

### Diffraction Data in Context

Metadata approaches

Simon Coles (s.j.coles@soton.ac.uk)

Director, UK National Crystallography Service

# What can structures be used for? School of Chemistry

- We don't capture valuable information from every part of the "study"
- Good crystallographic practice requires information about the experiment not covered wholly by CIF
- Cant predict how a structure might be used in the future measure AND CATALOGUE everything you possibly can!

• Data management is key: its all about context and linking...



### Why bother?

- Scientific rigour, reproducibility, long-term preservation
- Enable reuse
  - Amongst collaborators
  - For teaching & software development
  - New science
- Reanalysis as new techniques develop

• Some case studies/examples...

### **Operating a Service**

### Southampton

School of Chemistry



### In-house Management

Average publication lead-in time = 3 years

Publication

2015

Analysis

♦ 2000

2010

1250

2000

♦ 750

1990

80% not published
 Proposal

**Annual Data Collection Capability** 

¢<sub>5</sub>250

1980

2500 2000

1500

1000

500

0

1960

1970

# School of Chemistry

BUICS	UK National Crystallogra	aphy Service	Ú.		Cancel User Index Down	1040 1040
ome						
eta .	Your Sungkys					
New Sample			The Sample	•		
Messages	Township to our of					
Applications	Sample Cash	Day Balances	Tax	Printe	Thereit.	ites.
True Transit	BYUNCON200	1111.000ku	Ful Distance	Lim	Processing	
ierth .	SALENCERSIA.	Octor	Ful Strates	6.000	Date Cathering	parent.
	And	DOJe	Full Detailes	100	Date Cattering	- 275
T THE POOL OF	BYCACIDETT.	687070-2	Ful Drahen	Manine	Bala Comunity	Santo.
0	2013ACCOM#17	100/100	Fait Deartain	i.m	segged in (Descel)	1607-
Associat	30134CSRAM	INTERNAL I	Full Disation	High	Halwart to DLB	CANAL CONTRACT
	bid indication	100.001	Full Description	- Martine	Logard in Channel	- and the lot
	20-ChClimbty	Me12108	Ful Displays	Madlant	Lagest to Chemical I	10.00
	AD CHECKED PS	APR01/1	Fut Deatures	Volum	Logard to (Element)	19.09/
	NONCIMUM	period care	Tal Deater	Maillion	Laged in (General)	19.04
	2010ACTINUSA	Monthly Providence	Full Stration	Madian	Legged N (Graned)	112
	and a state of the	101111	1.00.000.000		competer in the second	_
		N		ġi.	1	
		IN.				
missio	n (			=	GE.	
	24.44				ALL.	
		the second se				
	Contraction of the local division of the loc	Contraction of the				
		121				
	-	14			AL.	
			-			
			6			
			6			

Approval

Experiment

S

Admin

Reporting

UK National Crystallography Service

### Working Across Facilities







### Availability & Reuse

- Cant publish everything in the traditional way
- WWW alternative route
- Data needs to 'speak for itself'
- Provide enough raw data & metadata so that
  - Error understood
  - Reproducible
  - Reanalysis possible
  - Reuse facilitated

THE ADDLE BOWINED ATTACK	Browse by Propin	-	
In Create Account			Start
6,7,9,10,12,13,1	5,16-Octahydro-benz	to-1,4,7,10,13-pentaoxacyclop	pentadecin
mple Originator: Esther Rousey <sup>a</sup> a	nd Jeremy G. Frey <sup>e</sup> .	al la	24
ta Collection: Simon J. Coles <sup>a</sup>			
ructure Determination: Simon J. D	les <sup>#</sup> and Michael B. Hursthouse*.		<b>S</b>
venety of Southempton <sup>®</sup>			1
eH20O5		Veril .	1
CHI=1/C14H20O5/c1-2-4-14-13(3-1) h1-4H,5-12H2	8-11-9-16-7-5-15-8-8-17-10-12-19		
entification 10.5258/ecrystals/145 Number:		3.0	Jmol
Controlled crown ethers, crown		Available Files	
Date 07 October 2004		Final Result	
Crepted		04ajc0831.clf	13k
Ore Ore		04sjc0831.cml	64
Deposited Dr Simon J Coles		04sjc0831.fcf.txt	156k
Dy. !		- Validation	
positor Comments	redetermined for a incel research	04ajc0831_ctwckcif.htm	78
ject.		Refinement	
ita collection parameters		04sjc0831.res	61
nemical formula	C14 H20 O5	04sjc0831_xList	34k
vatal morphology	Plate	Solution	
yatal system	Orthorhombic	04ajc0831.prp	Bk
ace group symbol	Peca	04ajc0831_xa.lat	39k
eil length a	18.4963(18)	Processing	
el lengos o	8.325(3)	Oddesing Oddesing	7004
na mondan se	20.061(6)	04sic0631.htm	104
er unger alpha	90.00	04sic0631_0kLing	578
of angle bits	00.00	04sic0831 h0Lipg	85×
ate collection tomotrat m	120/21	04sjc0831_hk0.jpg	B8k
Engrand marine	and al	Data Collection	
nine rent result	0.0409	04sjc0831_crystal.jpg	17k
Factor (Obs)	0.0487	Other Eiler	1.04
Factor (All)	0.0077	0444-0831 fac	785
weithed R Factor (Ots)	0.1008	044ic0631 ins	/of
eichted R Factor (All)	0.1192	04sic0831 mol	39
		Odeir0831 nén	
ation: Rousay, Esther and Frey, Jen rathouse, Michael B. (2004) Univers	my G. and Coles, Simon J. and ity of Southempton, Crystal	04440831 and ret	24
ucture Report Archive. (doi:10.5258 port as: oreChem EndNote BibTeX	ASCII Citation	Odsir0831 ellipsoid ail	104
and a second second second second		- alcoss _ empositedin	144

School of Chemistry



### Supporting future reanalysis

• Model unclear – need to come back to the raw data later

- (Z' = 30, 60, 120 or 360?!?)

- What's going on in the system unstudied non-Bragg regions
  - Disorder
  - No long range order
  - Solvent effects
  - Phase transitions



School of Chemistry

Sout



### Approaches in the Future



- What is structural information and what is background? Contributions to a reflection can only really be understood when we have a structural model! Second cycle of extracting structure factors once the model is complete...?
- The next generation will be addressing the information that we collect that is not Bragg scattering.
- Cant necessarily just do the experiment again
  - Interesting samples synthesised now may not be available in the future...
  - Facilities time difficult to get
  - Experiments performed on current equipment which may not be easily available in the future



# But what about process and context?

Detail behind the diffraction experiment itself.

Motivation behind the study as a whole.

Results in terms of the bigger scientific picture.

### Getting Mobile in the Lab

# School of Chemistry

• Every crystallographic experiment has 'unstructured' data



ample: 2013NC	S0516 :: Experimen	t: Examinati	on		1
is Experiment	Basic Information				
Change Status	Experiment Started By:	Dr. Graham Tizz	and Experiment Started:	21/08/13 15:33	
periments	Experimental Report			The new set	
Eirst Examination	Packaging			H add image	
<b>First Examination</b>	Bulk Sample			H add image	
Examination	Crystal 1			🖐 add.image	
	Experiment Log				
	Date	Status	User	Note	
	21/08/13 15:33	Queued	Dr Graham Tizzaid		
	*Experiment Files			boolqu 👸	

COMMUNICONCE 2013

### Getting Mobile in the Lab



### Package & Sample Tube









Bulk Sample & Manipulation

> Mounted Sample

> > Trial Diffraction Pattern







### Getting Mobile in the Lab

# School of Chemistry















13

(more)



	r' BlogMyData
HIGEM Blog	
esting POI works	Seath 14
76.August 2010 @ 13.40	This Post
latyge statickky@ithPaint egionafiniarest ROHTI21.794815 - 38.195867)) master HiCZW, KBITA, NOHTI21MEMI	Permakerik GRI URI caberi
arlablest temp Investigen: CO-1.4 anlablestandandmaine: COT/ Temperature.	Revisions Export XVV. (Mith Fired
www.exemus.cey. bec POLYGON(13339375 -35489785,95425 -15489785,91425 8.580235.13	319175 PMC Image
540235,13.359375 - 55.6887802 nr: 1910 4126	This Blog
ime: 2010-06-16700-00-00.0002 alendarsystem: 160_day levation: 1.050000100734863	New Post Bog Settings Timeline View
evalionumits: m	Exceptions
Sadina' test sie he MagingSada - Md2W X015 Maddig waars - 2021 Tenperateer Take 2010-so tetre of sa sing Depti: 5.0004025474281 +	October 2000 (3) September 2000 (2)
No. of the local division of the local divis	August 2010 (M
100 C	sections abriddat/ 0.38
	Viztype
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	RaticMap (5) SectoMapNithPore (2) Animation (6)
an sg	Dataset
	HIGH YEAR (28) HIGH YEAR (28)
0	Variablestandardname
the fact that the state of the	OCh) Temperature, #8 OCh) Selvety, @5 Denisty (\$5
1. THE	Calendarsystem
Testig HD webs	202 208 (23)
eEC aims to promote 4-Scenie methods in the eminimmental iscence community by de enconstrator projects with colladorators in academia, government adjectives ing, the Ne enconstrator. These projects will showcase the givest potential of 4-Science to be a yaellu	wioping Elevationumics r. Offick) W (13) method
car environmental scientists in many disciplines can use in their everyday research. To one previously official and time-community, such as sharing and working with very la its, can be made much more efficient.	sks that Bevationpositive ope data (13)
Anached Files	Tools
	Show/Hide QR Cade Show/Hide Kess

School of Chemistry

d Connert uner Abdrew Hitstel   Loo Gut Deshboard   the	10	School o	f Chemistry
🕲 Laser Environment - Mozilla Firefox		ų ču:	Al Eugs ( Holp
<u>Eile E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp	0		✓ BlogMyData
C X 🟠 http://xray.orc.soton.ac.uk/laser_env	🗟 🏠 र 🔽 Google 🔎		/ Dioginij Duna
Most Visited P Getting Started S Latest Headlines			
Laser Environment Environmental data from the laser	http://xray.orc.soton.ac.uk/data/2064.html		Seeth Sector
<u>46:1047 - 2009-7-20</u>			
21st July 2009 @ 00:00	46:1047 - 200	)9-7-20	
Auto blog	[m		
Data collected from 46:1047 on 2009-7-20	36         1:00         2:00         3:00         4:00         5:00         6:00         7:00         8:00         9:00         10:0011:0012:00	13:00 14:00 15:00 16:00 17:00 18:00 19	:00 20:00 21:00 22:00 23:00
<u>Auto Blogger   View Source   46:1047   Comments (0</u>		and in the state of the state o	and the second second
46:1047 - 2009-7-19 20th July 2009 @ 00:00			
Auto blog	28		
Data collected from 46:1047 on 2009-7-19			
<u>Auto Blogger   View Source   46:1047   Comments (0</u>	24	Manne	
<u>46:1047 - 2009-7-18</u> 19th July 2009 @ 00:00			
Auto blog			
Data collected from 46:1047 on 2009-7-18	Chamber (46:1047/3/typeKthermocouple)	Black	
Auto Blogger   <u>View Source</u>   <u>46:1047</u>   <u>Co</u> mments (0	Plant Room (46:1047/0/typeKthermocouple)	Red	
	Shelf top (46:1047/2/typeKthermocouple)	Green	
<u>46:1047 - 2009-7-17</u>	Laser box (46:1047/1/typeKthermocouple)	Blue	
18th July 2009 @ 00:00			

a Corrent	Deshboard i Help	,
er En Edit St Visit	So UltraFast Xray Group	uthampton
10 10	<< Next Post Previous Post >>	Search 🔍
o blo a co	Reconstructing the new data	This Post
10	Just a quick post to show some pictures from the reconstructions of the latest data and update everyone on progress.	URI Label Revisions Export:
July blc	Looking at data taken on 2nd August – the old 2um sample taken with a single wavelength. The diffraction patterns look v clean, with good detail right out to the edge – haven't worked out the resolution yet.	XML (With Files) PNG Image This Blog
10	Tried to get Ben's CDI routine working, but it broke badly, and I couldn't fix it immediately, so went back to my old routines (from 2 years ago! can't believe it's been that long) and got 'hio6.m working. This is running on Boxer under Windows, but not using anything clever.	New Post Blog Settings Timeline View Exhibit View Export Blog
July		Archives
10	<u>1000_iterations1.jpg</u> This one is a very plain HIO reconstruction, 1000 iterations after binning the data to 512x512. Shrinkwrapping every 100 (I think - notes at work right now). You can see that it reconstructs beautifully, and this happens every time - no need for multiple phases to get it to work.	November 2011 (18) October 2011 (40) September 2011 (38) August 2011 (40) July 2011 (31) June 2011 (35)
July	(Once I'd remembered to background-subtract properly, of course - before that it all failed	May 2011 (23) April 2011 (17)

# Corrent user: Andrew Histed | Log Out

**Dashboard** / Help

School of Chemistry

### ど Laser En Eile Edit

### Transformation of plasmid JRH4712/66 into BW25141 by electroporation

11th December 2006 @ 14:31

Transformations were set up according to the following protocol:

LB Ampicillin arabinose plates and SOC medium were warmed to 37 °C briefly before the arabinose plates were spread with X-glu (80 µL, 1:1 X-glu and LB) and allowed to continue warming.

BW25141 cells, plasmid JRH4712/66, p042, and electroporator cuvettes were cooled on ice. Items were added to the cuvettes as follows

-	1	+ve ctrl	-ve ctrl
BW25141	40 µL	40 µL	40 µL
plasmid 4712/66	4 µL	0 µL	0 µL
p042	0 µL	4 µL	0 µL

Cuvettes were electroporated at 1.75 kV, immediately had SOC medium (950 uL) added and the transformant transferred to eppendorf. The transformants were incubated at 37 °C for one hour with shaking. The transformants were diluted 1 in 20 with LB and 100 µL added to LB amp arabinose plates and incubated at 37 °C overnight.

Data



### Archives

January 2007 (24) December 2006 (11) November 2006 (5)

### Sections

beta-galactosidase preparation and assays (18)

### Test digestions to check the activity of two batches of EcoRI and Ncol 22nd January 2007 @ 11:57

Lab Book Ref: jrh4712-89 Sample Parent: jrh4712-80\_blue Sample Parent2: irb4712-80\_white Digestions were set up as follows:

-	1	2	3	4	5	6	7	8	9	10	11
4712/80 blue	8 µL	-	-	8		-	-	8 µL	•	-	-
4712/80 white	-	8 µL	-	-	8 µL	- /	-	-1	8 µL	- 1	-
p042	÷	-	S µL	-	-	S µL	5 µL	-1.	+	5 µL	S µL
water	7.5 µL	7.5 µL	10.5 µL	7.5 µL	7.5 µL	10.5 µL	10 µL	7.5 µL	7.5 µL	10.5 µL	10 µL
EcoRI buffer	2 µL	2 µL	2 µL	-	-	-	2 µL	2 µL	2 µL	2 µL	2 µt
NES buffer 4	-	-	-	2 µL	2 µL	2 µL	-	-	5	7	•
8SA	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
EcoRI (a)	0.5 µL	0.5 µL	0.5 µL	-	-	-	0.5 µL	-		-	-
Ncd	-	-	-	0.5 µL	0.5 µL	0.5 µL	0.5 µL	-	•	-	0.5 µL
EcoRI (b)	-	-	-	-	-	-	-	0.5 µL	0.5 µL	0.5 µL	0.5 µL

EcoRI (a) assay date 2/05 EcoRI (b) assay date 7/05

Digestions were incubated in a waterbath at 37 °C for 3 hours.

### Southampton

Ad Elega I Halts

Data

January 2007 (24) December 2006 (11) November 2006 (S)

### Sections

Archives

beta-galactonidase preparation and #83.89K (18) Nets-plucuronidase (18) Data (Formatting) (13 Software discussions [2] Starting materials and responds (1)

### Lab Book Ref

3044712-63(1) JRH#712-64 (2) JRH4712-66 (1) 204712-76(1) jeh4712-77 (1) jih#712-78 (1) jrh4712-86 (1) yh4712-01 (1) jih4712-83 (1) yh4712-83 (15 JH4712-84 (1) )mA712-85(1) 4712-88 (1) wb4712-89(1) 4712-86 (1) 284712-87(1) 4712-50+(1)

### Product

)/h4712-74 (1) 3/64712-76 (1) Jih4712-76a [1]

Jennifer Hale | Beta-glucuronidase | Comments (3)

jrh4712-76a(1) jrh4712-77 (1) ich (717, 70 (1)







EcoRI (a) assav date EcoRI (b) assay date

Digestions were incl

References:

Attached Files

doi: 10.1016/0040-4020(80)80102-5

**BO Linked Posts** 

Laura White | View Source | Procedures | Comments (0)

For data files, refer to Synthesis of ethyl 2,5-dimethyl-1-phenyl-1H-pyrrole-3-carboxylate (LMW

8-1) and Synthesis of ethyl 2-acetyl-4-oxopentanoate intermediate. (LMW 7-1)

irh4712-77 (1) ich 4717 70 (1)

jrh4712-76a (1

School of Chemistry All Blogs | Help | Support | About

ourexperiment

Search

Archives

April 2013 (2)

July 2012 (20)

June 2012 (4)

May 2012 (21)

April 2012 (12)

October 2012 (2)

August 2012 (22)

September 2012 (6)

Older Posts >>

MNR26

### ourexperiment

All Blogs | Help | Support | Abou 🚨 Login

### Synthesis of 4-substituted methylidene oxindoles

Project E-Lab Notebook for the synthesis of five 4-substituted methylidene oxindole from oxindole and their corresponding aromatic aldehydes.

Older Posts >> Search

### Single Crystal X-ray crystallography

27th June 2012 @ 23:09

🚨 Login

Single-crystal X-ray diffraction analyses were performed using a Bruker APEXII CCD diffractometer mounted at the window of a Bruker FRS91 rotating anode (MoKa = 0.71073 Å) and equipped with an Oxford Cryosystems cryostream device. Data were processed using the Collect package and unit cell parameters were refined against all data. An empirical absorption correction was carried out using SADABS . The structures were solved by direct methods using SHELXS-97 and refined on Fo<sup>2</sup> by full-matrix least-squares refinements using SHELXL-97. All non-hydrogen atoms were refined with anisotropic displacement parameters. All hydrogen atoms were added at calculated positions and refined using a riding model with isotropic displacement parameters based on the equivalent isotropic displacement parameter (Ueq) of the parent atom. Figures were produced using OLEX2

Graham Tizzard | View Source | Analytical Procedures | Comments (0)

### MS Spectrum of (3E)-3-(4-Nitrobenzylidene)-1,3-dihydro-2H-indol-2-one

6th May 2012 @ 18:16

### Spectroscopic Method: MS-ESI

Substituent: Nitro MS Spectrum of (3E)-3-(4-Nitrobenzylidene)-1,3-dihydro-2H-indol-2-one:



Tools Show/Hide Keys

The mass spectrum of (3E)-3-(4-Nitrobenzylidene)-1.3-dihydro-2H-indol-2-one has been obtained by negative electrospray ionization (ESI). The peak at m/e = 265.2 confirms the molecular mass of this compound as the molecular ion gains a proton.

Interpretation of MS Spectrum of (3E)-3-(4-Nitrobenzylidene)-1,3-dihydro-2H-indol-2-one:

Peak	Diff. between molecular mass	Suspected molecules or	Inference		
Position	and peak	ions			
265.2	265 - 264 = 1	(M+H)+	Parent compound plus a proton		

### Archives

June 2012 (1) May 2012 (15) March 2012 (29)

Sections

Analytical Procedures (B)

**Condensation Products** 125 Experimental Procedure

(1) Spectroscopic Data (31)

### Substituent

Nitro (8) Methoxy (8) Bromo (8) Chlore (B) Methyl (B)

### Spectroscopic

Method DSC (5) ATIR-FT-IR (5) HPLC (5) MS-ESI (5) PXRD (1) C-NMR (5) H-NMR (5)

### Procedure:

MNR11-8 (3.96 g, 10.63 mmol) was dissolved in EtOH (21 mL) and HCI (1M) (110 mL) and heated to reflux for 2.5 hours. The solution was allowed to cool to room temperature then cooled in an ice bath, basified with NaOH pellets (approx B g, pellets used to minimise volume of aqueous material) to pH 12-13 and extracted with DCM (4 x 100 mL). The organic fractions were combined, dried over magnesium sulphate, filtered and concentrated under reduced pressure to give an orange crispy solid (1.008 g).

### TLC

Reaction mixture after 2.5 hours ran in 100% EtOAc



Hydrolysis of MNR11-18 23rd April 2013 @ 01:09	to give MNR2	6-7
Mnr: 21-30 As for Hydrolysis of MNR11-16 to MNR26-4 Hydrolysis of MNR11-17 to MNR26-5 Starting material from Synthesis of MNR11-1B		
	HCI, EtOH	

Pictet-Spengler route to Praziguantel

0	Compound	FW (gimoti	Density (g/mL)	mmol	Atass (g)	Volume (mL)	PQ.	(mmolineL)
	AIN(211-18)	372.46		10.63	3.56	1.0.0.0.0	1.00	1
	TM HC3	1.	2.97	106.32		110	10.00	
	EIOH .				2	21		8.50
	MMR26-7	262.30						-

MNR11





### Mnr

11-20 (4) 41-50 (15) 1-10 (4) 21-30 (2) Sc

1-10 (5) Tools

Show/Hide Keys

Ists

line

xed

1.1

n at

20

504

pate

riv.)

and

nL).

der

lica

ited

not

(sx,

53.

MW

12



• ELN as Supplementary Information for conventional publication (*Chemistry Central Journal* 2013, 7:182 )





• ELN as Supplementary Information for conventional publication (*Chemistry Central Journal* 2013, 7:182 )



21



• ELN as Supplementary Information for conventional publication (*Chemistry Central Journal* 2013, 7:182 )





• ELN as Supplementary Information for conventional publication (*Chemistry Central Journal* 2013, 7:182 )

		Crystals			Southampton			
7	COM	Home About Browse by Year	Browse by People					
-		Login   Create Account			Secon		Home	About
			Bromo Oxin	dole (2010src0792)				
φφ	<b>39/010</b> sis	Sample Originator: Romanus O Or Spencer <sup>4</sup> .	vyeabo <sup>a</sup> , Mark Edwards <sup>a</sup> and John	1		MS.	SXRD	PXRD H
		Data Collection: Graham J Tizzard	and Simon J Coles <sup>6</sup>	-	r i			
		Structure Determination: Graham	J Tizzard <sup>b</sup>	L			17 CF	
3a	0-140 <sup>20</sup>	University of Generalist <sup>®</sup> University of Southempton <sup>®</sup>		K	2000 12	1.00	the second	
		(C <sub>19</sub> H <sub>10</sub> BrNO) (H <sub>2</sub> O)		21				
		InChi=1/C15H10BrNO.H20.H2/c15 14(12)17-15(13)18_h1-9H.(H.17.18	11-7-5-10(6-8-11)9-13-12-3-1-2-4- (;1H2;1H/b13-9+;;	1	19-000		and the	
3b	್ಷ-ನಿಕ್ಷಾನನ್ನಿ	Controlled Organic Keywords:		Augusta Files	Jmal	-	ATH.	)
		Date 21 September 2010 Created:		Final Result		Arres	· ·	
		Deposited 19 Mar 2012 14:24 On:		2010sro0792ra.off	13k			
		Deposited Or G J Tizzant By: Data collection parametera		2010src0792ra.cml	64		×	TU BAL
3c	and early			2010src0792ra.fcf	142k	34	18 H	E Land
_				Validation		100	A CT	- that all
		Chemical formula	C15 H12 Br N O2	2010src0792ra_checkcif.htm		and the second	1 A W	B entropy
		Crystal morphology	Block	Deferment			1 ×	
		Crystal system	moroclinic	President in the second s	-		4. 25	
3d	0-0005	Space group symbol	C 2%	20108/02/92/92/92	58.		1274	
-		Cell length a	19.623(3)	Solution		and the second	201	
		Cell length b	4.0710(5)	2019src0792r.prp	71	A Company of		
		Cell length c	32.979(4)	Processing		-	5.0	
	1. Mar. 1. Mar	Cell angle alpha	90.00	2010src0792r.htm	118	_	V. FC	
3e	00-deol	Cell arigle beta	101.699(3)	2010src0792r_0klipg	34k		TTH	
		Cell angle gamme	90.00	2010src0792r_h0Lipg	29k		AL.	
		Data collection temperature	120(2)	2010src0792r_hk0.jpg	42k	Constructions.		
		Refinement results		2010src0792ra.hkd	2948			



# Metadata support for solutions.

Invariably the context for a crystal structure

Or

The context a crystal structure has to fit into!

### Datastores, ICAT & CSMD



The Core Scientific Metadata model forms the information model for ICAT & is designed to describe facilities-based experiments



### CSMD as a Starting Point

- Doesn't cover all, but...
- Forms the basis for extensions:
  - To derived data
  - To laboratory based science
  - To secondary analysis data
  - To preservation information
  - To publication data





### Recording process

• Plan (Prospective provenance)



• Enactment (Retrospective provenance)



Realisation





### oreChem Plan for eCrystals

- Machine-readable representation of methodology
- Describes requirements for software and data products

Cry	stals				Southampto
Home Abo	ut Browse by Year Brows	e by People			
Besit.					60070
	Top are viewing at HTML	version of the dep	artent. To and Per-	ndetyng XML vax your ees intenens "Vi	ev Page Seurce" option
3	4-Diphenyl-1H-pyr	role-2.5-	ficarboxy	lic acid bis-I/3.5-din	itro-phenyl)-amidel
		and alle i		ne ann an Kele an	
łók	entitien: <u>http://eczystals.chem.s</u>	oton at uk/21	<u>M</u>	(RMIIH, Rput) (RMIH	Advantation (See See ) (IIIIII III III)
6	Treator: Camiolo, S.				1 _ 1
G	Steator: Gale, Phil A.			(1311104A4) (1031	TOALE CortEnas
c	treator: Light, Mark E.				
C	Steator: Hursthouse, Michael B.			(flastin LST) (flas	eth ATA
	Date: 2002-04-12				
Proven III				(Elser (B.FRF)	>/
stages (a)	N				
Stage:	Plan-stage URL:	Num, Piles Used:	Num. Files Emilted:	(Ibal@381	2
1	plan.rdl#SADABS	0	2	Files (11)	
3	plan.ntMXPREP	1	. 1	Filename:	Plan-object URL:
3	plan.nt#SHELX	2	2	02sot038.HKL	atan.miWHKL
4	plan.rdf#FinalSolution	2	2	0250T039.htm	plan.oddHTML
5	plan.nt/PCheckCIE	1	2	02xot038.PRP	plan.off#PRP
	plan.rdttOpenBabe/ToMOL	1	1	02sot039.LST	plan.rdf#LST
7	plan.rdf9OpenBabelToCML	1	1	02xot039.855	plan.com RES
8	plan.ntf#DepositRecord	11	0	02SOT039.CIF	plan.mtMC1F
				02SOT039.doc	
				destruction of the state of the	
				62801039 checkof.htm	plan.rdMCHECKCIF_HTML
				02SOT039_ellipsoid.gif	plan.rdWCHECKCIF_HTML plan.rdWCHECKCIF_GIF
				02507039_ellipsoid.ptf 02507039_ellipsoid.ptf 02501039.moj	plan.ndWCHECKCIF_HTML plan.ndWCHECKCIF_GIF plan.ndWMOL



### SIMS:



### Sample Information Management System

- A standard/format for crystallographic sample and experiment data management and archival
- Supported by CrystalClear and NCS Portal, providing interaction between facility, instruments and CIF, ImgCIF etc

#Twel version="L.8"7>
#FileType Type = "Sumple">
#User Nove + "Buss\_Sthey"/> destructions being a 199300-2 dispirical Teas Teas - "0.66"/> ScreenProgram Program + "Crystalfloor].104"/> "Environment Tempi - "128.00" Pressure - "29.89" Humidity - "62.38"/\* Twince clwin Washer = "1" Twining = "5lide 1/2"/>
clwin Washer = "2" Twining = "Twist 1/3"/> «ProjectNose Nose - "91" /s. coldStrageTamp Tempt - "-148.48"/> «Samplehame Hume = "EPSIRS11" /» «Instrument Inchrumant - "AFC12 Soburn 724+"/» cleageDirectory Directory = "CitizegesManaCyt"/s
cleageTemplate Template = "IPSIMSIL"/> "Bovelength Wovelength - "8. 2272020"/s Twitter "Element Element - "Moly"/s «/Process» Gelvo «ProjectInfa» «ScreenScene» -SolutionProgram Program = "Dise2/Disis"/> -DotablaseID ID - "AN3145"/\* •Scan Scankador = '1' Axis = 'Owege's •Scankange thart = '-60.8' Ind = '-50.8' With = '1.8' Ing = '1.8' cher10 10 = "A32456"/> -SolutionMethod Hethod - "Direct"/+ Restruints+ cfacility10 10 + "CAMENES"/> -GonioAxes ChiO-Kappa = "54.8" PhiOrOmega = "8.8"/s "Restraint Parameter - "Parami" Value - "8.0000" Tolerande - "8.1008"/> efficity Priority - "Medius"/s -detectorisitings Distance - "48.8" September - "10.0" TeeTheta - "21 "Restroint Parameter = "Param2" Value = "1.9900" Televence = "0.2000"/> -SubwissionForm Farm - "Standard OnLine"/-"DetectorWode Blanking = "I" Deringer = "Yes"/s «/Restraints» -BulkProperties Properties - Weltiple Crystals in «/Scim Constraintso caulkMorphology Morphology - "Westly rods"/s «/ScreenScans» «Constraint Parameter - "Parami" Value - "8.0000" Talerance - "8.1000"/> «SolutionEnal Goal + "Fublishable Structure"/# diank Buth = "135"/> «Construint Parameter = "Paramet" Value = "-1.0000" Tolerance = "0.0010"/+ «UserFormula Formula = "CBOSN3W13"/» «IndesProgram Program = "CrystolClear3,104"/» -/Constroints--LubelingScheme Scheme + "Standard"/s «IndexAlgorithe Algorithe - "#"TREE"/+ diructure lyne - "Solve"s clementiousty Sensitivity - "Light Sensitive"/> «CSOMotiches» dti Percentase - 74.307/5 "Sofety Sofety - "Countie weatlow"/" -CiDMatch Hatch - "Someone in Combridge did this years ago"/> cluteProjectSubmission Date = "2012.01.04"/+ -Atoms--CSONatch Match - "St Andress did this in 1994"/> 4100 cDotsProjectReport Sate = "2812.91.96"/+ «/ISIMutches» cationCoord latel="01" (ywini="0" =="0.000000" y="0.100000" y="0.20000" Octubercy="1.00000"/> Prujectlafm--State Type - "Screen"> "Atominian UT1-"8.000000" UZ2-"8.010000" UZ3-"8.020000" UZ3-"8.000000" UZ3-"8.000000" UZ2-"8.00000"/-«SampleInfo» cfil A = "5 123" B = "13.067" C = '14.704" Alona = "80.000" Bata = " <CellE50 A = "0.000000" B = "0.000000" C = '0.000000" Alona = "0.00000</pre> -/ALC#-«DrystalNowber Humber - "4"/s aktows <frystallographer Name - "Graham"/> «CrystalSystem System - "Orthorhombic"/~ "extendioted label="M1" lymbul="M" a="8.000000" y="8.200000" a="8.400000" Occupancy="1.000000"/> «PreparationScheme Scheme - "Chig of Large crystal -SpaceGroupNore Nume - "P 2 2 21/> shipederies UT1-"8.000000" UTI-"8.02000" UTI-"8.640000" UTI-"8.00000" UTI-"8.00000" UTI-"8.00000" /-«FragmentEut Cut = "Yes"/». -tmeClass Class - 'm m m'/s «InvatalColor tolar - "Colorings"/> - 743 cm-«ErystolSize X = "0.20" Y = "0.20" 2 = "0.20"/> distitut Latting - "P"/> «/Atoms» ctrystalWarphology Humphology = "Priss"/s
«Solvent Solvent = "PES200"/s «Centricity Cambricity = "acentric"/> «/Structures «SpaceGroupNumber Humber = "16"/s n/Solver History -Reerge Emerge - "4.32"/s -ManipulationMedium Hadium = "Stuff"/s -Scantfype Type = "Fiber"/s -ScepteMountID ID = "MC52765"/s -Completeness Completeness = "98.70"/> -Step Bets = "12.81.11" Type = "Streen"s-/Step-Redundancy Redundancy = "4.30"/+ citep Data - '12.01.11' Type - 'Neta's The sample was abandoned here becaue the reflections were very weak and -fotalReflections Heflections - "34987"/> "WiscolQuality Quality - "Unknown"/w it could not be indexed-/StepopticalProperties Properties - "Transporent"/--UniqueReflections Reflections - "27654"/e/Historys discally Twinned Tataned - "No"/> -dejectedieflections Reflections + "2456"/> dissults. «ValidutionProgram Program = "CharddlP"/w «DoteDatabaseExtraction Date - "2012 01.06"/v «MaxResolution Hesolution - "\$.83"/> -VallantionErrors--DoteMounted Bats - "2012.01.07"/> s/States eValidationError Error - "Emerge too high"/s -DataScreened Data = "2012.01.077/s «/Screent» «ValidationIrror Frenz » "Resolution too low"/» cOntsCollected Date = "2012 81 88"/> «Collect» "EnteProcessed Date - "2012.01.09"/\* -WolldationErrars-«StrotegyProgram Program - "Crystallleur 3.104"/» «/Results» «EnteSolved Data = "2812.01.09"/dollectSchedule Schedule + "Default\_Collect"/> -dentilAddress fruil - "Anne user@incoltest"/> <DeteDetabaseDeposit Nate - "2012.01.10"/> «PredictedCompleteness Completeness = "99:80"/» "Status Status - "Ready"/> -/Sampisinfas ProdictedRedundancy Endurdancy = "4,48"/> «LestError Error » "Note"/\* Screen -CollectProgram Program - "CrystalClear 3.164"/diarik flame - "135"/s «ScreenProgram Program » "CrystolClear3.1mt"/s dinvironment TempE + "128:00" Pressure - "20.80" flatidity - "62.30"/s



### A semantic framework for chemistry

• Describes and relates different types of process information



### elnItemManifest



31

- Layered metadata model for description, export & packaging
- This is the first (information) layer leads into knowledge
- Published through Dial-a-Molecule at<u>http://wp.me/p2JoQ6-xF</u> & in J. ChemInf 2013, 5:52

```
<?xml versions"1.8"/>
                                                                                                                          cxs:complexType nume="contentInformation">
wstschend wnlmstxs="http://www.w3.org/2001/XMLSchend">
                                                                                                                            <xs:annotation>
                                                                                                                              -xs:documentation-Text that describes what the item is, with a descriptor of the «/xs:documentation»
<xs: annotation>
                                                                                                                              -xs:documentotion=digital type, or "undefined" if no corresponding MIME format=/xs:documentation=
  <ks:documentation>Change history/xs:documentation>
                                                                                                                            </xs:onnotation>
  <xs:documentation>====/xs:documentation>
                                                                                                                            -xs:sequences
  cxs:documentations15 June 2012 - [CL8] Created from enDataDescription prototype</ss:documentations
                                                                                                                             <xs:element nome="description" type="ks:string"/>
«/xs:annotation»
                                                                                                                             <xs;element nome="mimeType" type="xs:string" default="undefined"/>
                                                                                                                            «/astsequence»
cxs:cnnotation>
                                                                                                                          «/xs:complexType»
  costdocumentation-Definitions of data types used in the manifest of the ELN item
 cxs:documentation>
                                                                                 ---/xs:docomentation>
                                                                                                                          cxs:complexType nome="related10">
</xs:annotation>
                                                                                                                            <xs:prinotations</pre>
                                                                                                                              -xs:documentotion-Nature of the related information, for example, publication or related work
<xs:complexType name="keywordSet">
                                                                                                                              -xs:documentation-Id can be any string, but DOI preferred if the related item is a publication-/ks:documentation-id can be any string.
  <xs:unnotation</pre>
                                                                                                                            «/xs:onnotation»
     <xs:documentation>A list of terms</xs:documentation>
                                                                                                                            <X51SEQUELCEN
  </xs:ennotetion>
                                                                                                                             <xs:element nome="relationship" type="%S:string"/>
  cits:sequence>
                                                                                                                             cas:element nome="id" type="sa:string"/>
   <s:element name="keyword" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
                                                                                                                            «/xs:sequence»
  «/xs:sequence»
                                                                                                                          «/xs:complexType»
</xs:complexTypes
                                                                                                                          cxs:complexType name="relatedltemSet"s
«xs:complexType none="identifierSet">
                                                                                                                           -xs:amotations
  ext: sequences
                                                                                                                              <xs:documentation>Zero or more item(s) of related information</xs:documentation>
    <xs:element name="primuryLocalIdentifier" type="ks:string">
                                                                                                                            </r>
     exs: annotations
                                                                                                                           -cost sequences
        <xs:documentation>Primary string, URI, or item in any other format that enables
                                                                                                                             os:element none-"item" type="relatedID" minOccurs="0" maxOccurs="unbounded"/>
        <xs:documentation>this record to be located uniquely in the originating system
                                                                                                                            «/xs1sequence»
      «/astennotation»
                                                                                                                          </www.complexTypes
    e/xs:elements
    -cxs:element nome="atheriocalIdentifier" type="xs:utring" minOccurs="8" maxOccurs="unbounded">
                                                                                                                          <xs:complexType nume="contributorInformation">
     <xs:enotation-</pre>
                                                                                                                            xxs:phonototiom
       <xs:documentation>[Optional] Alternative means of locating record in the originating system
                                                                                                                              «xsidocumentation»For example, Author, Funding Body, PI, institution, ....</xsidocumentation»</pre>
                                                                                                                              <xs:documentation>Flain text, but name ideally complemented by unique identifiers
      </siconnotation>
    «/ks:element>
                                                                                                                            </xs:annotation>
    <xs:element none="accessIdentifier" type="ksionyURI" #inOccurs="0">
                                                                                                                            cx5:Sequence>
                                                                                                                             kxs:element nume="role" type="xs:string"/>
      <xs:annotation>
        <xs:documentation>[Optional] URI that provides a direct link to the content.</ks:documentation>
                                                                                                                             <xs:element name="name" type="xs:string"/>
        <xs:documentation>If included, must be a 'linked data' URI giving open access</xs:documentation>
                                                                                                                            «/xs:sequence»
                                                                                                                          e/xs:complexType>
      </signation>
    «/xs:element»
                                                                                                                          cms:complexType name="contributorSet">
  «/xs:sequence»
                                                                                                                           xs:annotation>
</xs:complexTypes
```

### Standards for reactions: S88



- Group arising from Dial-a-Molecule consisting of Mettler Toledo, Pfizer, GlaxoSmithKline, AstraZeneca, Johnson & Johnson, Southampton University, NextMove, Royal Society of Chemistry looking to:
  - Provide guidance for S88 implementations for synthetic organic chemistry reaction procedures
  - Provide example set
  - Agree on controlled vocabularies for elements
  - Generate a schema
  - IUPAC uptake?

### Standards for reactions: S88



```
<?xml version="1.0" encoding="utf-8"?>
 Application="RscDataRepository" ApplicationVersion="1.0.0.0" Experimentid="ReactionRunIDFromDataRepository" ExperimentStatus="?????" FormatVersion="1.0">
    <LinkedExperiment>
         <Experiment Name="SIH-01-211" Description="Another run of the same reaction"/>
         <Experiment Name="SJH-01-211" Description="The previous step in the reaction which created &lt;b&gt;16&lt;&#47;b&gt; as a product"/>
     <LinkedExperiments/>
    «Process»
         <ProcessStage>
         <ProcessOperation Name="Reaction">
              <ProcessAction Name="Start Experiment" Description="Add &it;b&gt;16&lt;&#47;b&gt; (0.101 g, 0.132 mmol) to a round bottom flask.">
                  <ProcessParameter Name="&lt;b&gt;16&lt;&#47;b&gt;TotalMass" Value="0.101" Unit="g"/>
                  <EquipmentUnit Unitid="I"/>
              </ProcessAction>
              <ProcessAction Name="Add Cu(OTf)&it;sub&gt;2&it;&#47;sub&gt;" Description="Add Cu(OTf)&it;sub&gt;2&it;&#47;sub&gt; (0.006 g, 0.01 mmol) under a N&it;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&i
atmosphere" >
                  <ProcessParameter Name="Cu(OTT)&it;sub&gt;2&it;&#47;sub&gt; TotalMass" Value="0.005" Unit="g"/>
                  <ProcessParameter Name="Atmosphere" Value="N&lt;sub&gt;2&lt;&#47.sub&gt;"/>
                   <EquipmentUnit Unitid="1"/>
                 «/ProcessAction»
                 <ProcessAction Name="Dissolve & It.b&gt:2&it;&#47;b&gt; in C&it;sub&gt;2&it:&#47;sub&gt;4&it;sub&gt;4&it;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub&gt;2&it;&#47;sub
vial, dissolve <b&gt;2&lt;&#47;b&gt; {0.155 g, 0.753 mmol} in C&it;sub&gt;2&lt;&#47;sub&gt;H&it;sub&gt;4&it;&#47;sub&gt;C1&lt;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&#47;sub&gt;2&lt;&=11;sub&gt;2&lt;&=11;sub&gt;2&lt;&=11;sub&gt;2&lt;&=11;sub&gt;2&lt;&=11;sub&gt;2&lt;&=11;sub&gt;2&lt;&=11;sub&gt;2&lt;&=11;sub&gt;2&lt;&=11;sub&gt;2&lt;&
                       <ProcessParameter Name="&lt.b&gt;2&it;&#47;b&gt;.TotalMass" Value="0.155" Unit="g"/>
                        <ProcessParameter Name="&lt;b&gt;2&lt;&#47;b&gt;.TotalVolume" Value="1.3" Unit+"mL"/>
                        <EquipmentUnit UnitId="2"/>
                   «/ProcessAction»
                  __ etc __
         </ProcessOperation>
      «/ProcessStage>
    «/Process»
                                            16 (0.101 g, 0.132 mmol) and Cu(OTf)_2 (0.006 g, 0.01 mmol) were added to a round-bottom flask under a N<sub>2</sub> atmosphere.
    <Unit UnitId
    «/Unit»
                                         In a separate vial, 2 (0.155 g, 0.753 mmol) was dissolved in C_2H_4CI_2 (1.3 mL) and transferred to the reaction flask.
   <Unit UnitId
                                         CF<sub>3</sub>CO<sub>2</sub>H (0.030 mL, 3 equiv) was added to the reaction mixture, which was refluxed at 100 ° C for 1 h. The reaction
   </Unit>
                                        mixture was washed with saturated NaHCO<sub>3</sub> (15 mL) and extracted with C<sub>2</sub>H<sub>4</sub>Cl<sub>2</sub> (3 x 5 mL). The organic fractions were
 </Esperimen
                                          collected, dried (MgSO<sub>4</sub>), and filtered to give a dark red solution. The solvent was removed, and the product was purified by
```

column chromatography (SiO<sub>2</sub>, 30:70 CH<sub>2</sub>Cl<sub>2</sub> : hexane) to yield 17 as a pale yellow powder (0.096 g, 68% yield).

### **Allotrope Foundation**



• Standards for analytical process in (pharma) industry



Drganization Member Companies Framework Architect Partner Network Members Focused on Delivery FAQ Contact Us

Membership in Allotrope Foundation is open subsidiary or division thereof, which di pharmaceutical or biological products. Fo Foundation, please use <u>this</u> link.

Current Allotrope Foundation members include

- AbbVie
- Amgen
- Baxter
- Bayer
- Biogen
- Boehringer Ingelheim
- Bristol-Myers Squibb
- Eli Lilly
- Genentech/Roche
- GlaxoSmithKline
- Merck & Co.
- Pfizer



### Allotrope Foundation

Southampton

School of Chemistry

The tasks and deliverables to achieve the goals of Allotrope Foundation are carried out by Working Groups who work in close collaboration with our <u>Framework Architect</u>. These groups include:

- Core team Coordinates activities within the working groups, ensuring the forward progress of the project.
- Architecture & Development Responsible for defining the architecture, underlying data model, and specifications for the software components of the Allotrope project built by our <u>Framework Architect</u>.
- Business Model Development Develops business strategies and business models for the project both near- and long-term.
- Communications Develops and implements communication strategies to ensure consistent, timely and efficient messaging to all relevant audiences, both internal and external.
- **Membership** Reaches out to potential new members to support expansion of the project base through increasing Allotrope membership
- **Proof-of-Concept (PoC) Delivery** Gathers requirements for integrating the Allotrope Framework into a member company's daily business practices.
- Quality and Compliance Ensures that quality, validation, and other regulatory-relevant requirements and best practices are "built-in" at all levels during the development of the Allotrope Framework.
- Regulatory Affairs Interacts with regulatory agencies regarding activities of Allotrope Foundation in order to educate, to provide a means for the Allotrope Foundation to gain a better appreciation of the regulatory perspective and to provide a channel for regulators to offer feedback and influence the development of the Framework
- Standards (Document Standards, Metadata, and Test Data) This group, along with subject matter experts from our <u>Framework Architect</u> evaluates existing data standards, and defines appropriate controlled vocabularies and ontologies to use with the Framework. Provides real and simulated test datasets for use in Framework development, and in close collaboration with our <u>Framework Architect</u> maps the metadata associated with the test data to existing standard definitions.

### **Allotrope Foundation**



		Ноте	FAQ	Contact Us
oundation	Q Search site ** This site is <u>NOT</u> compatible with IE8 **			
Our Shared Vision	What are we building?	Information Center	Jc	oin Allotrope
	Allotrope oundation ilding an Open Framework for Laboratory Data Our Shared Vision	Allotrope oundation Iding an Open Framework for Laboratory Data Our Shared Vision What are we building?	Allotoppe oundation Inding an Open Framework for Laboratory Data Our Shared Vision What are we building? Information Center	Allotrope       FAQ         Q. Search site       ** This site is NOT compate         Our Shared Vision       What are we building?       Information Center       Journal

### What are we building? Metadata Repositories

Open Document Standards Metadata Repositories Class Libraries Proofs-of-Concept What is out of scope?

CONTRACTOR OF

The Metadata Repository provides a single location for maintaining accurate, pre-approved metadata that applications will use as input to drive scientific and business processes. The controlled vocabulary in the Metadata Repository allows for the standardization of input to software applications, thus ensuring that complete, correct, and consistent metadata enters the system from the very start. This eliminates the need to clean up, reconcile, or rename metadata values after data creation and enables many more possibilities for downstream automation in the analytical laboratory than exist today.

The controlled vocabularies used by the Framework will be defined using <u>Document</u> <u>Standards</u>. The Allotrope public domain vocabulary facilitates interoperable exchange of laboratory data; Vendor vocabularies will allow for application-, technique-, or platform- specific data that are not strictly vendor neutral, and Enterprise vocabularies will provide enterprise specific data.

### **CREAM:**

Collaboration for Research Enhancement using Active Metadata

- How to collect and use metadata actively to capture tacit information
- Active metadata: assemblage of metadata and annotations used actively within the process that generates it (capable of being reused by another process).
- Central Facilities; Chemistry; Geosciences; Art; Music...
- Uptake: CODATA; Research Data Alliance







https://blog.soton.ac.uk/cream/





**Research Data Sharing** 

### **Research Data Alliance**

- Chemistry data interest group
- without barriers
- Joint RDA/IUPAC Charter drafted •
  - Characterise chemical data types
  - Leverage to establish standards
  - Examine workflows in disciplines interacting with chemistry
  - Cultivate a sharing culture
- Inaugral meeting: Paris Sept 23<sup>rd</sup> September 2015 (hosted by Ian Bruno, Richard Kidd and Dave Martinsen)