



CCDC Metadata Initiatives

Suzanna Ward, Ian Bruno, Colin Groom

Workshop on Metadata for raw data from X-ray diffraction and other structural techniques

Session IV: Data in the Wider World - From Laboratory to Database

Sunday 23rd August 9:35-10:00

From data to knowledge

CSD 50 1965 - 2015

Experimental Data



Structural Knowledge

From raw data to model



Raw	
Image data	Positions and intensities of diffracted beams
Processed	
Structure factor data	Quantities needed to derive electron density
Derived	
Atom coordinates	Modelled and refined from electron density

Primary emphasis has typically been on deposition of **derived** data only









The Cambridge Structural Database





Metadata



A set of data that describes and gives information about other data

- Data that describes the substance studied
 - Important for discovery, data analysis and mining
- Data that describes the dataset as a whole
 - important for provenance (structure of record), attribution (citation)
 - Data that describes the associated publication
- Data that describes the experiment
 - where was it done, how it was done, who did it, who funded it

7

Recent metadata developments at CCDC

Web Deposit

- Updated Web Interface with greater interactivity
- Aim to capture as much input from depositor as possible

CSD-Xpedite

- Manages depositions and creation of CSD entries
- Designed to adapt to changing internal and external needs

CCDC DOIs

Implemented 2014

Launched 2014

Launched 2014

Deployed 2013

- Unambiguous and persistent identification of datasets
- Foundation for formalising data citation and interoperability

Get Structures

- Update to previous "structure request" service
- Enhanced web interface for accessing individual structures

audit creation method SHELXL-97 _chemical_name_systematic 33 {5-[(7-chloroquinolinium-4-yl)amino]-2-hydroxybenzyl}diethylammoni dihydrate _chemical_name_commor 'Amodiaquine dihydrochloride dihydrate 'C20 H24 C1 N3 O 2+, 2(C1 -), 2(H2 O) chemical formula moietv 'C20 H28 C13 N3 O3' chemical formula sum hemical formula iupac 'C20 H24 C1 N3 O 2+, 2C1 -, 2H2 O hemical formula weigh: 464.80 chemical melting point monoclini













Crystallographic Information Framework (CIF)

Acta Cryst. (1991). A47, 655-685

International Union of Crystallography

Commission on Crystallographic Data Commission on Journals Working Party on Crystallographic Information

The Crystallographic Information File (CIF): a New Standard Archive File for Crystallography*

BY SYDNEY R. HALL

Crystallography Centre, University of Western Australia, Nedlands 6009, Australia

FRANK H. ALLEN

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, Engloop

• A standard format for crystallographic data

- derived results
- raw and processed data
- experimental conditions
- publication data

AND I DAY	VID BROWN	_atom_site_label
	BROWN	_atom_site_type_symbol
Institute for Materials Research, McMaster Un	iversity, Hamilton, Ontario L8S 4M1, Canada	_atom_site_fract_x
		_atom_site_fract_y
(Received 8 April 1991;	accepted 28 June 1991)	_atom_site_fract_z
		_atom_site_U_iso_or_equiv
Abstract	Introduction	_atom_site_adp_type
1105tract	THE OULCED	_atom_site_occupancy
	There is an increasing need in many branches of	_atom_site_symmetry_multiplicity
The specification of a new standard Crystallographic	for a uniform but flexible method of archiving	atom site_calc_flag
Information File (CIF) is described. Its development is	changing data in electronic form. Rapid advances	atom site refinement flags
based on the Self-Defining Text Archive and Retrieval	puter technology, coupled with the expansion	atom site disorder assembly
(STAR) procedure [Hall (1991). J. Chem. Inf. Comput.	national and international networks, have fuelled	
Sci. 31, 326-333]. The CIF is a general, flexible and	for such a facility. The variety and relative infl	_atom_site_disorder_group
easily extensible free-format archive file; it is human	of existing data exchange formats have inhibited	C11 C1 0.5993(2) 1.0007(7) 0.8131(17) 0.044(3) Uani 0.50 1 d PDU A 1
and machine readable and can be edited by a simple	fective use. This is true even in fields where t	S1 S 0.5321(3) 0.8260(6) 0.9322(3) 0.0327(11) Uani 0.50 1 d PDU A 1
text editor. The CIF is designed for the electronic	data requirements are well defined. Problems of	C2 C 0.5529(4) 0.8802(9) 0.8184(9) 0.029(4) Uani 0.50 1 d PDU A 1
transmission of crystallographic data between individ-	change are further exacerbated if the number and	C3 C 0.5286(7) 0.8174(18) 0.7440(7) 0.031(4) Uani 0.50 1 d PDU A 1
ual laboratories, journals and databases: it has been	data types change rapidly and continuously. Unc	H3A H 0.5350 0.8343 0.6771 0.037 Uiso 0.50 1 calc PR A 1
adopted by the International Union of Crystallogra-	conditions specialized and local file formats hav	C4 C 0 4918(8) 0 7220(19) 0 7782(8) 0 027(4) Harri 0 50 1 d DDH & 1
phy as the recommended medium for this purpose.	erated. This diversity was tolerable when electro	
		C5 C 0.4900(6) 0.7171(14) 0.8779(5) 0.029(4) 0811 0.50 1 d PD0 A 1
Acta Cryst 1991 A47 655 doi:10.11	07/\$010876739101067X	C12 C1 0.3202(2) 0.4982(6) 1.0830(5) 0.0586(15) Uani 0.50 1 d PDU A 1
(icia crysti) 1991 , (47, 055 doi:10.11	0,,00100,0,0010100//	S2 S 0.38755(19) 0.6658(5) 0.9578(5) 0.0400(10) Uani 0.50 1 d PDU A 1

The growth of electronic deposition



Hand-typed tables of coordinates

					J. CHE	M. SOC. DALTON	trans. 1985	
I. Crystallograp	hic data and detai	ls of data collectio	n and processing	for ML(NO3)2, with	1 M = Cu [i	n (1)], Ni [in (2)], ar	d Cd [in (3)]	
		(1)		(2)		(3)		
Stoicheiometry M		C ₁₈ H ₂₇ Cul 529.01	N ₉ O ₆	C ₁₈ H ₂₇ N ₉ NiO ₆ 524.18		C ₁₈ H ₂₇ CdN ₉ O ₆ 577.88		
Lattice type Space group a/Å	J. CHEM. SOC.	DALTON TRANS	1985					
b/Å c/Å	Table 2. Atomic	co-ordinates with	estimated standar	d deviations in parent	heses			
α/" β/"	Atom	X/a	Y/b	Z/c	Atom	X/a	Y/b	Z/c
Y/~	(a) Compo	und (1) (×10 ⁵ for	Cu, × 10 ⁴ for oth	ers)				
U/A-	Cu	7 734(2)	11 519(5)	62 931(2)	C(27)	2 251(3)	671(8)	4 844(4
Z D (n mm-3	N(01)	1 493(1)	-122(3)	6 454(2)	C(30)	2 044(2)	483(5)	6 905(2
$D_{\rm e}/{\rm g} {\rm cm}^2$	C(10)	1 359(2)	-1 289(5)	6 758(3)	N(31)	1 888(1)	1 149(3)	7 431(3
u(Mork)/cm ⁻¹	1 N(11)	708(1)	-1 425(3)	6 566(2)	N(32)	1 386(1)	1 906(3)	7 246(2
Approximate of	N(12)	379(1)	- 356(3)	6 561(2)	C(33)	1 395(2)	2 512(4)	7 809(2
dimensions	C(13)	-162(2)	- 764(4)	6 558(2)	C(34)	1 907(2)	2 157(5)	8 332(2
Number of set	C(14)	-171(2)	-2078(4)	6 570(2)	C(35)	2 220(2)	1 302(4)	8 094(2
θ range/° (cell		386(2)	-24/5(4)	0 580(2)	C(36)	918(3)	3 421(6)	/ 819(
(data	C(10)	-651(2)	142(7)	6 606(4)	N(40)	2 766(2)	2 265(5)	5 5000
h range	C(20)	1 549(2)	-339(5)	5 777(2)	0(41)	136(1)	2 399(3)	6 1630
k range	N(21)	1 509(2)	902(3)	5 458(2)	0(42)	-717(2)	3.031(5)	5 4860
/ range	N(22)	1 099(1)	1 738(3)	5 555(2)	O(43)	-272(2)	1 385(5)	5 2170
Number of ref	1 C(23)	1 123(2)	2 745(4)	5 177(2)	N(50)	1 960(2)	3 011(5)	3 3460
measured	C(24)	1 539(3)	2 539(6)	4 854(2)	O(51)	1 968(2)	4 072(5)	3 604(
independent	C(25)	1 786(2)	1 358(5)	5 035(2)	O(52)	1 526(2)	2 345(6)	3 201(
observed	C(26)	743(3)	3 859(6)	5 139(3)	O(53)	2 417(2)	2 613(5)	3 337(-
Final R								
Final R'	(b) Compound	(2) (×10 ⁴)						
	Ni	5 110(1)	3 357(1)	7 658(1)	N(31)	2 270(6)	1 531(4)	7 437(
	N(01)	4 516(6)	1 860(4)	6 658(3)	N(32)	2 763(6)	2 672(4)	7 868(
	C(10)	5 885(9)	995(6)	7 036(5)	C(33)	1 964(8)	2 706(6)	8 528(4
	N(11)	6 173(6)	936(4)	8 049(4)	C(34)	897(8)	1 577(7)	8 516(
	N(12)	6 420(6)	2 047(5)	8 549(3)	C(35)	1 109(8)	827(6)	7 812(4
	C(13)	6 802(8)	1 740(7)	9 452(5)	C(36)	2 223(9)	3 805(7)	9 165(
	C(14)	6 405(0)	4/6(8)	9 522(6)	N(40)	7 521(3)	- 423(6)	8 2600
	C(15)	7 092(10)	2 741(8)	10 200(5)	0(41)	6 084(6)	4 914(4)	8 519/
	C(16)	6 208(11)	-1 311(6)	8 281(6)	0(42)	8 610(7)	5 982(5)	8 553(
	C(20)	4 684(10)	2 395(6)	5 790(5)	O(43)	7 666(5)	4 302(4)	7 6900
	N(21)	3 777(7)	3 513(5)	5 656(4)	N(50A)	1 902(9)	8 557(6)	5 0480
	N(22)	3 928(6)	4 195(5)	6 453(4)	O(51A)	1 307(7)	7 942(6)	4 339(
	C(23)	2 971(9)	5 137(6)	6 121(6)	O(52A)	3 333(10)	8 780(8)	5 620(
	C(24)	2 311(10)	5 059(8)	5 134(6)	O(53A)	573(14)	8 512(6)	5 431(
	C(25)	2 805(9)	3 985(7)	4 857(5)	N(50B)	2 032(9)	8 882(6)	4 874(
	C(26)	2 759(11)	6 060(7)	6 797(6)	O(51B)	1 353(7)	8 323(6)	4 169(
	C(27)	2 488(11)	3 386(8)	3 912(5)	O(52B)	3 281(10)	9 440(8)	5 053(3
	C(30)	2 652(8)	1 313(6)	6 551(4)	O(53B)	1 492(14)	8 544(6)	5 645(6

Kleywegt et al. (1985) J. Chem. Soc., Dalton Trans, 2177-2184 doi:10.1039/DT9850002177

Structures Digitally Deposited cf Published



CSD entries with 3D coordinates (1936-2013) CCDC Number used as an indication of digital deposit

The growth of electronic deposition



Hand-typed tables of coordinates

178					J. CHEN	1. SOC. DALTON	trans. 1985	
able 1. Crystallograph	ic data and detai	ls of data collectio	n and processing	for ML(NO3)2, with	M = Cu [in	(1)], Ni [in (2)], a	nd Cd [in (3)]	
		(1)		(2)		(3)		
Stoicheiometry		C18H27Cul	N ₉ O ₆	C18H27N9NiO6		C18H27CdN9O6		
M Lattice turns		529.01		524.18		577.88		
Space group	J. CHEM. SOC.	DALTON TRANS	. 1985					21
a/A								
c/Å	Table 2 Atomic	co.ordinates with	estimated standar	d deviations in parent	herer			
a/°	Table & Atomic	co-ordinates with	estimated standar	a deviations in parent	Incaca			
β/°	Atom	X/a	Y/b	Z/c	Atom	X/a	Y/b	Z/c
γ/°	(a) Compo	und (1) (×10 ⁵ for	Cu, × 10 ⁴ for oth	ers)				
U/A ³	Cu	7 734(2)	11 519(5)	62 931(2)	C(27)	2 251(3)	671(8)	4 844(4)
2	N(01)	1 493(1)	-122(3)	6 454(2)	C(30)	2 044(2)	483(5)	6 905(2)
$D_{e}/g \text{ cm}^{\circ}$	C(10)	1 359(2)	-1 289(5)	6 758(3)	N(31)	1 888(1)	1 149(3)	7 431(2)
r(000)	N(11)	708(1)	-1 425(3)	6 566(2)	N(32)	1 386(1)	1 906(3)	7 246(2)
Approximate of	N(12)	379(1)	- 356(3)	6 561(2)	C(33)	1 395(2)	2 512(4)	7 809(2)
Approximate c	C(13)	-162(2)	- 764(4)	6 558(2)	C(34)	1 907(2)	2 157(5)	8 332(2)
Number of cett	C(14)	-171(2)	-2 078(4)	6 570(2)	C(35)	2 220(2)	1 302(4)	8 094(2)
A range/° (cell	C(15)	386(2)	-2 475(4)	6 586(2)	C(36)	918(3)	3 421(6)	7 819(3)
o range/ (cento	C(16)	-651(2)	142(7)	6 546(4)	C(37)	2 788(2)	622(6)	8 412(3)
h range	C(17)	650(3)	-3 767(5)	6 606(4)	N(40)	- 304(2)	2 265(5)	5 590(3)
k range	C(20)	1 549(2)	- 339(5)	5 777(2)	0(41)	136(1)	2 399(3)	6 163(2)
/ range	N(21)	1 509(2)	902(3)	5 458(2)	0(42)	- /1/(2)	3 031(5)	5 480(3)
Number of refl	N(22)	1 (099(1)	2 745(4)	5 333(2)	O(43)	- 272(2)	1 383(3)	3 21/(2)
measured	C(23)	1 530(3)	2 539(6)	4 854(2)	O(51)	1 968(2)	4 072(5)	3 604(3)
independent	C(24)	1 786(2)	1 358(5)	5 035(2)	0(57)	1 526(2)	2 345(6)	3 201(2)
observed	C(25)	743(3)	3 859(6)	5 139(3)	O(52)	2 417(2)	2 613(5)	3 337(4)
Final R	0(20)	(10(0)	5 007(0)	0 105(0)	-()	(-)	(-)	
Final R'	(b) Compound ((2) (×10 ⁴)						
	Ni	\$ 110(1)	3 357(1)	7.658(1)	N(31)	2 220(6)	1 531(4)	7 437(3)
	N(01)	4 516(6)	1 860(4)	6 658(3)	N(37)	2 763(6)	2 672(4)	7 868(3)
	C(10)	5 885(9)	995(6)	7.036(5)	C(33)	1 964(8)	2 706(6)	8 528(4)
	N(11)	6 173(6)	936(4)	8 049(4)	C(34)	897(8)	1 577(7)	8 516(5)
	N(12)	6 420(6)	2 047(5)	8 549(3)	C(35)	1 109(8)	827(6)	7 812(4)
	C(13)	6 802(8)	1 740(7)	9 452(5)	C(36)	2 223(9)	3 805(7)	9 165(5)
	C(14)	6 818(11)	476(8)	9 522(6)	C(37)	326(10)	-423(6)	7 459(6)
	C(15)	6 405(9)	-15(6)	8 625(5)	N(40)	7 521(8)	5 107(5)	8 260(4)
	C(16)	7 092(10)	2 741(8)	10 200(5)	O(41)	6 084(6)	4 914(4)	8 518(3)
	C(17)	6 208(11)	-1 311(6)	8 281(6)	O(42)	8 610(7)	5 982(5)	8 553(4)
	C(20)	4 684(10)	2 395(6)	5 790(5)	O(43)	7 666(5)	4 302(4)	7 690(3)
	N(21)	3 777(7)	3 513(5)	5 656(4)	N(50A)	1 902(9)	8 557(6)	5 048(5)
	N(22)	3 928(6)	4 195(5)	6 453(4)	O(51A)	1 307(7)	7 942(6)	4 339(5)
	C(23)	2 971(9)	5 137(6)	6 121(6)	O(52A)	3 333(10)	8 780(8)	5 620(7)
	C(24)	2 311(10)	5 059(8)	5 134(6)	O(53A)	573(14)	8 512(6)	5 431(6)
	C(25)	2 805(9)	3 985(7)	4 857(5)	N(50B)	2 032(9)	8 882(6)	4 8 74(5)
	C(26)	2 759(11)	6 060(7)	D /97(D)	O(51B)	1 333(7)	8 323(6)	4 169(5)
	C(27)	2 488(11)	3 386(8)	5 912(5)	O(52B)	3 281(10)	9 440(8)	5 053(7)
	C(30)	2 652(8)	1 313(6)	6 331(4)	O(33B)	1 492(14)	8 244(6)	o 645(6)

Kleywegt et al. (1985) J. Chem. Soc., Dalton Trans, 2177-2184 doi:10.1039/DT9850002177

Structures Digitally Deposited cf Published



CSD entries with 3D coordinates (1936-2013) CCDC Number used as an indication of digital deposit



http://www.ccdc.cam.ac.uk/NewsandEvents/News/Pages/NewsItem.aspx?newsid=40

IUCr checkCIF and CSD-Deposit

- Deposition to CCDC now includes:
 - Generation of checkCIF reports
 - Available to depositor during submission
 - Stored at CCDC at submission
 - Soon to be available to publishers and referees during peer review process
 - An easy way to embed validation responses into CIF
 - Retrieval of edited CIFs and checkCIF reports
- Launched July 2015



Cambridge Crystallographic Data Centre	lidation Add Publication Ei	nhance Data Review	Submit
Validation	istency and integrity of t	your structures	
Structure	🕄 💟 IUCr	checkCIF	
MultipleCifs.cif			
data_structure_1	View report	Enter response	
data_structure_2	View report	Enter response	
data_structure_3	View report	Enter response	
Single.cif			
data_structure_4	View report	Enter response	
Go Back V Proce	ed to Next Step		

Level A	Most likely a serious problem - resolve or explain
Level B	A potentially serious problem, consider carefully
Level C	Check. Ensure it is not caused by an omission or oversight
Level G	General information/check it is not something unexpected



Capturing additional metadata



PICK A STRUCTURE TO EDIT	3D VISUALISER	CSD FIELDS
Example_1.cif		Compound name 😧
• data_5_93GPa		2-hydroxybenzaldehyde oxime
• data_tBu10kbar		Synonyms/other names 😵
		salicylaldoxime
		Crystal colour 😧
		Crystal habit 🕄
		block
	JSmol	Space group 😧
CIF		P 1 21/n 1
1 data_5_93GPa 2	A	Study temperature (K) 😵
3 _audit_creation_method 4 _chemical_name_systematic	CRYSTALS_ver_12.80 '2-hydroxybenzaldehyde oxime'	293
5 _chemical_name_common 6 _chemical_formula_moiety 7 _ shemical_formula_sum	salicylaldoxime 'C7 H7 N1 O2'	Formula moiety 😧
8 _chemical_formula_weight 9 _chemical_melting_point	137.14 59-61	C7 H7 N1 O2
10 _symmetry_cell_setting 11 _symmetry_space_group_name_H-M	Monoclinic 'P 1 21/n 1 '	Formula sum 😌
12 _symmetry_space_group_name_Hall 13 loop_	'-P 2yn'	C7 H7 N1 O2



How you capture metadata is important

- Important to capture metadata in a semantic or defined way.
- Free text is not necessarily a good thing
- There can also be confusion over meaning for example the morphology as grown or as cut?



Study Temperature relative to Melting Point



Plot based on data as initially deposited and before validation at CCDC

CSD-Xpedite







- Syntax errors
- Inaccuracies
- Revisions
- Republications
- Timely release

Over 70,000 structures deposited annually

The CCDC's internal Systems automate most steps involved in processing deposited CIFs but manual intervention sometimes needed.



CSD-Xpedite- Metadata used internally

CSD-Xpedite

- Deployed 2013
- Based on CRM
- Used to manage all the transactions that go on behind the scenes
- Number of different entities/record types
- Each entity has metadata associated with it
- Entities include contact, deposit, CSD entry, journal, publication, documentation

Deposition metadata

1374873	-DEP							
Name * 1374873-DEP Depositor Thomas McGuire Email New web deposit				Deposition Coordi CCDC Deposit	nator"			
Status Details								
Status Reason Processing Complete Structure Factor Matchin Yes	g Completed			User Supplied Data Required	Chec	k		
Import								
Catdep Email Reference				Catdep Original To	tal			
Curated Data								+ =
Name Deposit	CCDC Number	Source File (Externa	l Format)	Status Reason	Re	gister Results.	Register Decisions Present	Created On
1939893 1374873	DEP 141013	file1_NR9003.nf.cif		Added to WebCSD	No		No	02/07/2015 1
Documentation								
Name 🛧	Deposit	File Type	File Format	Status Reason	Cre	ated On		
citation_file.citx	1374873-DEP	Citation	chemical/	Valid	02/	07/2015 1		
file1_NR9003.nf.cif	1374873-DEP	Structure File	chemical/	Converted	02/	07/2015 1		
Administration								
Administration Details								
Created By	CCDC Deposit			Created On		02/07/2015	10.56	
Modified By	CCDC Deposit			Modified On		02/07/2015	10:57	
Notes								
NOTES								
Enter a note								

Publication metadata

PUBLICATION : IF	FOR	MATION		
Private	C	ommunication (2015)		
Name		Private Communication (2015)		
Journal *		Private Communication	DOL	
Year		2015	DOI Service search of	Yes
Volume			JDS Number	1.00
First page		**	Last page	
IUCr Data Validatio	ē.		IUCr Co-editor Code	**
Manuscript			CFC	
Authors		Tracey Hewat, Neil Robertson, Nick Funnel		
Ref				
Internal Publication	n		Ref Type	
Status				
Created On		02/07/2015 10:56	Created By	CCDC Deposit
Modified On		30/07/2015 13:25	Modified By	Seth Wiggin

CSD entry metadata

DEACTIVATE															
CURATED DATA : IN	FORM	NOITAN													
1939893	3-(CUR									Activ	/e		■ 1410	.33
Name*															
1939893-CUR Is-Published CCDC N	lumbi	tr .				Ad	ditional CCDC	Number							
- int To Re Processed						 Ni	ot To Re Proces	sed Reason							
						-	or to be truces	and means							
imbargoed Date						10 10	nbargoed Emai	I Sent Date							
	A 1	3.8097(9)	ь			16.4640	(10)		c		8	13.7683(2)		
Alpha	₽ 9	0	Beta			121.445	i(5)		Gamma			90			
Space Group	■ (2/c	R-Fact	De .		7.78			Temperature		8	150.00			
formula Jompound Name		28 H32 Cu1 N4 1+,Cl1 O iis(4,4',6,6'-tetramethyl-2	14 1- ?,2'-bipyridine)	copper	perchlora	te unknov	vn solvate								
Active Publications	×													+	
Search far records			Q												
Name ↑ Private Communicatio 1. Refcodes by State Refcode ↑	in us Cu	Authors Tracey Hewat; Neil Robe rated D Y	ertson, Nick Fur Curated	nel Y	Publicat	t page	Journal Private Comm Status	Acce	Vol Year	Cre 2015 Modified C	02/07	0n /2015 10:5 Modifie	6 30) d By	dified On 107/2015 : Create	
Name 1 Private Communicatic 1. Refcodes by State Refcode 1 YUNPOO	in	Authors Tracey Hewat, Nei Robe rated D Y	rtson, Nick Fur Curated 1939893-	nel Y Record CUR	Publicat	t page ion	Journal Private Comm Status Active	Acce	Vol Year	Cre 2015 Modified C	02/07	0n /2015 10.5 Modifie CCDC De	Mo 6 30 d By eposit	dified On 107/2015 : Create 02/	
Name + Private Communicatic I. Refcodes by State Refcode + YUNPOO CCDC DOI	in	Authors Tracey Hewat, Neil Robe rated D Y	ertson, Nick Fur Curated 1939893-	l c nel Y Record CUR	OOL Firs es Publicat Private C	t page	Journal Private Comm Status Active	Acce 0	Vol Year	Cre 2015 Modified C	oz/07)n /2015 10.5 Modifie CCDC De	Mo 6 30, d By sposit	dified On 107/2015 : Create 02/	
Name 1 Private Communicatio 1. Refcodes by State Refcode 1 VUNPOO CCDC DOI CCDC DOI	in us Cu	Authors Tracey Hewat, Neil Robe rated D Y 10.5517/cc1jbc5d	Curated 1939893-	l Record CUR	OL. Firs	t page	Journal Private Comm Status Active	Acce 0	Vol Year	Cre 2015 Modified C	o2/07	Dn /2015 10.5 Modifier CCDC De	Mo 6 30, d By eposit	dified On 107/2015 : Create 02/	
Name \uparrow Private Communicatio 1. Refcodes by State Refcode \uparrow YUNPOO CCDC DOI CCDC DOI Source Information	en us Cu	Authors Tracey Hewat, Neil Robe rated DY	rtson, Nick Fur Curated 1939893-	nel Y Record CUR	OGL. Firs es Publicat Private C	t page ion	Journal Private Comm Status Active	Acce 0	Vol Year	Cre 2015 Modified C 32/07/2015	eated 0 02/07	n /2015 10.5 Modifie CCDC De	Mo 6 30, d By eposit	dified On 107/2015 : Create 02/	
Name \uparrow Private Communicatio	us Cu	Authors Tracey Hewat, Neil Robe rated D Y 10.55317/ec.ljbc5d 1174871-DEP	ertson, Nick Fur Curated 1939893-	nel Y Record CUR	OCL. Firs es Publicat Private C	t page ion iomm	Journal Private Comm Status Active	Acce 0	Vol Year	Cre 2015 Modified C 32/07/2015	eated 0 02/07	n /2015 10.5 Modifier CCDC De	Mo 6 30, d By eposit	dified On 107/2015 : Create 02/	
Name \uparrow Private Communicatio 1. Refcodes by State Refcode \uparrow VUNPOO CDC DOI CCDC DOI Source Information Deposit Source File	in us Cu	Authors Tracey Hewat, Neil Robe rated D V 10.55317/cc1jbc5d 1174873-08P fibe1 NR3003.nf.off	ertson, Nick Fur	tecord	OCL. Firs	t page ion iomm	Journal Private Comm Status Active Web Deposit Id	Acce 0	Vol Yeas ssion Date 2/07/2015 (data_NR900	Cre 2015 Modified C 20/07/2015	eated 0 02/07	In V2015 10-5 Modifier CCDC De	Mo 6 30, d By eposit	dified On 107/2015 : Create 02/	
Name ↑ Private Communicatio 1. Refcodes by State Refcode ↑ YUNPOO CCDC DOI CCDC DOI CCDC DOI Source Information Deposit Source File Administrati	as Cu as Cu a a a a	Authors Tracey Hewat, Neil Robe rated D	rtson, Nick Fur	CUR	OCL. Firs	t page ion	Journal Private Comm Status Active Web Deposit Is \$\$ Source File Sit	Acce d ction	Vol Yeaa asion Date 2/07/2015 (data_NR900	Cre 2015 Modified C 22/07/2015	eated 0 02/07	n /2015 10.5 Modifie CCDC D	Mo 6 30, d By eposit	dified On 107/2015 : Create 02/	
Name A Private Communication 1. Refcodes by Stath Refcode A YUNPOO CCDC DOI CCDC DOI CCDC DOI Source Information Deposit Source File Administrati Status Details	as Cu as Cu a a a	Authors Tracey Hewat, Neil Robe rated D	rtson, Nick Fur	CUR	OL., Firs	t page	Journal Private Comm Status Active Web Deposit Is \$Source File Sit	Acce d c	Vol Yeaa asion Date 2/07/2015 (data_NR900	Cre 2015 Modified C 22/07/2015	02/07	n /2015 10.5 Modifie CCDC De	Mo 6 30, d By eposit	dified On 107/2015 : Create 02/	* d On)77/20
Name A Private Communication 1. Refcodes by State Refcode A VUNPOO CCDC DOI CCDC DOI CCDC DOI Source Information Deposit Source Information Deposit	as Cu	Authors Tracey Hevat, Neil Robe ated D	ertson, Nick Fun Cunsted 1938893-	CUR	XOL., First	t page	Journal Private Comm Status Active Web Deposit Id	Acce C	Vol Year asion Date 2/07/2015 data_NR900 Decifer	Cre 2015 Modified C 22/07/2015	02/07	In /2015 10.5	Mo 6 30, d By eposit	dified On 07/2015 : Create 02/	e d On 07/20
Name * Private Communicatio Refloade * YUNPOO CCDC DOI CCDC DOI CCDC DOI Source Information Deposit Source Information Deposit Source File Addministrati Stato Detaib Stato Detaib Stato Detaib	as Cu as Cu a a a ion	Authors Tracey Hevat, Nell Robe rated D v 10.5532/vc18e5d 1324821-069 fin1 100003.art.of	Unated 1939893-	CLFer Te: No	DOL., First es Publicat Private C	f page	Journal Private Comm Status Active Web Deposit Id	Acce C C	Vol Year ssion Date 2/07/2015 (2/07/2015 (data_NR900 Decifer 2.0 Resistors	Cre 2015 Modified C 32/07/2015 3 Version by Score	eated C 02/07)n /2015 10.5 Modifie cCDC D4	Mo 6 30, d By eposit	dified On Create 02/	
Name + Private Communication Private Communication Refoode + VUNPOO CCDC DOI CCDC DOI CCDC DOI CCDC DOI Source Information Deposit Source Information Status Reason # Added 5 Wolfs Visited by Editor # Added 5 Wolfs	in us Cu a a ion sp	Autors Taray Head, Kel Robe and D	ertson, Nick Fun Curated 1939893- De De Va Va	CLFer Te: No No	DOL. First es Publicat Private C	f page	Journal Private Comm Status Active Web Deposit Is \$50urce File St	Acce C	Vol Year ssion Date 2/07/2015 0 2/07/2015 0 data_NR900 Beliabit M 1	Cre Cre Cre Solution Compared Co	eated C 02/07	In V2015 10.5	Moo	dified On Create 02/	
Name + Private Communication Reflocade y team vuniPCO CCDC DOI CCDC DOI CCDC DOI CCDC DOI Source Information Deposit Source File Addministrati Status Details Datas Reason & Addministrati Status Reason & Addministrati & Addministrati Status Reason & Status Reason & Sta	as Cu as Cu a a a a s o n	Autors Taray Head, Nel Robe and D	Cursted 1936893- 1936893- De Person Va Pa Pa Pa Pa Pa Pa Pa Pa Pa Pa Pa Pa Pa	CLFer Te No No No	es Publicat Private C Private C Deferred Cisions Pn	f page ion iom	Journal Private Comn Status Active Web Deposit Ia	Acce C C	Vol Year asion Date 2/07/2015 (2/07/2015 (data_NR900 Decifer 2.0 Reliabiti Begietz Registra	t Cre 2015 Modified C 22/07/2015	ated 0 02/07	In V2015 10.5	Moo	dified On 07/2015 : 02/	d On 077/20
Name + Private Communication Refunde + VUNPOO CCDC DOI CCDC DOI CCDC DOI Source File Address Volkans Status Reason # Added 5 Wolkans Refuger Perults # New Series	as Cu as Cu as Cu as cu so so	Austron Taxay Head, Kei Rabd and D V 10.3337/sclackd 1174072-002 Micl. N90003 and d	Landon, Nick Fur Curated 1939893-	CLFerTe No gisterDe No	es Publicate stopping topping	f page ion omm	Journal Private Comn Status Active Web Deposit Id ¥Source File St	Acce 0	Vol. Year asion Date 2/07/2015 data_NR900 Decifer 2.0 Reliabit B registr B res	Crack C	lete	In V2015 10.5	Moo 6 30,	dified On 02//2015 : 02/	d On 07/20
Name + Na	as Cu	Autors Taray Head, Kel Robe and D	Prtson, Nick Fun Curated 1939893- 1939893- De Que Que Que Que Que Que Que Que Que Qu	CLFerTe No Sister Do No Synchronis Synchronis Synchronis	es Publicat Private C Private C Private C Conly Deferred cisions Pn ation Status	t page	Journal Private Comn Status Active	Acce C	Vol. Year asion Date 2/07/2015 data_NR900 data_NR900 data_NR900 data_NR900 data_NR900 data_NR900 data_NR900 data_NR900 data_NR900 data_NR900	Cracial Control Control Control Control Control Compared Control	eated C 02/07	In V2015 10.5	Moo	dified On 007/2015 : 02/4	* d On 07/20
Name + Name + Next Communication Retools by Stat Retools + VUNPOO CCCC DOI CCCC DOI CCCCC DOI CCCC DOI CCCC DOI CCCC DOI CCCC DOI CCCC DOI CCC	as Cu	Autors Taray Head, Nel Robe and D	ertson, Nick Fur Curated 1939893- 1939893- De B B B B B B B B B B B B B B B B B B	CLFer Te: No Synchroniss Synchroni	es Publicat First es et Publicat control et Conly et Conl	t page	Journal Private Comn Status Active	I ection	Vol Vea ssion Date 2/07/2015 (2/07/2015 (2/07/20) (2/0)	t Creation Modified Control 202025	eated C 02/07) n Modifie	Moo	dified On 07/2015 : 02/4	* 0 0n 077/200
Name + Name + Name + Name + Name + Name + Name - Na	as Cu as Cu a a a a son so	Autors Tracy Head, Nel Robert and Du V IBASSIT/AcQueSel BIASSIT/AcQueSel Bia INDODUATO Bia INDODUATO	ertson, Nick Fur Curated 1939893- 1939893- De De De De De De De De De De De De De	CIFer Te No Synchronis Synchro	boL. First es Publicat Private C Private C Dublicat Conly Deferred cisions Pn ation Status	t page	Journal Private Comm Status Active Web Deposit la ¥Source File St	I ection	VoL. Veara ssion Date Vol. Veara 20207/2015 (Vol. Veara Relativity B 1 Registra B 1 Registra 1 Registra B 1 Registra 1 Registra B 1 Re	Cree Cree Control Cont	eated C 02/07) In	Moo	dified On 107/2015 : 02/	* d On 07/20
Name + Na	ion SD	Autors Taray Head, Nel Robe 10.5532/scillectel 13.5	Curated 1939893- 1939893- De Que No No No No No No No No No No No No No	CLFer Te No Synchronis Synchron	boL. First es Publicat Private C Private C Daferred ation Statu	t page	Journal Private Comm Status Active Web Deposit Id ¥Source File St Created On Modified On	I Acce	Vol Vear ssion Date 2/07/2015 (data_3x8000 Decifer 2.8 alabi 1 alabi 2/07/2015 2 alabi 2/07/2015 2 alabi 2/07/2015 2 alabi 2/07/2015	Cree Cree Control Cont	stelled	In Modified	Moo	dified On 107/2015 : 02/	• d On)77/20:

Assignment of chemistry



- Automated processes based on structural knowledge
- Manual validation by expert editorial staff







1965 - 201

loop_ atom_site_label atom site type symbol atom site fract x atom site fract y atom site fract z atom site U iso or equiv atom site adp type atom_site_occupancy atom site symmetry multiplicity atom site calc flag atom site refinement flags atom site disorder assembly atom site disorder group Cl1 Cl 0.5993(2) 1.0007(7) 0.8131(17) 0.044(3) Uani 0.50 PDU A 1 0.8260(6) 0.9322(3) 0.0327(11) Uani 0.50 1 d PDU A 1 0.8802(9) 0.8184(9) 0.029(4) Uani 0.50 1 d PDU A 1 0.5286(7) 0.8174(18) 0.7440(7) 0.031(4) Uani 0.50 1 d PDU A 1 C3 C H3A H 0.5350 0.8343 0.6771 0.037 Uiso 0.50 1 calc PR A 1 0.4918(8) 0.7220(19) 0.7783(8) 0.027(4) Uani 0.50 1 d PDU A 1 0.4900(6) 0.7171(14) 0.8779(9) 0.029(4) Uani 0.50 1 d PDU A 1 C12 C1 0.3202(2) 0.4982(6) 1.0830(5) 0.0586(15) Uani 0.50 1 d PDU A 1 S2 S 0.38755(19) 0.6658(5) 0.9578(5) 0.0400(10) Uani 0.50 1 d PDU A 1

Assignment of chemistry is important for discovery, data mining , analysis and interoperability

Data that describes the substance







Data that describes the dataset as a whole

- Important for provenance and attribution
- Includes:
 - Publication data
 - Could be in the scientific literature and/or repository
 - Authorship data
 - Creation of the sample, data collection, data analysis









Streed or out a 27.501(6/4 add 350(0/4 cald 350(0/4 cald



Metadata - Associating CSD data to publications



Metadata exchanges

- Automated workflows with:
 - Publishers Third Parties

Publication metadata at CCDC

Journal *	CrystEngComm	DOI	10.1039/C4CE02564D	
Year	2015	DOI Service search co	Yes	
Volume	17	JDS Number		
First page	5377	Last page		
IUCr Data Validation		IUCr Co-editor Code		
Manuscript		CFC		
Authors	Conrad W. Ingram, Geoffr Varma H. Rambaran, Bran Zhang	rey Kibakaya, John Bacsa, Stej Idon Dennis, Esmeralda Casta	phan R. Mathis,II, Alvin A. Hok neda, Julianne S. Robbins, Z. J	ler, ohr

Metadata from publishers

<doi>doi:10.1107/S2056989015014383</doi>

- <title></title>

Crystal structure of 6,6,12,12-tetrachlorotricyclo[8.2.0.04,7]dodecane-5,11-dione

<journal>urn:issn:2056-9890</journal>

<coro>xu5862</coro>

<author>Turan Akm, E.</author>

<author>ThKelek, T.</author>

<volume>71</volume>

<issue>9</issue>

<firstpage>1000</firstpage>

<lastpage>1002</lastpage>

<published>2015-08-06</published>

<ccdc>1415865</ccdc>

Publication metadata from third parties

cros	ssref.org
ABC	DUT CROSSREF) (FOR PUBLISHERS) (FOR LIBRARIES) (FOR AFFILIATES) (FOR RESEARCHERS)
A DOI qu	very 0
Select re	sult format xml-xsd: QR unixref:
DOI:	
	Search
XML Result 2	<pre>cfult_tite>CrystBngCommc/full_tite></pre>
Bil Thi inte obt	Iliographic metadata search s form is a guest query interface to the CrossRef system for individual DOI retrieval. This interface is no ended for automated querying. If you would like to query CrossRef on an automated batch basis, please tain an account on our system.
of	j must supply either author or first page and we recommend using journal title as well as ISSN, for a lis journal titles in the CrossRef holdings please visit <mark>our browsable journal list.</mark>
Lie	
	Southan OK Southan OK Southerence Proceeding
Fir	st Author ISSN
Fir	ISSN ISSN
Fir	st Author ISSN
Fir Joi Ari	st Author ISSN urnal Title ticle Title Urme Issue Rade Year
Fir Joi Ari Vo	st Author ISSN urnal Title ticle Title lume Issue Page Year
Fir Joi Ar Vo	st Author ISSN urnal Title liume Issue Page Year BN Component Number
Fir Joi Ar Vo IS	st Author ISSN urnal Title lume Issue Page Year BN Component Number

Search

Clear Form



Metadata enabling referee access to data

From: 418e7b.0b9f8125@editorialmanager.com Sent: 04 March 2015 03:45 Subject: Manuscript

In view of your expertise I would be very gratefu following manuscript ...

If this paper includes new structural information, available from one of the following databases:

CCDC (organic and metal-organic compounds): http://www.ccdc.cam.ac.uk/data_request/referee.access

Fiz Karlsruhe (inorganic crystal structures): <u>http://www.fiz-informationsdienste.de/en/DB/icsd/depot.html</u>

PDB (proteins): <u>http://www.rcsb.org/pdb/</u>





Metadata used to link articles and data





Elsevier, Wiley and IUCr journals

CCDC crystal structure DOIs

- Over 560,000 dataset DOIs registered since April 2014
- DataCite metadata includes related article DOIs
- Foundation for formalising data citation and interoperability
- Data Citation Principles Data should be considered legitimate, citable products of research...

 Synthesis, Structure, and Catalytic Studies of Palladium and Platinum Bis-Sulfoxide Complexes

 E. Postgrad, A.Crystallographer and R. Supervisor

 Abstract
 DOI: 10.1021/jacs.5b03972

 Coordination-directed self-assembly has become a well-established technique for the construction of functional supramolecular structures. In contrast to the most often exploited transition metals, trivalent lanthanides Ln^{III} have been less utilized in the design of polynuclear self-assembled structures despite the wealth of stimulating[...]

 Supporting Information

🧐 Crystallographic Information File 🛛 🔁 PDF

ja5b03972_si_001.cif (2.93 MB) > ja5b03972_si_002.pdf (2.48 MB)

Crystallographer doesn't always get credited for their contribution.

DataCite

Data Citation Principles

10.5517/CCPHZ37

https://www.force11.org/datacitation

Article DOI primarily identifies the paper and not the data.

Crystal structure data often buried amongst the supporting information.



Provenance and attribution







Sign In 👻 Help

DOIs – facilitating interoperability

Active filters (X clear all):	×			
datacentre BL.CCDC - The	e Cambridge Crystallogr	aphic Data Centre		
560521 documents found i Page 1 of 56053 🔶 🦛 🖨	n 754ms ▶ 📦			
CCDC 651833: Expe \$doi:10.5517/CCPW8WP Zakharova, V.M. • Tanarro	o, C.M.G. • Gutschow, N	ucture Determination# 11. • Hennig, L. • Sieler, J. • (et.		
ai.)	doi:10.5517/CCP	W8WP		
CCDC 651836: Expe	This page represents Data	Cite's metadata for <i>doi:10.5517/CCPW</i>	8WP.	
CCDC 658558: Expe &doi:10.5517/CCQ38TW	Citation	Zakharova, V.M.; Tanarro, C.M.G.; Gu	Itschow, M.;	
CCDC 672176: Expe		Experimental Crystal Structure Deterr	08): CCDC 651833:	
 doi:10.5517/CCQKG3T Sun, Yun • Wang, Xiaozh 		Cambridge Crystallographic Data Cer	ntre. http://dx.doi.org	
CCDC 644817: Expe	Descriptions	TU.55T//CCPW8WP RIS BIBTEX		
In the second	Abstract	An entry from the Cambridge Structur	al Database, the	_
		world's repository for small molecule of	crystal structures.	
		The entry contains experimental data	from a crystal	
		diffraction study. The deposited datas	et for this entry is	
		freely available from the CCDC and ty	pically includes 3D	
		coordinates, cell parameters, space q	roup, experimental	



InCites™ Journal Citation Reports® Essential Science Indicators M EndNote

entry contains experimental data from a crystal diffraction study. The deposited dataset for this entry is freely available from the CCDC and typically includes 3D coordinates, cell parameters, space group, experimental conditions and quality measures.

Keywords

Author Keywords: Crystal structure; Experimental 3D Coordinates; Crystal System; Space Group; Cell Parameters



Coverage of CCDC data by the **Thomson Reuters Data Citation** Index achieved via the DataCite metadata store.



Facilitates formal citation AND interoperability. Currently adding indication of full citation and resourceType to metadata

conditions and quality measures

25

DOIs – facilitating interoperability





OpenAIRE CrossRef OpenAIRE Resolver PubMed





UK Research Data Discovery Service: Aim is to aggregate metadata for research data held within UK universities and national, discipline specific data centres.



RDA-WDS Publishing Data Services prototype taking a feed of CCDC article-dataset links from DataCite.



Data literature inter-inking service (beta) that ingests data article -inks. Inspection portal : <u>http://dliservice.research-</u> <u>infrastructures.eu/</u>

Linking based on chemistry



Bulk identification of links between ChemSpider molecules and CSD Crystal Structures (~50,000) facilitated by InChIs.



Standard InChI:

InchI=1s/C33H3606/c1-7-10-37-31-19-25-13-23-17-29(35-5)33(39-12-9-3)21-27(23)15-24-18-30(36-6)32(38-11-8-2)20-26(24)14-22(25)16-28(31)34-4/h7-9,16-21H,1-3,10-15H2,4-6H3

Standard InChIKey: IZHKSTHBLQRIOW-UHFFFAOYSA-N 1965 - 2015

Other crystal data repositories



• PDB

- match PDB ligands to best representative CSD molecules
- working group on ligand validation

29 July, 2015

Data correspondences between the PDB and CSD archives now available

The Worldwide Protein Data Bank and the Cambridge Crystallographic Data Centre (CCDC; http://www.ccdc.cam.ac.uk) are pleased to announce the availability of a new data resource containing correspondences between the biopolymer components and ligand molecules found in the PDB archive that exactly match small-molecule X-ray structures in the Cambridge Structural Database (CSD) archive.

The chemical structure of every unique molecule in the Protein Data Bank is described in the **PDB Chemical Component Dictionary**. The new PDB Chemical Component Model data file complements information in the PDB by providing the following CSD information for matching molecular entries: accession code correspondences, Cartesian coordinates and R-value, data-collection temperature and a disorder flag, SMILES and InChI descriptors, and a Digital Object Identifier (DOI) for the citation associated with the CSD entry.

At present, there are 20,077 chemical components in the PDB Chemical Component Dictionary, and for 1,418 of these exact match structures have been identified in the CSD. The new PDB Chemical Component Model file is available from the PDB FTP archive via:







Linking experiment to data

ISIS: Neutron and muon instruments at the UK Science and Technology Facilities Council Rutherford Appleton Laboratory.



ICAT

🔆 Experiment Data

The ISIS data catalogue

All data produced at ISIS is catalogued into ICAT – the ISIS data catalogue. This catalogue is made available as a searchable website, via Mantid and as an API.

ISIS will issue citable DOIs for all experiments carried out at ISIS. You are encouraged to cite these in publications relating to ISIS experiments. For more information see <u>DOIs</u>



CCDC and STFC investigating the potential for linking between:

- raw data stored at STFC
- modelled structure stored at CCDC

Challenge: reliably associating results with original experiment





Metadata about people and institutions

ORCID	FOR RESEARCHERS	FOR ORGANIZATIONS	ABOUT	HELP
Connecting Research and Researchers	SIGN IN REGISTER FOR AN O	RCID ID LEARN MORE		789,22
Suzanna C. Ward	* Employment			
(D) http://orcid.org/0000-0002-6062-7492	Cambridge Crystallog	raphic Data Centre (2006-	01 to present)	0
Also known as: Suzanna Ward	Manager, Cambridge Stru	ictural Database		
Keywords: crystal	* Works			
	CCDC 1008189: Expe Determination 2014	rimental Crystal Structur	e	0

Assigning DOIs through DataCite allows researchers to add their data to researcher IDs such as ORCiD





Funders are wanting to know the output of their funded research. This will require linking between people, publications, data and funders



Could be used to find institutions research output. Ringgold IDs now in our internal systems and we want to work towards reliably identifying institutions at point of deposition

What next?



Your under pent structure	re deposition and dissemination
develo	
Email Password	Email or iD Password
Sign in	Sign in
Register with CCDC	Register for an ORCID iD

in more than one to be	ent	ind records	containing all th	ie terms ente	ered. More inform	nation
		mber(s) or CS	D refcodes(s)			
deven	A single pu	ublication DOI	or CSD DOI			
Journal 😯	e.g. Journa	al of the Amer	ican Chemical S	ociety		
Publication details	Year	Ø	Volume	0	Page	
Authors 😯	e.g. F.H.Al	llen				
Compound name 😯	e.g. sulfad	diazine				

CSD Deposition Portal

- Integration of Researcher and Institutional IDs
- Exposure of CCDC services during deposition
 - Ability to edit CSD entry metadata such as 2D chemical diagram
 - Extend validation and integrity checks
- Capturing other metadata
 - Topological descriptions, funders, experimental data, links to associated data, etc
- Integration with third party software

Interoperability, links and programmatic access

- Extension of links....
- Extend CSD Python API to access deposited data

CSD-Community Access

- Extension of search functionality
- Addition of InChI lookups

_





Thank you