57k
Lipg		85k
0.jpg		88k
on		
ystal.jpg		17k
<u>e</u>		78k
		5k
1		3k
2		1k -
f.txt		2k
ipsoid.gif		19k
	Repository Staff Only: item co	entrol page
of Southanning	D	- Province

## Crystallography Service

### eCrystals metadata (discovery)

- Using simple Dublin Core protocol (OAI-PMH)
  - Crystal structure; Title (Systematic IUPAC Name); Authors; Affiliation; Creation Date
- Additional chemical identifiers via Qualified Dublin Core
  - Empirical formula; International Chemical Identifier (InChI); Compound Class & Keywords
- DOI (one for the whole record); Citation
- Specifies 'datasets' present in an entry; Version
- Application Profile <u>http://www.ukoln.ac.uk/projects/ebank-uk/schemas/</u>
- Licensing

## **Crystallography Service** Portal TNG

- Basic functionality for 'data managing' any facility (core),
  - Communication, user-facility interface
  - Set up, background information gathering
  - Collect and process data
  - Archive & Curate
  - Data 'dissemination'
- With add-ins for...
  - Applications / Peer review
  - Queueing and prioritisation
  - Instrument integration
  - Usage statistics
- Impact Acceleration Account project
  - By Research Software Group / Software Sustainability Institute
  - Core freely available on GitHub
  - 'Premium' model for add-ins



Engineering and Physical Sciences Research Council

# Service What about RAW data?

- Now becoming common to deposit structure factors
- We cater very well for routine structures
- If...
  - have accounted for everything in a raw image properly
  - have an acceptable model
  - Is the raw data required any longer?



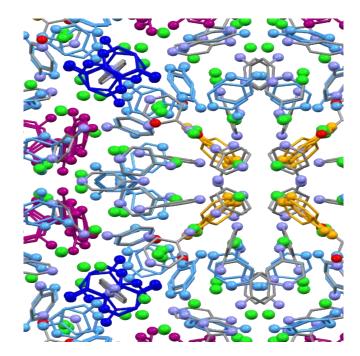
## Crystallography Service

# The case for 'publishing' (Chem Cryst) raw data?

- Validation: clear contribution to chemical knowledge, but poor quality; support a 'grand' claim
- Disorder, twinning, incommensurates
- Diffuse scattering
- 'Advanced Experiments': Charge density, high pressure, phase transition, gas environment, excited state
- Clear that future improvement may be possible; somebody else can do a better job
- Training sets for developers

# Solutional Crystallography Service A real example...

- Exceptional structure ?? Submitted to journal...
- CIF, with structure factors, deposited with CCDC
- eCrystals data
- Still doesn't help others reinterpret in a different way?
- Publish raw data
- Need a procedure...
- Open up our archive?
- Supported community repository?



## Crystallography Service

### Supported, Open, Large data volumes

zenodo	Search Q	Upload	Communities	≗s.j.coles@soton.ac.uk	•
🗎 Delete				Save Vublish	J

### New upload

Instructions: (i) Upload minimum one file or fill-in required fields (marked with a red star ). (ii) Press \*Save\* to save your upload for editing later. (iii) When ready, press \*Publish\* to finalize and make your upload public.

Files 💙		Choose	e files ③ Start upload
Filename (101 files)	Size	Progress	Delete
2012rm001z601001.img md5:e3db9458b5df6a5f3f2e2408e01aa724 @	13.7 Mb	~	ê
2012rm001z601002.img md5:bead64a970d87f2fd663ba61de286e92 @	13.7 Mb	~	Ê
2012rm001z601003.img md5:895361ad9cff53c9b609a3ced9180b6e @	13.7 Mb	~	Ê
2012rm001z601004.img md5:d42967ace9f592cdf5d83fad70e2e266 @	13.7 Mb	~	Ê
2012rm001z601005.img md5:a05415d535e47237d1df1176261c1884 @	13.7 Mb	~	ê
2012rm001z601006.img md5:991ad7655a823fa9f444c49cd6baeeed 🚱	13.7 Mb	~	Ê
2012rm001z601007.img md5:b7116a60afbe1ffae3831560982410a4 @	13.7 Mb	~	ê
2012rm001z601008.img md5:08005135cc9693f5eb0a7c26fb1b3885 @	13.7 Mb	~	ê
2012rm001z601009.img md5:c8c30a65269c00943386ef241dcec98c	13.7 Mb	~	ê

			S	UK N Crys <sup>-</sup>	lation tallog	al raph	y Se	rvic	е									
_			C					Note: File addition, re	moval or modification are not all							staCite for each uploa	d.	
	י סוי	ntv/	$\cap t$	mo	tad	ata	ר י			(min	mum 1 file required,	,max 50 GB per d	ataset - contact u	is for larger datasets	¢			
	ICI	icy	UI	me	tad	ata	;	Upload type	Publication	Poster O	Presentation	Dataset O	image	Video/Audio	<b>≻</b> Software ○	Lesson O	requ	uired 🂙
								Basic information	n								requ	uired 💙
									Digital Object Identifier	10.5281/zer	nodo 557092							
										others to easily		y cite your uploa				register a new DOI fo DOI once it has been		
										IIII Pre-reserve	× 100							
									Publication date *	2017-04-24								
										Required. Form	at: YYYYMM-DD. In	n case your uploi	ad was already p	ublished elsewhere	, please use the d	ite of first publication	L	
									@ Title *	4-aminopyri Required.	dine hydrochlorid	ie dodecihydra	te raw diffractio	on data (SPIDER I	diffractometer)			
Communities							reco	ommended 🔉						1	10 11 1			÷ ×
Funding							reco	ommended >	Authors*	Simon Cole				University	of Southampton			Ŧ×
Related/alterr	nate identifiers						reco	ommended 🔉		+ Add anothe	rauthor							
Contributors								optional 🔉	Pescription *	6 6	B I <del>S</del> ×,	×" ;: ::	: "	$I_x \Sigma G$	2 Source	×		
References								optional 🔉		Diffraction	data in Rigaku Cry	stal Clear .img	format					
Journal								optional >										
Conference								optional >										
								optional 👂										
Book/Report/	Chapter							optional >		Required.								
Thesis								optional 🔉	Seywords									÷ ×
Subjects								optional 🔉	· Reywords									•
🕆 Delete							🗅 Save	✓ Publish	Additional notes	+ Add anothe	rkeyword							
									y Additional holes									
						Funded	d by			Optional.								
About About	<b>Blog</b> Blog	Help FAQ	Developers REST API	Contribute		CERN		****										
Policies		Features	OAI-PMH	C <sup>®</sup> Donate		N/		****									requ	uired 💙
									Access right *	P Open A     Ø Embarg     A Restrict     A Closed     Required. Open	poed Access ted Access	ve considerably	higher visibility o	n Zenodo.				
									License *	Creative Co	mmons Attributio	on 4.0						

Required. The selected license applies to all of your files displayed in the top of the form. If you want to upload some files under a different license, please do so in two separate uploads. If you think a license is missing from the list, please inform us at info@zenodo.org

## **Crystallography Service** Discovery OK, but need more...

Q Search

REST API

Entities

OAI-PMH

Privacy policy

Terms of Use

Contact

### <isReferenceQuality>truc</isReferenceQuality> Zenoco Developers <schemaVersion>2.1</schemaVersion> <datacentreSymbol>CISTLJOE</datacentreSymbol> - <payload> <resource xsi:schemaLocation="http://datacite.org/schema/kernel-2.1 http://schema.datacite.org/meta/kernel-2.1/metadata.xsd\*> <identifier identifierType="DOI">10.5072/WDCC/CCSRNIES\_SRES\_B2</identifier> - <creators> - <creator> <creatorName>Toru, Nozawa</creatorName> Metadata formats </creator> <creator> <creatorName>Utor, Awazon</creatorName> <nameIdentifier nameIdentifierScheme="ISNI">1422 4586 3573 0476</nameIdentifier> </creator> Metadata for each record is available in several formats. The available formats </creators> <titles> include: - <title> National Institute for Environmental Studies and Center for Climate System Research Japan </title> oai datacite <title titleType="Subtitle">A survey</title> </titles> <publisher>World Data Center for Climate (WDCC)</publisher> OAI DataCite (latest schema version) - This metadata format has been <publicationYear>2004</publicationYear> <subjects> specifically established for the dissemination of DataCite records using <subject>Earth sciences and geology</subject> <subject subjectScheme="DDC">551 Geology, hydrology, meteorology</subject> OAI-PMH. In addition to the original DataCite metadata, this format contains </subjects> several other elements describing the version of the metadata, whether it is of <contributors> - <contributor contributorType="DataManager"> reference quality, and the registering datacentre. For more information about <contributorName>PANGAEA</contributorName> </contributor> this format and its schema please see the DataCite OAI schema website. - <contributor contributorType="ContactPerson"> <contributorName>Doe, John</contributorName> <nameIdentifier nameIdentifierScheme="ORCID">xyz789</nameIdentifier> This metadata format will always deliver metadata according to the latest </contributor> </contributors> available DataCite schema version - <dates> <date dateType="Valid">2005-04-05</date> <date dateType="Accepted">2005-01-01</date> See example </dates> <language>eng</language> <resourceType resourceTypeGeneral="Image">Animation</resourceType> - <alternateIdentifiers> <alternateIdentifier alternateIdentifierType="ISBN">937-0-1234-56789-X</alternateIdentifier> We recommend you harvest using the "oai\_datacite" metadata format. The </alternateIdentifiers> - <relatedIdentifiers> <relatedIdentifier relatedIdentifierType="DOI" relationType="IsCitedBy">10.1234/testpub</relatedIdentifier> <relatedIdentifier relatedIdentifierType="URN" relationType="Cites">http://testing.ts/testpub</relatedIdentifier> </relatedIdentifiers> - <sizes> <size>285 kb</size> <size>100 pages</size> oai datacite3 </sizes> - <formats> <format>text/plain</format> OAI DataCite - This metadata format has been specifically established for the </formats> <version>1.0</version> dissemination of DataCite records using OAI-PMH. In addition to the original <rights>Open Database License [ODbL]</rights> - <descriptions> DataCite v3.0 metadata, this format contains several other elements - <description descriptionType="Other"> The current xml-example for a DataCite record is the official example from the documentation. describing the version of the metadata, whether it is of reference quality, and <hr/> the registering datacentre. For more information about this format and its Please look on datacite.org to find the newest versions of sample data and schemas. </description> schema please see the DataCite OAI schema web site. </descriptions>

</resource> </payload> </oai datacite>

- <oai datacite xsi:schemaLocation="http://schema.datacite.org/oai/oai-1.0/ oai datacite.xsd">

## **Crystallography Service**

🖭 globus				👤 blaiszik -
Describe Describe Uplo	ad Verify License Complete			
Submit: Desc Please fill further information abo				
Enter appropriate subject keywor	ds or phrases below.			
Subject Keywords	self-healing	🛱 Remove Entry	cicuit	B Remove Entry
	microcapsules	â Remove Entry		+ Add More
Enter the names of any sponsors	and/or funding codes in the box below.			
Sponsors	This material is based upon work supporte Interfaces, an Energy Frontier Research C Office of Basic Energy Sciences under Aw	enter funded by the U.S. Dep	partment of Energy, Office of Science,	
Enter a description for this item ir	n the box below.			
Description	Thermomechanical failure of conductive p often impossible to repair and remains a lo approaches to restoration of conductivity r of relatively low conductivity materials. He pearly full recovery of conductance (cs. 99	ong-standing problem hinderir ely on external intervention in re, we demonstrate autonomi	ng advanced electronic packaging. Prior the form of heating or manual delivery c healing of an electrical circuit with	
Enter the name of experiment for	this item below.			
Experiment	self-healing-10wtper			
Enter the names of materials use	d in this experiment below.			
Material	Gallium	a Remove Entry	Gold	Remove Entry
	Indium		circuitboard	+ Add More
Enter the energy density used in	this experiment.			
Energy Density (mAh/g)	2000			
Estable Association (110 the set				
Enter the Argonne GUP that this	experiment was conducted under. 345-455-2543			

# In the matrix of the service of the

- Layered metadata model for description, export & packaging
- Top discovery / information layer well understood
- Leads to processing / knowledge layer – poorly understood
- Published through Dial-a-Molecule at <u>http://wp.me/p2JoQ6-xF</u> & J. ChemInf 2013, 5:52

	version="1.0"?> hema xmlns:xs="http://www.w3.org/2001/XME.Schema">
	notation>
	documentation>Change history
	documentation>
	<pre>documentation&gt;15 June 2012 - [CLB] Created from enDataDescription prototype</pre>
<td>notation&gt;</td>	notation>
<xs:an< td=""><td>notation&gt;</td></xs:an<>	notation>
<x5:< td=""><td>documentation&gt;Definitions of data types used in the manifest of the ELN item</td></x5:<>	documentation>Definitions of data types used in the manifest of the ELN item
<x5:< td=""><td>documentation&gt;</td></x5:<>	documentation>
<td>nnotation&gt;</td>	nnotation>
xxs: co	mplexType name="keywordSet">
	annotations
	xs:documentation>A list of terms
e/x5	annotation>
	sequence>
	selement name="keyword" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
	sequence>
<td>omplexType&gt;</td>	omplexType>
-	molexType nome="identifierSet">
	sequence>
	<pre>selement name="primaryLocalIdentifier" type="xs:string"&gt;</pre>
	xs: anotation>
	<pre><x:documentation>Primary string. URI. or item in any other format that enables</x:documentation></pre>
	<pre><xs:documentation>this record to be located uniquely in the originating system</xs:documentation></pre>
	xs:element>
	sielement name="otherLocalIdentifier" type="xs:string" minOccurs="0" maxOccurs="unbounded">
	exs:anotation>
	<pre><xs:documentation>[Optional] Alternative means of locating record in the originating system</xs:documentation>[Optional] Alternative means of locating record in the originating system</pre>
</td <td>xs:element&gt;</td>	xs:element>
	sielement name="accessIdentifier" type="xs:anvURI" minOccurs="0">
	exs:gnnotation>
	<pre><xs:documentation>[Optional] URI that provides a direct link to the content.</xs:documentation></pre>
	<pre><xs:documentation>If included, must be a 'linked data' URI giving open access</xs:documentation></pre>
<	xs:element>
<td>sequence&gt;</td>	sequence>
«/xs:0	omplexType>

```
<xs:complexType name="contentInformation">
  <xs:annotation>
     <s:documentation>Text that describes what the item is, with a descriptor of the</xs:documentation>
     <xs:documentation>digital type, or "undefined" if no corresponding MIME format</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="description" type="xs:string"/>
    <xs:element name="mimeType" type="xs:string" default="undefined"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="relatedID">
  <xs:annotation>
     <xs:documentation>Nature of the related information, for example, publication or related work</xs:documentation>Nature of the related information, for example, publication or related work
     <xs:documentation>Id can be any string, but DOI preferred if the related item is a publication</xs:documentation>Id can be any string.
  </xs:annotation>
 <xs:sequence>
   cxs:element name="relationship" type="xs:string"/>
    <xs:element name="id" type="xs:string"/>
 </xs:sequence>
</xs:complexType>
<xs:complexType name="relatedItemSet">
 <xs:annotation>
    <xs:documentation>Zero or more item(s) of related information</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="item" type="relatedID" minOccurs="0" maxOccurs="unbounded"/s</pre>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="contributorInformation";</pre>
     <xs:documentation>For example, Author, Funding Body, PI, institution, ...</xs:documentation>
     <xs:documentation>Plain text, but name ideally complemented by unique identifiers</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="role" type="xs:string"/>
    <xs:element name="name" type="xs:string"/>
  </xs:sequence>
</xs:complexType>
```

<xs:complexType name="contributorSet">
 <xs:complexType name="contributorSet">

# So do we really comply with RDM mandates?

 I make the assumption that raw data format and metadata is "all sorted out", but in regards to the whole picture...

- Technically oh yeah!
- Philosophically kinda
- Real world usefulness work to be done...