UK National Crystallography Service

Administration, raw diffraction data, structure factors and coordinates at the UK's National Crystallography Service

Management of people, process and data

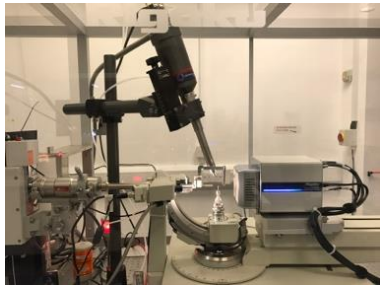
Prof. Simon Coles (s.j.coles@ncs.ac.uk)

Director, UK National Crystallography Service

UK National Crystallography Service

The NCS

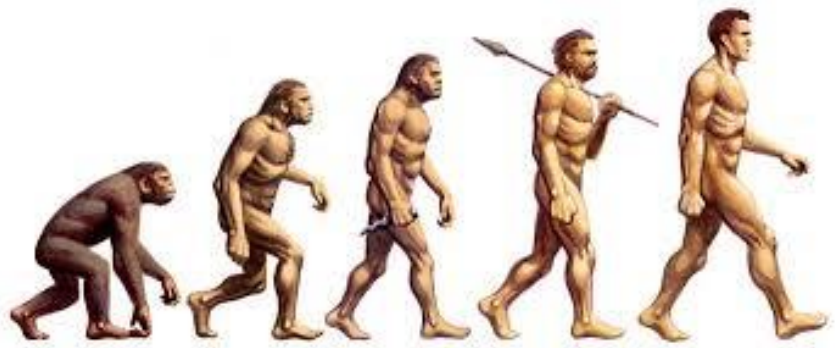
- The most powerful and highest throughput chemical crystallography laboratory
- National Research Facility, Est 1981
- Expert research staff
- Synchrotron component
- Around 100 users from UK academia
  - Chemists, crystallographers, materials scientists
  - Advanced techniques



UK National 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...



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 available to enable other researchers to understand the research and re-use potential 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.

ncs | UK National Crystallography Service

Interpreted by my research council as…

EPSRC
Engineering and Physical Sciences Research Council

Search

FUNDING | RESEARCH | INNOVATION | SKILLS | NEWS, EVENTS AND PUBLICATIONS | ABOUT US

- About us
- Service standards and policies
- EPSRC policy framework on research data
- Principles
- Scope and benefits
- Exploitation of research results and collaborative research
- Impact, timescales and support
- Responsibility for costs
- Expectations

Home / About us / Service standards and policies / 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. 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 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 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.

The policy reflects the principal UK legal provisions intended to assure public access to publicly held information, the most relevant of which to EPSRC-funded research data are contained in the Freedom of Information Act (2000) and the Freedom of Information (Scotland) Act (2002) (other relevant legislation includes the Data Protection Act 1998, the Environmental Information Regulations 2004 and the Environmental 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 data and a range of guidance is available to help universities understand their obligations (See for example advice published by Joint Information Service Committee.) **Note:** the exemptions in Scotland differ in certain important respects from those in the remainder of the UK.

See also

Access to research publications

Related links

Freedom of Information Act 2000 [ICO.]

Freedom of Information (Scotland) Act 2002 [GOV.UK]

Freedom of Information and research data: Questions and answers

UK National 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.

UK National 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.

ii. 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 (for example as available through the DataCite organisation).

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.

UK National Crystallography Service

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 scaffolds for pyrotechnic materials

L. H. Blair,[a] A. Colakel,[b] R. M. Vrcelj,[c] I. Sinclair,[b] and S. J. Coles*[a]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

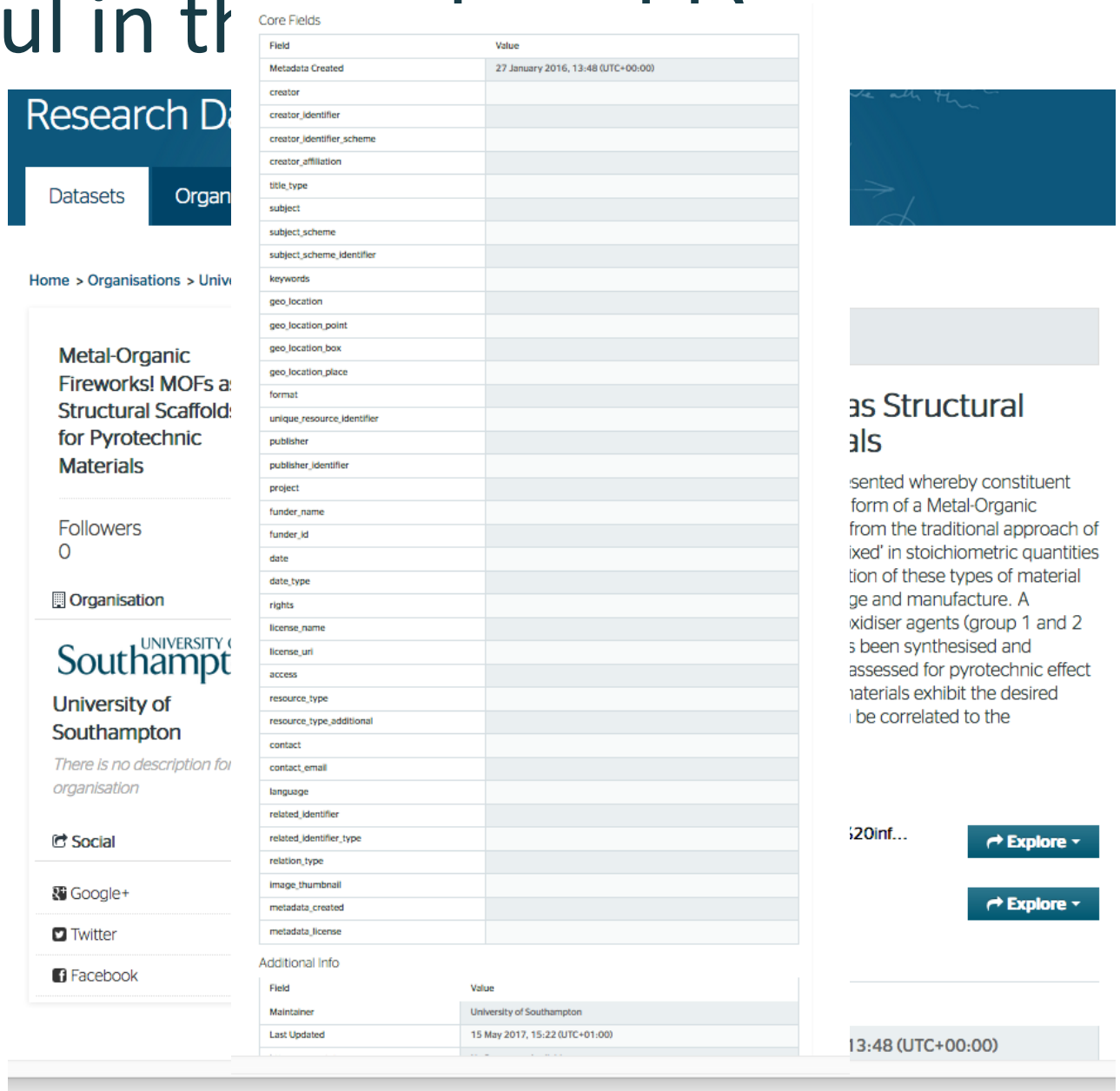
Description/Abstract

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.

DOI: 10.1039/x0xx00000x

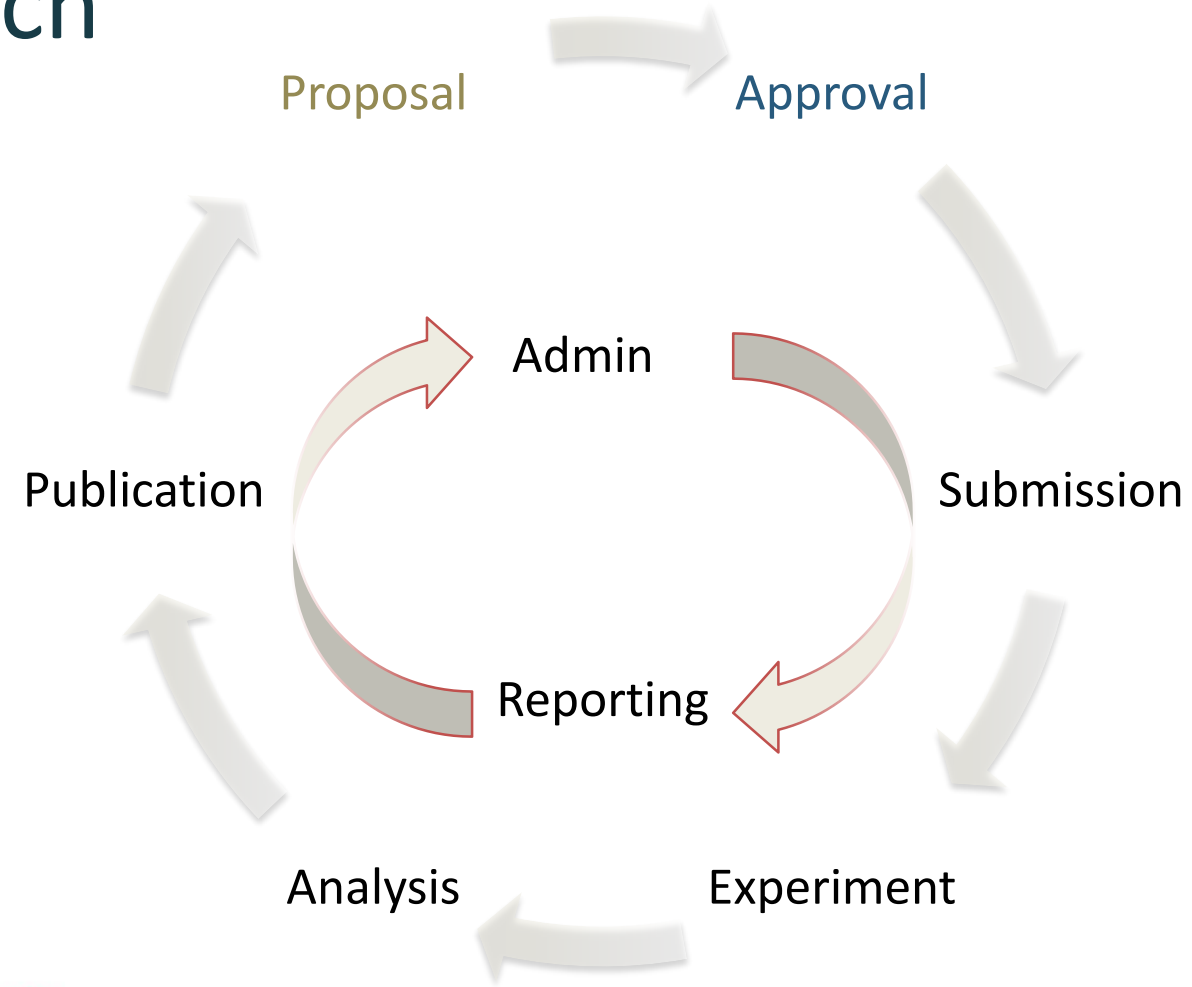
ncs | UK National Crystallography Service

Useful in th



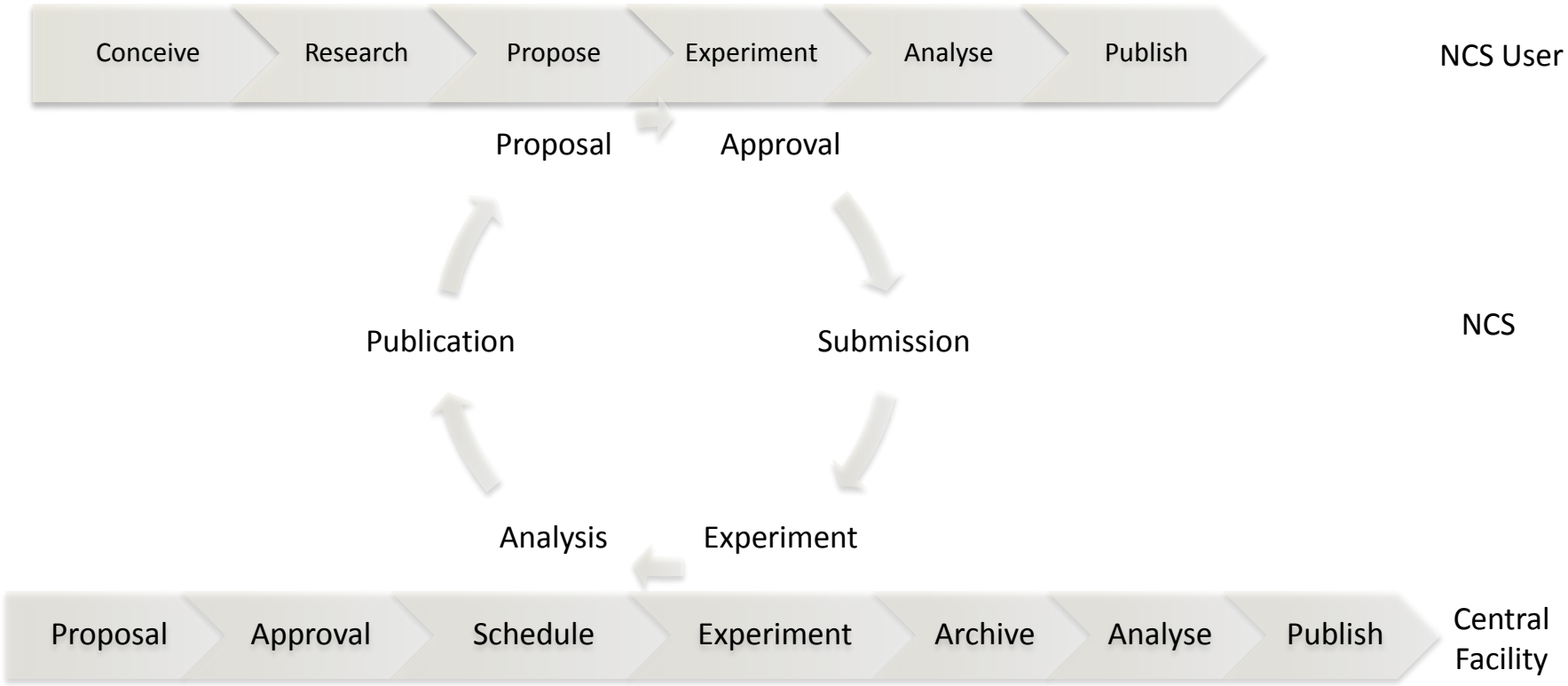
UK National Crystallography Service

NCS Portal – an Integrated Approach



UK National Crystallography Service

Operation Across Organisations



I2S2 Infrastructure for Integration in Structural Sciences

JISC

UK National Crystallography Service

Sample Information Management System

```
<?xml version="1.0"?>
<FileType Type = "Sample">
  <User Name = "Russ.Athay"/>
  <ProjectName Name = "P1" />
  <SampleName Name = "EPSIMS11" />
  <ImageDirectory Directory = "C:\Images\Rus
  <ImageTemplate Template = "EPSIMS11"/>
  <ProjectInfo>
    <DatabaseID ID = "AX3T45"/>
    <UserID ID = "AX2456"/>
    <FacilityID ID = "CAMSM33"/>
    <Priority Priority = "Medium"/>
    <SubmissionForm Form = "Standard OnLine"
    <BulkProperties Properties = "Multiple C
    <BulkMorphology Morphology = "Mostly rod
    <SolutionGoal Goal = "Publishable Struct
    <UserFormula Formula = "C9O5N3H13"/>
    <LabelingScheme Scheme = "Standard"/>
    <Sensitivity Sensitivity = "Light Sensit
    <Safety Safety = "Caustic medium"/>
    <DateProjectSubmission Date = "2012.01.0
    <DateProjectReport 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.
```

```
<Screen>
  <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"/>
  <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"
      <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"/>
    <Rmerge Rmerge = "4.30"/>
    <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>
  <Solve>
    <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="O1" Symbol="O" 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>
  </Solve>
  <History>
    <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"/>
  <Rank Rank = "135"/>
</FileType>
```

Rigaku

Application (individual)

ncs | UK National Crystallography Service

Current User Log out
Service Account
(info@ncs.ac.uk)

Application for an Allocation

Period: 1st May 2013 - 31st October 2013

Personal Details

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
United Kingdom

Funding

Funding Source: Not Funded

Local Facilities

Local Facilities: No

Reason(s) why additional facilities are required:

No local facilities, this is a new department just starting up and while we intend to buy a diffracto this will not be until next year at the earliest.

Last Allocation

	FSA Full Structure	DSO 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	5	5

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.

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

UK National Crystallography Service

Application (service crystallographer)

Last Allocation

	FSA Full Structure	DSO 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:

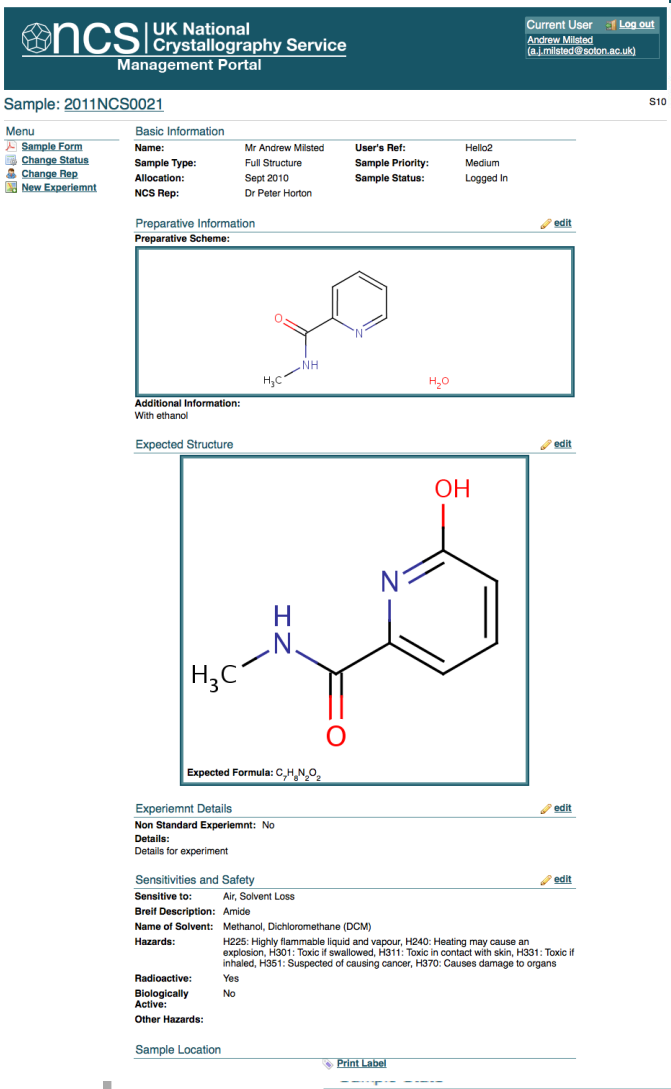
Publications:

Research Keywords:

Carbohydrate chemistry, organofluorine chemistry, total synthesis

ncs | UK National Crystallography Service

Sample submission



UK National Crystallography Service

Queued

UK National Crystallography Management Portal

Sample: 2012NCS0650

Menu

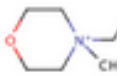
- Sample Form
- Change Status
- Change Rep
- New Sample (User)

Basic Information

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

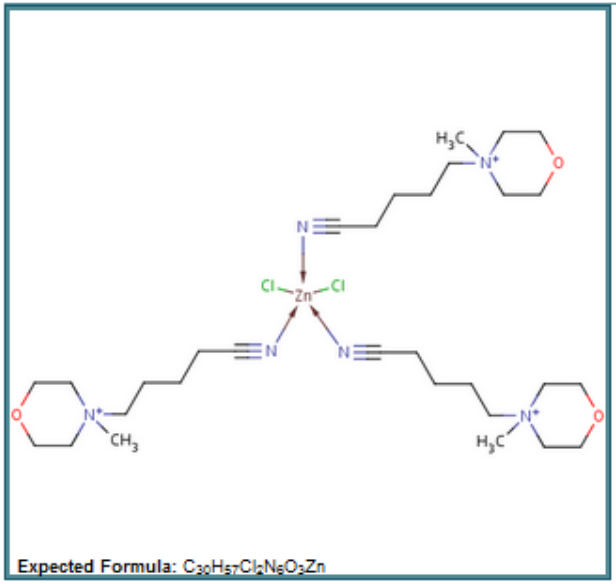
Preparative Information

Preparative Scheme:



Additional Information:
crystallized from melt, no solve

Expected Structure edit



Expected Formula: $C_{30}H_{37}Cl_2N_5O_2Zn$

Experiment Details edit

Non Standard Experiemnt:	No
Details:	

Sensitivities and Safety edit

Sensitive to:	
Breif Description:	
Name of Solvent:	
Hazards:	
Radioactive:	No
Biologically Active:	No
Other Hazards:	Unknown treat as toxic

Sample History

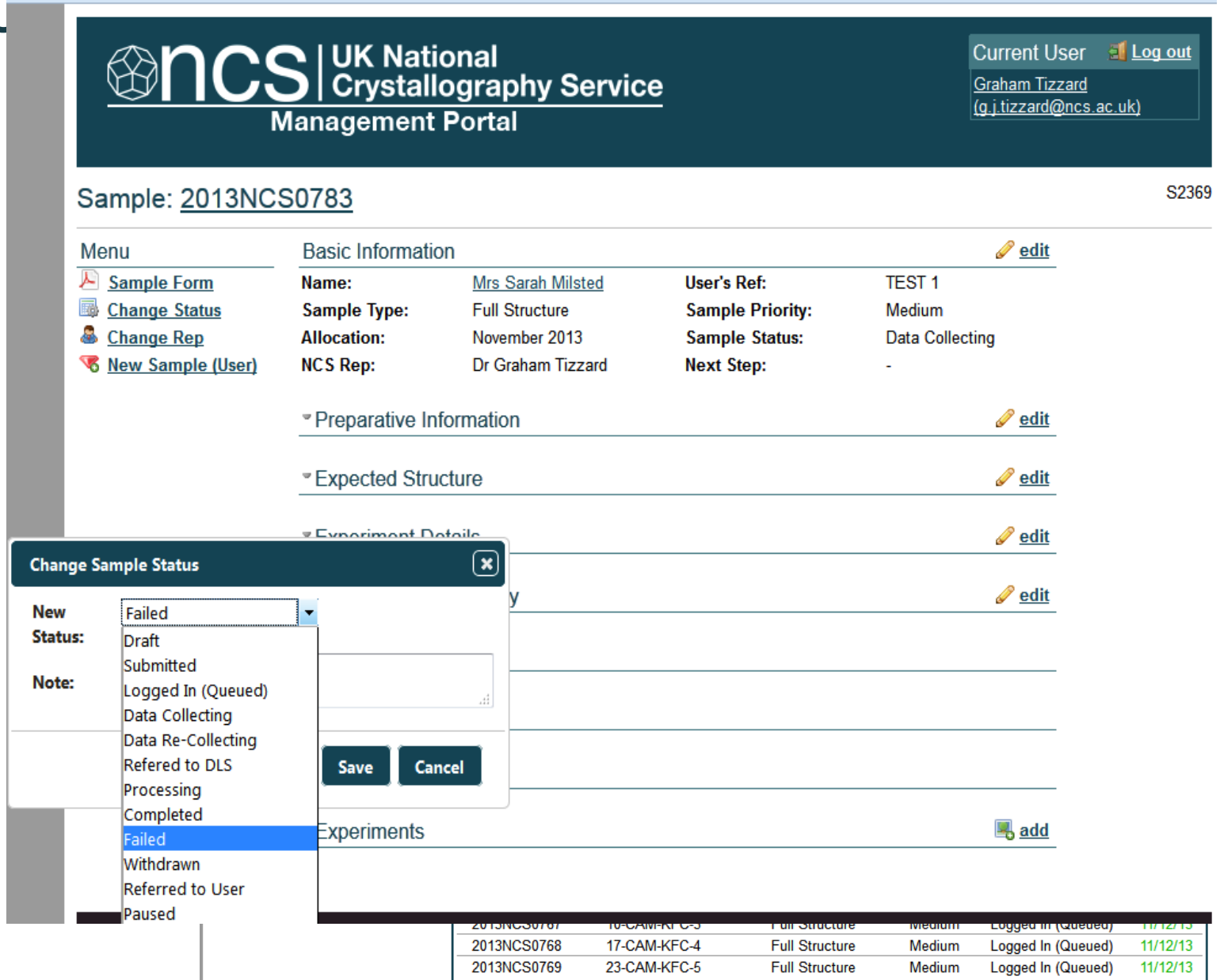
Sample Stats

Sample Location

Experiments add

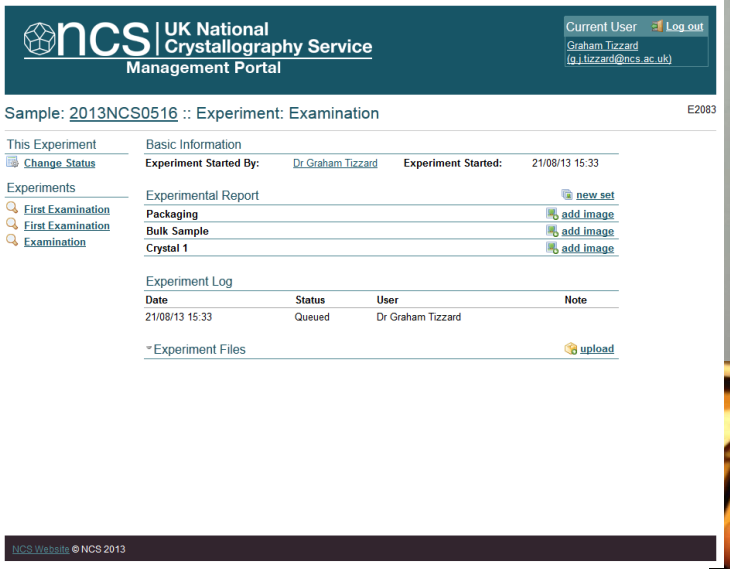
UK National Crystallography Service

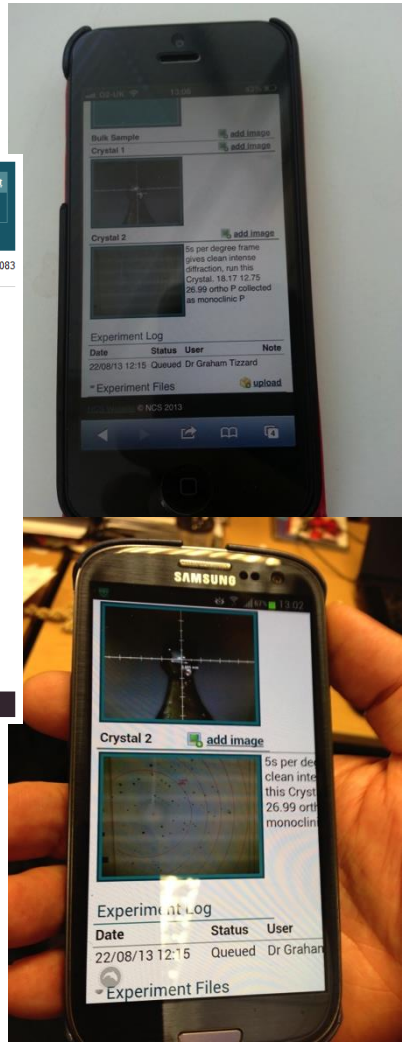
Data Collection

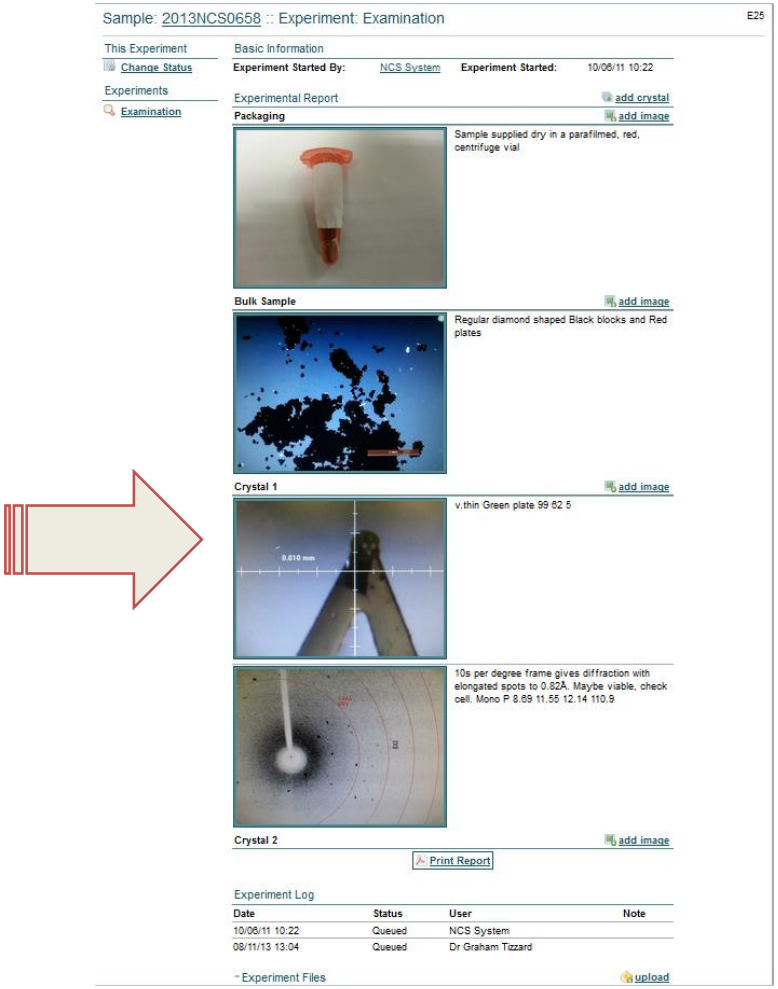


UK National Crystallography Service

Lab metadata







ncs | UK National Crystallography Service

User View

ncs | UK National Crystallography Service
Management Portal

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

Home

Menu
- New Sample
- Messages
- User Admin
- Applications
- Stats
- Stats Admin

Search

Current Allocation

Period: November 2013
01/11/13 - 30/04/14

Sample Type:	
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 Awarded)

Your Samples

Sample Code	Your Reference	Type	Priority	Status
2013NCS0775	2013sot0091	Full Structure	High	Completed
2013NCS0776	2013sot0094	Full Structure	High	Logged In (Queued)
2013NCS0777	2013sot0093	Full Structure	Medium	Logged In (Queued)
2013NCS0778	2013sot0092	Full Structure	Medium	Logged In (Queued)
2013NCS0779	2013sot0090	Full Structure	Medium	Logged In (Queued)
2013NCS0780	2013sot0089	Full Structure	Medium	Logged In (Queued)
2013NCS0801	2013SOT0095	Full Structure	Medium	Logged In (Queued)

ncs | UK National Crystallography Service
Management Portal

Sample: 2013NCS0775

Menu
- Sample Form

Basic Information

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 Structure
- Experiment Details
- Sensitivities and Safety

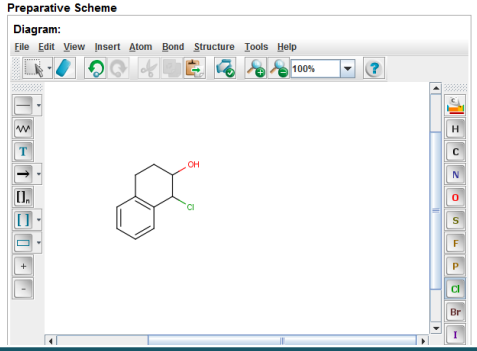
ncs | UK National Crystallography Service
Management Portal

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

Edit Sample: S2400 1 of 4

Preparative Scheme

Diagram:



ncs | UK National Crystallography Service
Management Portal

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

Edit Sample: S2400 4 of 4

Sample Sensitivty

Sensitive To:
- [ ] Air
- [ ] Water
- [ ] Light
- [x] Solvent Loss
- [ ] Other

Sample Safety

Brief description of sample: (e.g. organic amide)

Name of Solvent:
- [ ] No Solvent
- [x] Acetonitrile
- [ ] Dichloromethane (DCM)
- [ ] Diethyl ether
- [ ] Dimethyl Sulfoxide
- [ ] Dimethylformamide
- [ ] Ethanol
- [ ] Hexane
- [ ] Methanol
- [ ] Trichloromethane (Chloroform)
- [ ] Other

(if in mother liquor)

Is Sample Radioactive?: No

Is Sample Biologically Active?: No

Hazard Phrases:
- [ ] H200: Unstable explosive
- [ ] H201: Explosive; mass explosion hazard
- [ ] H202: Explosive; severe projection hazard
- [ ] H203: Explosive; fire, blast or projection hazard
- [ ] H204: Fire or projection hazard
- [ ] H205: May mass explode in fire
- [ ] H316: Causes mild skin irritation
- [ ] H317: May cause an allergic skin reaction
- [ ] H318: Causes serious eye damage
- [ ] H319: Causes serious eye irritation
- [ ] H320: Causes eye irritation
- [ ] H330: Fatal if inhaled

UK National Crystallography Service

Data Availability

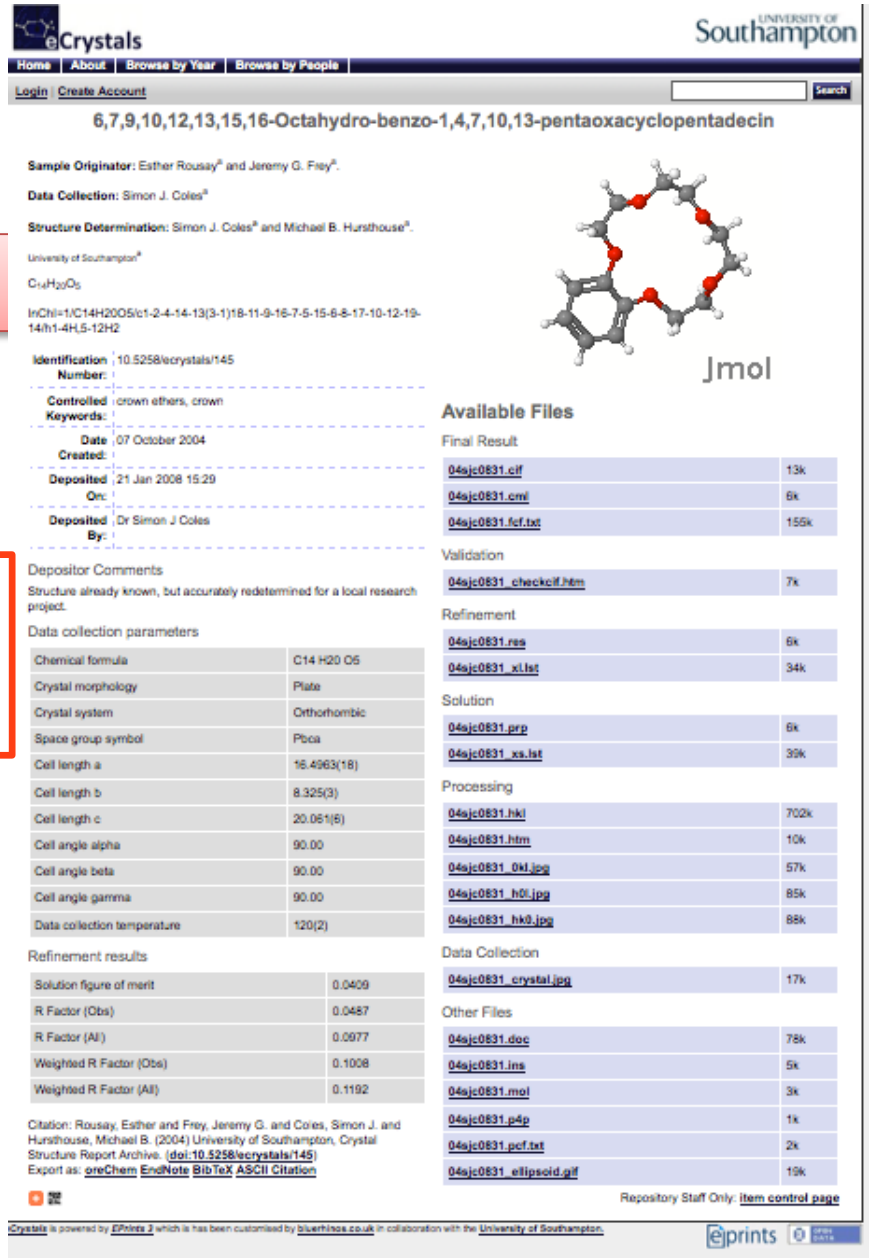
http://ecrystals.chem.soton.ac.uk

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

DataCite
Helping you to find, access, and reuse data

eCrystals

JISC



UK National 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 http://www.ukoln.ac.uk/projects/ebank-uk/schemas/
- Licensing

UK National 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

EPSRC
Engineering and Physical Sciences Research Council

UK National Crystallography 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?

UK National Crystallography Service

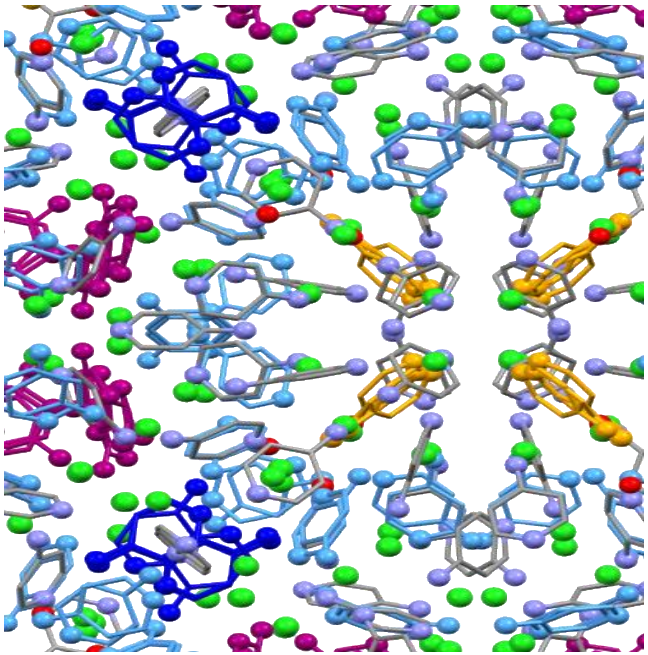
The case for 'publishing' (Chem Cryst) raw data?

- 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

UK National Crystallography Service

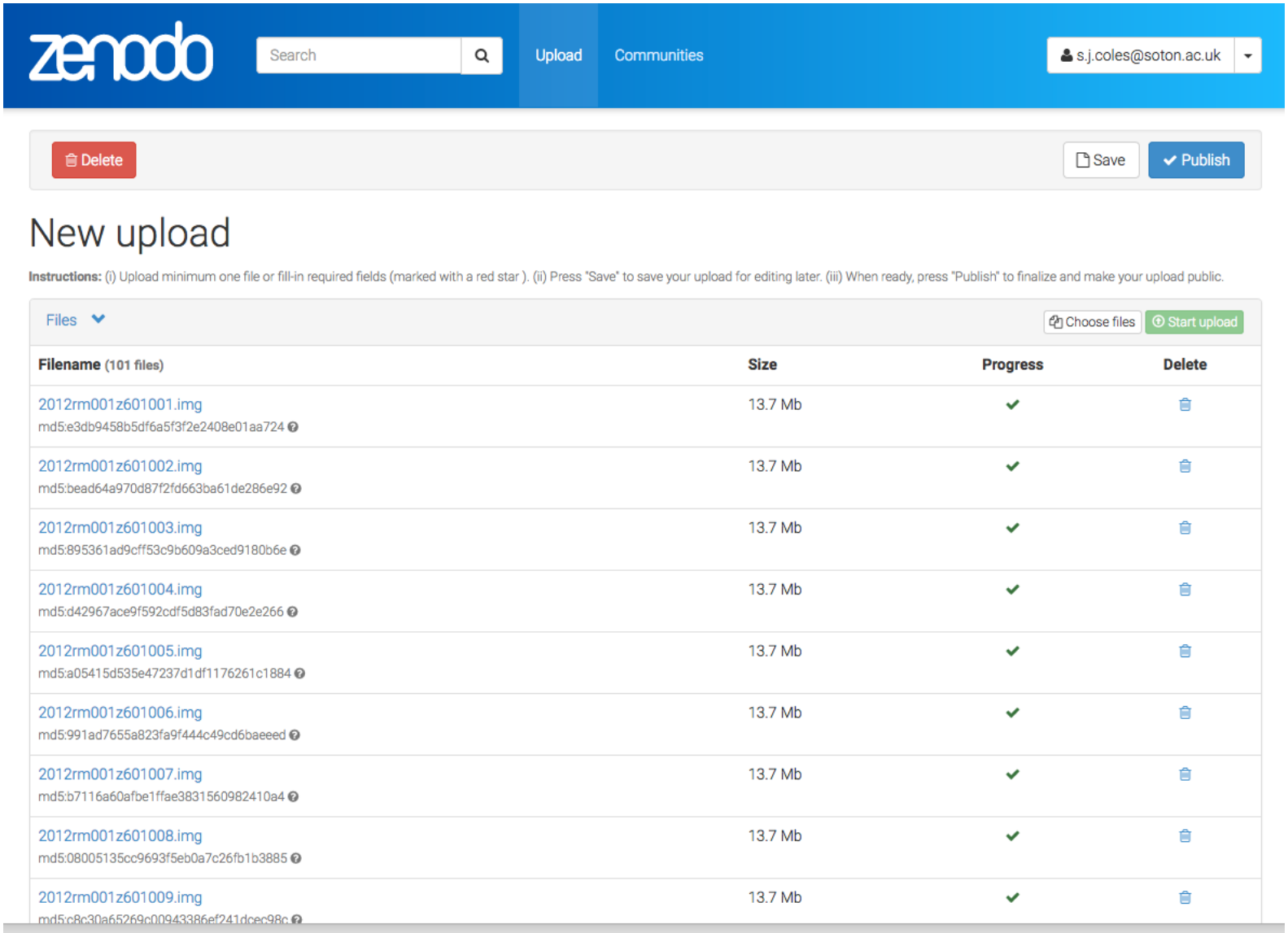
A real example…

- 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?



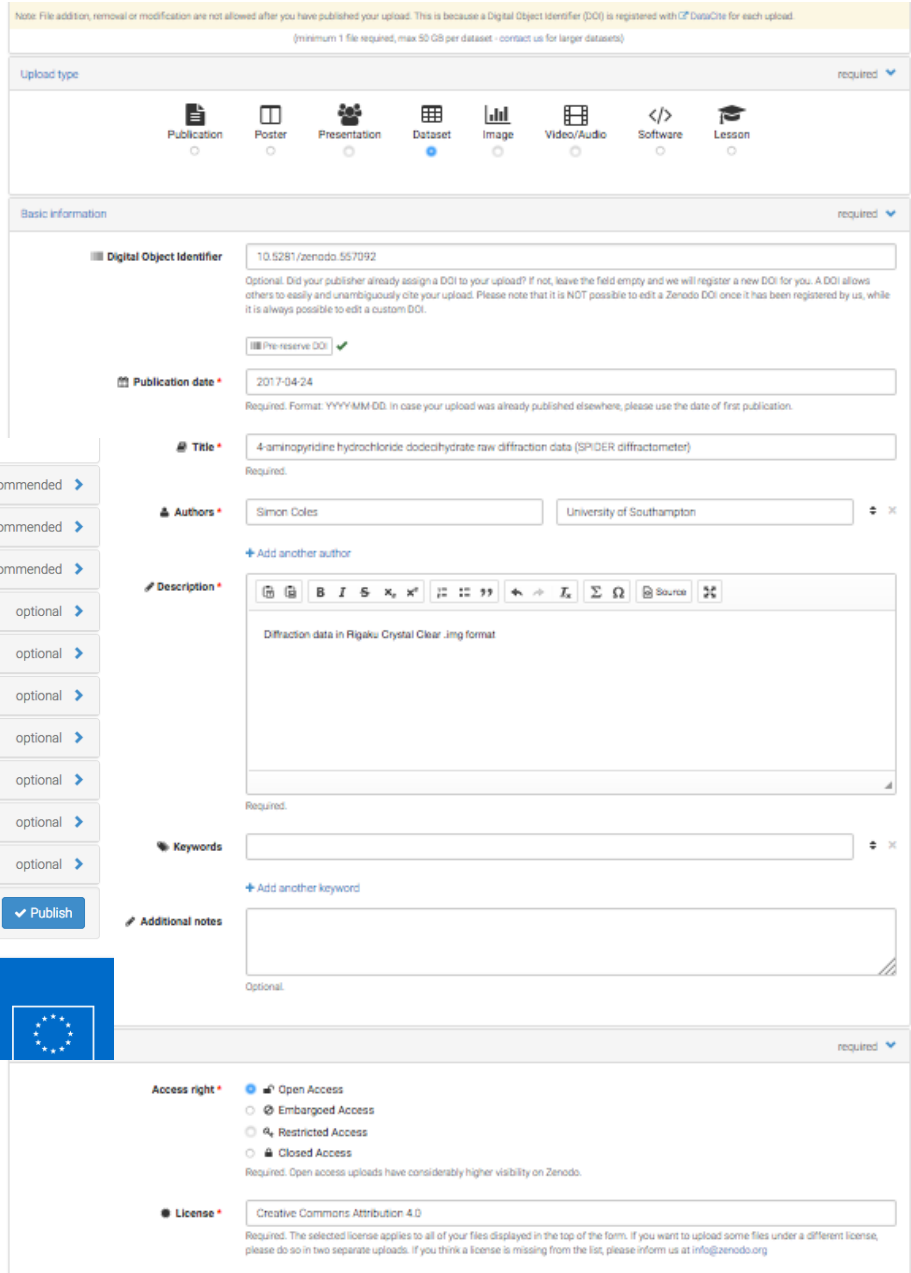
UK National Crystallography Service

Supported, Open, Large data volumes



UK National Crystallography Service

Plenty of metadata?



UK National Crystallography Service

Discovery OK, but need more…

zenodo | Developers

Search

REST API
Entities
OAI-PMH
Base URL
Resumption tokens
Rate limit
Metadata formats
Sets
Update schedule
Changes

Privacy policy
Terms of Use
Contact

Metadata formats

Metadata for each record is available in several formats. The available formats include:

`oai_datacite`

OAI DataCite (latest schema version) — This metadata format has been specifically established for the dissemination of DataCite records using OAI-PMH. In addition to the original DataCite metadata, this format contains several other elements describing the version of the metadata, whether it is of reference quality, and the registering datacentre. For more information about this format and its schema please see the DataCite OAI schema website.

This metadata format will always deliver metadata according to the latest available DataCite schema version.

See example

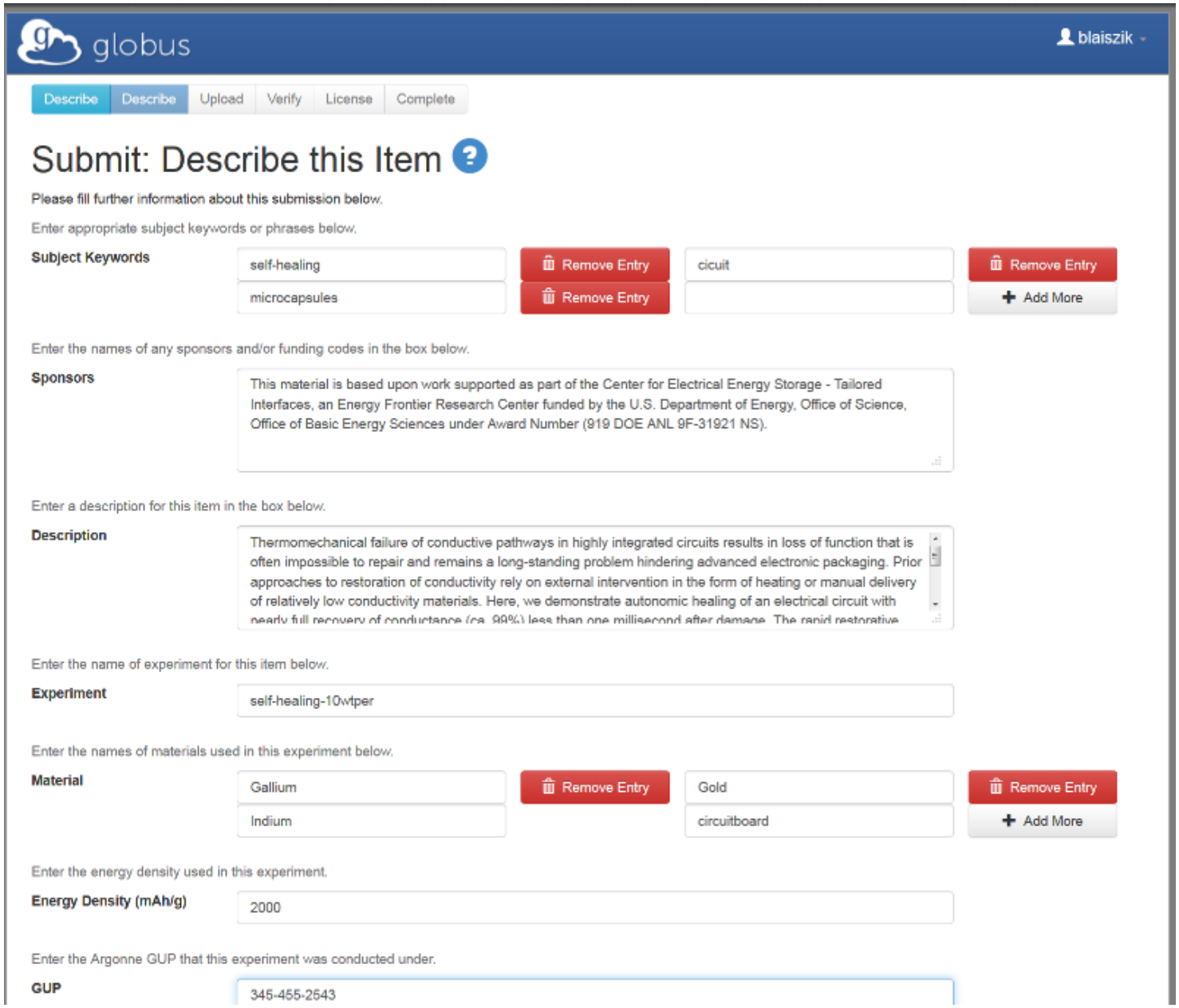
We recommend you harvest using the "oai_datacite" metadata format. The format contains the most complete metadata and is our primary supported export format.

`oai_datacite3`

OAI DataCite — This metadata format has been specifically established for the dissemination of DataCite records using OAI-PMH. In addition to the original DataCite v3.0 metadata, this format contains several other elements describing the version of the metadata, whether it is of reference quality, and the registering datacentre. For more information about this format and its schema please see the DataCite OAI schema web site.


```
<!-- -->
<oai_datacite xsi:schemaLocation="http://schema.datacite.org/oai/oai-1.0/ oai_datacite.xsd">
  <isReferenceQuality>true</isReferenceQuality>
  <schemaVersion>2.1</schemaVersion>
  <datacentreSymbol>CISTI.JOE</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>
        </creator>
        <creator>
          <creatorName>Utor, Awazon</creatorName>
          <nameIdentifier nameIdentifierScheme="ISNI">1422 4586 3573 0476</nameIdentifier>
        </creator>
      </creators>
      <titles>
        <title>
          National Institute for Environmental Studies and Center for Climate System Research Japan
        </title>
        <title titleType="Subtitle">A survey</title>
      </titles>
      <publisher>World Data Center for Climate (WDCC)</publisher>
      <publicationYear>2004</publicationYear>
      <subjects>
        <subject>Earth sciences and geology</subject>
        <subject subjectScheme="DDC">551 Geology, hydrology, meteorology</subject>
      </subjects>
      <contributors>
        <contributor contributorType="DataManager">
          <contributorName>PANGAEA</contributorName>
        </contributor>
        <contributor contributorType="ContactPerson">
          <contributorName>Doe, John</contributorName>
          <nameIdentifier nameIdentifierScheme="ORCID">xyz789</nameIdentifier>
        </contributor>
      </contributors>
      <dates>
        <date dateType="Valid">2005-04-05</date>
        <date dateType="Accepted">2005-01-01</date>
      </dates>
      <language>eng</language>
      <resourceType resourceTypeGeneral="Image">Animation</resourceType>
      <alternateIdentifiers>
        <alternateIdentifier alternateIdentifierType="ISBN">937-0-1234-56789-X</alternateIdentifier>
      </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>
      </sizes>
      <formats>
        <format>text/plain</format>
      </formats>
      <version>1.0</version>
      <rights>Open Database License [ODbL]</rights>
      <descriptions>
        <description descriptionType="Other">
          The current xml-example for a DataCite record is the official example from the documentation.
          <br/>
          Please look on datacite.org to find the newest versions of sample data and schemas.
        </description>
      </descriptions>
    </resource>
  </payload>
</oai_datacite>
```

ncs | UK National Crystallography Service



ncs | UK National Crystallography Service

eInItemManifest

- 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 http://wp.me/p2JoQ6-xF & J. ChemInf 2013, **5**:52


```xml
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:annotation>
  <xs:documentation>Change history</xs:documentation>
  <xs:documentation>==============</xs:documentation>
  <xs:documentation>15 June 2012 - [CLB] Created from enDataDescription prototype</xs:documentation>
</xs:annotation>

<xs:annotation>
  <xs:documentation>Definitions of data types used in the manifest of the ELN item</xs:documentation>
  <xs:documentation>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~</xs:documentation>
</xs:annotation>

<xs:complexType name="keywordSet">
  <xs:annotation>
    <xs:documentation>A list of terms</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="keyword" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="identifierSet">
  <xs:sequence>
    <xs:element name="primaryLocalIdentifier" type="xs:string">
      <xs:annotation>
        <xs:documentation>Primary string, URI, or item in any other format that enables</xs:documentation>
        <xs:documentation>this record to be located uniquely in the originating system</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="otherLocalIdentifier" type="xs:string" minOccurs="0" maxOccurs="unbounded">
      <xs:annotation>
        <xs:documentation>[Optional] Alternative means of locating record in the originating system</xs:documentation>
      </xs:annotation>
    </xs:element>
    <xs:element name="accessIdentifier" type="xs:anyURI" minOccurs="0">
      <xs:annotation>
        <xs:documentation>[Optional] URI that provides a direct link to the content.</xs:documentation>
        <xs:documentation>If included, must be a 'linked data' URI giving open access</xs:documentation>
      </xs:annotation>
    </xs:element>
  </xs:sequence>
</xs:complexType>
```



```xml
<xs:complexType name="contentInformation">
  <xs:annotation>
    <xs: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:documer
    <xs:documentation>Id can be any string, but DOI preferred if the related item is a publication</xs:docume
  </xs:annotation>
  <xs:sequence>
    <xs: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"/>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="contributorInformation">
  <xs:annotation>
    <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:annotation>
```

UK National Crystallography Service

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…