

ASCA 2018 🥋 CRYSTAL 32

December 2-5 2018 | Auckland, New Zealand

The Cambridge Structural Database – Developments in deposition and access

Matt Lightfoot, Suzanna Ward

The Cambridge Crystallographic Data Centre (CCDC)

4th December 2018



The Cambridge Crystallographic Data Centre

International Data Repository Archive of crystal structure data High quality scientific database Scientific Software Provider Search/analysis/visualisation tools Scientific applications	Employer of ~60 staff Scientific editors Software developers Applications scientists Cambridge UK USA
Collaborative Research Organisation New methodologies Fundamental research	Originated in 1965 Financially self-supporting
Education and Outreach Conferences, Workshops, Bespoke Training, Teaching Materials	University Partner Institute www.ccdc.cam.ac.uk

Dedicated to the advancement of chemistry and crystallography for the public benefit through providing high quality information services and software.

The Cambridge Structural Database



https://www.ccdc.cam.ac.uk/CCDCStats/

 \mathbf{C}



What's in the CSD?

Organic 43%

Metal-Organic 57%

At least one transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po

Not Polymeric 89%

Single
Component
56%Multi
Component
44%

Organic

• Drugs

Polymeric: 11%

- Agrochemicals
- Pigments
- Explosives
- Protein ligands

Metal-Organic

- Metal Organic Frameworks
- Models for new catalysts
- Porous frameworks for gas storage
- Fundamental chemical bonding





$\mathbf{(0)}$

New Zealand









WAZHAJ

$\mathbf{(0)}$

Worldwide data sharing

- 100,000 datasets deposited annually
- >10,000 depositors
- ~ 1 million datasets
- >1,000 journals
 - >450,000 publications
 - >380,000 authors
- Peer review data access
- Data available to all at the point of publication

Countries contributing data



Analysis based on a sampling of e-mail address domains

CS

emistry for Life®





Description Springer



IUCr









Deposition guidelines

- Data deposited pre-publication
 - Deposition when crystallographer most engaged
 - Enables links to data sets at the point of publication
 - Tailored deposition service

Small molecule single crystal data

Authors should present their crystal data in a CIF (Crystallographic Information File) format and deposit any organic or organometallic structural information with the <u>Cambridge Crystallographic Data Centre</u> (CCDC) before they submit their manuscript to us. Data will be held in the CCDC's confidential archive until publication of the article, when data for organic and organometallic compounds will be entered into the Cambridge Structural Database. Authors are encouraged to deposit inorganic crystal structures with the <u>ICSD</u>, hosted by FIZ Karlsruhe.

During submission of a manuscript to the Royal Society of Chemistry using our online submission system, authors will be asked to provide CCDC reference numbers; CIFs should not be submitted with the manuscript (these should have already been deposited with the CCDC/ICSD, see above). Any revised CIFs obtained subsequently should be deposited directly with the CCDC before the revised manuscript is submitted to us. CCDC or ICSD numbers should be included in the manuscript prior to submission.





New CIF guidelines for deposition



The CCDC CIF Deposition Guidelines

When preparing your CIF for deposition please include as much information as possible and check it carefully. This is especially true for <u>CSD Communications</u> where there is no paper to describe the chemistry and experimental details leading to your structure. If you choose to publish your data as a *CSD Communication* please remember to provide all the authors/crystallographers/chemists who contributed to the crystallographic experiment as authors of the data. If we are unable to validate your structure from the information you have provided we may contact you. If we cannot resolve the issue, unfortunately, we may not be able to add your structure to the <u>CSD</u>.

All experimental <u>CIF files</u> (including those from powder diffraction experiments) should contain an R-factor. This should be consistent with the crystallography being performed correctly and to the best ability that would be expected from the material and equipment used. We would like all experimental CIFs to contain:

R-factors (R1, wR2, Rint)

https://www.ccdc.cam.ac.uk/Community/depositastructure/cif-deposition-guidelines/



cten careagan O

10 CR

Compound name O

IN LOCAL DRIVEN

Crystal colour 0

0

Synonyme/other names 😣

Incastor to the company of the Artist

COD Fields

10 G

rendered at 18000

contract and so we an ot the second

400,042,04

Agential (62-1913)

20. α 60. α

1.111.1.1000

6-110 http:

2,000

The Local Device Address

Without the design of the second

descenses of the

Deposit Structures Upload your data to the CCDC for inclusion in the Cambridge **Data Deposition** Structural Database Enhance Data Press sheet the internation behavior such structure extentities and odd os much additional information as possible "edite the CSD fields on the digit hand side other than the CIF deads. Any edite to the CSD fields will update the CIF automatically 0 NORMORON ENIMATED .0 0 hen you have checked each structure placese proceed to the next step. ±Save Changes ↓ Proceed to Not Step Check Syntax Path a descripted section 32 MAN The files that before a set in the left hand entered are shown for beed filing before presseding CCC CFIZ Karlsruhe Register 🚽 Sign 88 📰 checkCF_memoles.cF Plane with an one working spranch the full last relative works for consuming only any type dot the Serie & Serie & Planet Plan home being data_sections data_ac(2016a For state information on loss in the server planes and the contesting GPA longer 0 0 0 001_1025146 0 O. Animate Unit deter son the last Per Maria Mi Paronen Kestik, 243 CIF deposition and validation service TANK OF ANY OTHER DESIGNATION. Manual Incor This web sentce enables you're solonill CF they and associated shurture factor five to the OCDC and for your structures to be included in the College of a started out on the specific consistences Cambridge Disturbane Delaber Deposition attents you to correct against errors, check the integrity of your data and add additional data Control of the second of Control of the second of the s Piesse include exacture factor data for of shadtores. Play should be in CIF, FCF or HPL formed and may be included in a Z P file. dria_se2006c At least one CP, the most to includes in the submission. At this satisfield on one form should consepond to may path/earther only. - Date is a limit of 50 MB per the and a limit of 100 MB for the total size of flive uploaded Arrestory, adult, and an ext. Far more information plates see our Structure Experiment Montentine-page. 2 10.00 made Visit Intelle Q Datente Treat **COST FIZ** Karlsruhe ther more address Q markingly can at all 50 🖬 🖬 🖬 Press shid on the one provings to a single in the bankur of the sear in the CF. Institution (e.g. University/Company) @ octor CCCC CFIZ Karlsruhe 0 O 0 ECDC nembers for residenciane O Add Enhance Baylow Submit Publication Deta CERCEMIL 2 Pilks 0 o Select Has Date 12 Enhorace Data @ invalid_2.df 100% Validation in MultipleCh of 100% Add Publication Mow reports on the consistency and integrity of your structures Datablock: I Details 0 Il Remember my details and periods. 202 - 1.224 A. Please check and add/apdate the publication details shown below Ophiers 😡 10.00 Breed C.C. Line #1 web to fun the RAC checkGH-PLATOY eavise on my data 🏭 If you don't know the full publication details then please provide the current list of authors for the data you an IUCr checkCIF 🛛 🔛 Structure depositing salatan Karatan 00.000 irrealid_2.cif Party Street 10.00 where the set COLUMN TAXAB Authors 😣 John Smith, David Brown O the balance 0 0 0 . data_I 402.00 Enter Response Maliculies. Add Publication Enformer Date Journal Name 😡 Early Inorg, Chem. 278 MultipleCits.cit 0.000 10 M 1 Volume O CIF deposition and validation service ALC: 12 data_I Enter Response Doctorial States COLUMN AND Yeer 😣 2016 sale reas C.475, N. 206 C.475 data II Annual and a second of the second sec No Structure Factor Data have been uploaded. No Response Regi Structure Terror Data are an essential part of the departition. You should click 'So Teok' to add Structure Teccor Tetra to your departition. Him Page O compliand documbances you are anable to include Structure Fredor data with your deposition you may continue by clicking Proceed to Ned data III Resident And American Enter Response Par -- 111 Step Additional Information The interview of the transmission of the Action for the town of the transmission of the Transmission of the town of the town of the town of the town. Proceed to Next Ste If you do not intend to publish the data in the scientific literature and would like to share the data immediately AAlert level 5 Minute TOUGHER AND AND A COMMUNICATION OF THE COMUNICATION OF THE COMMUNICATIO through the Cambridge Structural Database then please click the 'CSD Communication' button below. • Retrieve Deposited Files COD Alert Level C SUTTO, A DT 2, Class 1028 British Average 20,5 Taxor -SUTTO, CLASS CONTRACT, CLASS BRITISH AVERAGE 20,5 Taxor -Deposit Structures Horne Access Structures CSD Communication

9



Ability to identify the data producer

0000	CSD Ent	try: UVESEV01
	3D viewer	Chemical diagram
1 Login 2 Upload 3 Check Syntax 4 Validation 5 Add Publication Add Crystallographer Details Please check and update the details of the main crystallographer associated with this data Use My Details		$HOOC \rightarrow H_{1}C \rightarrow H_{2}C \rightarrow H_{3}C \rightarrow H_{$
Crystallographer Name e.g. Frank Allen Publishing Name e.g. Frank H. Allen Email Address e.g. Frank H. Allen	H Disorder & Menu Open - Style Labels Packing Meas (Bail and Stock * No.Labels * None * None	JSmol H,C,H,H,COOH H,C Ulew group symbols key
Affiliation 3 e.g. CCDC, 12 Union Road, Cambridge, CB2 1EZ	Crystallographer(s)	
Go Back VProceed to Next Step	Crystallographer	Tanya le Roex 💿
	Affiliation	Stellenbosch University

- Details will also embedded into CIFs downloaded from CCDC
- Linked ORCID iD



Integrity and validation checks

	1 Login	2 Upload 3 Check Syr	tax 4 Validation	5 Add Publication	6 Enhance Data	7 Review 8 Submit	
V	alida	ation					
	Structure	e IUCr cl			Unit cell check 🚱 🌜		
	example_	1.cif					
	data_5_93	3GPa View	Report Enter Response		View Hits		
	example_	<u>2</u> .cif					
	data_tBu1	0kbar View	Report Ent tesponse		View Hits		
		Most likely a serious	CheckCIF Responses:		ABANEY01 : 2 (1.3 benzow Space Group: P21/n, Cell: a βΑ ο 13.2075(11)Α ο 13.4578(11)Α, α'90' β 110.9676(11)' γ 90'		
Le	evel A	problem, resolve or explain	PLAT027_diffrn_refins_theta_ful_value (loo) Low 13.60 Degree		3D viewer	Chemical diagram	
		A potentially serious	PLAT029 diffm measured fraction theta full va	ilue Low . 0.677 Note	0		
Le	evel B	problem, consider carefully			and the		
	evel C	Check. Ensure it is not caused	PLAT415 Short Inter D-H., H-X H1 ., H623 ., 2.00	Ang.	68965		
	evere	by an omission or oversight	> Level C		H Disorder di Menu Open -	JSmot	
Le	evel G	General information/check it is not something unexpected			Style Labels Packing Meaa (Ball and Stock V) No Labels V) None V) None	Urew group symbols key	



Data enrichment: By the depositor





Adding data to your Researcher IDs

Connecting Research and Researchers





Deposition portal

- Launched December 1st 2016
 - create a CCDC profile, log on and view and retrieve your depositions
 - deposit new data using your account
- April 2017
 - Edit and update basic information
 - Assign data as CSD Communications
 - Extend embargo periods
- July 2017
 - Share data with co-workers
- October 2017
 - Revise their structures
 - Add links to raw data files

- May 2018
 - Improve the deposition of large files
- July 2018
 - Launch of the joint deposition process for CSD and ICSD structures
- September 2018
 - Ability of add additional files after deposition
- Currently 29,000 users



0000

My Structures

Search by CCDC Number

Clear Filters

Q

After depositing structures to the CCDC they may take a few minutes to appear in the table. If yo please contact deposit@ccdc.cam.ac.uk

To change the displayed columns or filter results you should click on the down arrow of the relev To order the results by a particular column click on the column heading you wish to order your re

	CCDC Nu ~	Data block ~	Deposited On 🛛 🗸	Deposited By	Refcode
	1416019	data_amit63m	12/01/2017	mplightfoot74@g	AHUXUY
	1416026	data_mn467	12/01/2017	mplightfoot74@g	AHUXOS
	1416025	data_l	12/01/2017	mplightfoot74@g	AHUXIM
	1416016	data_l	10/01/2017	mplightfoot74@g	AHUXAE
	926930	data_eu_dotam	10/01/2017	mplightfoot74@g	
	1416015	data_eu_dotam	10/01/2017	mplightfoot74@g	
	1416010	data_1n_DMF_1	21/12/2016	mplightfoot74@g	AHUWIL
	1416009	data_12DMF_90K	21/12/2016	mplightfoot74@g	AHUWEH
	1416008	data 12DMF 180K	21/12/2016	mplightfoot74@g	AHUWAD
M	▲ 1 2 3	4 5 6 7	8 9 10	► ► 20	▼ items per
Dowr	nload Selected 0 Vi	iew 0 in Access Structu	rres Export - Cle	ar selection	

Ay Structure Details	6		
Data block: I Space Group: Pnma, Cell: a 13.167A & Formula: C6 H7 Cl1 N2 O1, Temperatur	b 6.463Å c 8.19Å, α 90.00° β 90.00° γ re: 113 K	90.00°	✓ Chemical Diagram
 Additional CCDC Details 			
CCDC Number	1416025		
CCDC Citation	Determination, 2016, DOI: 10.4124/co	dc.csd.cc1jjh7s	An Original Information
Refcode	AHUXIM		Source information
Compound Name	5-chloro-1,3-dimethyl-1H-pyrazole-4-	carbaldehyde	File Name
Deposited On	12/01/2017		invalid_2_file001.cif
Deposited By	mplightfoot74@gmail.com		File Section
Additional Depositors			data_i
Status	Published in the CSD		L Download Files View in Access Structures
✓ Associated Publications			Source Documents
John Smith, Inorganic Chemistry, 2016 Created on 12/01/2017 Modified on 12/01/2017	, 5, 5234		invalid_2-data_1_file002.html
+ Back to My Structures			
0000	C	SD Entry: AH	JXIM
Your query was: CCDC identifier(s);	1416025, Authors: John Smith and th	e search returned 1 record.	New Search
Results	AHUXIM : New structure under	rgoing enhancement 😧	
Refcode CCDC Number	Space Group: Pnma, Cell: a	13.167(9)Å b 6.463(5)Å c 8.190(6)Å	, α 90.00° β 90.00° γ 90.00°
✓ AHUXIM 1416025			
	3D viewer		Chemical diagram
Download -	H Deorder C Style Education Violander	LSmol Menu Open - Packing Measure Viter + Viter +	NG NG NG NG NG NG NG NG NG NG NG NG NG N
			• For group of mode not
	Additional CCDC details		
	CCDC Number	1416025	
	CCDC Citation	John Smith CCDC 1416025: Exper DOI: 10.4124/ccdc.csd.cc1jjh7s	imental Crystal Structure Determination, 2017,
	Deposited on	12/01/2017	
	Associated publications		
	John Smith, Inorganic	Chemistry, 2016, 5, 5234	



Coming soon in Deposition/My Structures

- Ability to create a subset of your structures
- Ability to share a subset of your structures with your colleagues
- Ability to create a CSD format database of your in-house structures to use in CSDS software.

The Cambridge Crys Data Cent	tallographic re							A Matt Lightfo	ot 🕶					
My St	ructures	6												
Search by D	eposition Number		(Clear Filter	s		🖹 Save Setting	s 🗙 Reset Set	ings					
After depositing contact deposi	g structures they ma t@ccdc.cam.ac.uk displayed columns	ay take a few minut or filter results you	tes to appear in th should click on th	ne table. If you are st ne down arrow of the	ill unable to see relevant colum	e your structures after you have rece n heading.	ived your Depositi	on Numbers pleas	e	Database Requests				
To order the re You can select subsets.	sults by a particular which structures ar	column click on th e displayed by usir	e column heading ng the dropdown t	you wish to order y box below. Different :	our results by. subsets can be	created using the 'Add to Subset' bu	itton and you can i	view or export the	e	Created 02/12/2018	Name Matts new structures	Status Complete	Download	
12			¥	My	Databases	< Associate More Structures	Add to Subset	Anare Struct	ures	09/10/2018	All structures	Downloaded		
	Deposition	Datablock	Deposited On	Deposited By	Refcode	Formula	Status	Subsets 🛓	:					
Details	1636536	data_4	06/02/2018	erogers@ccdc.c	AJIJUA	(C11 H14 N2 Ni1 O5)n,H2 O1	Published in the CSD	x Test Data x Pub data	^					
Details	1636534	data_3	06/02/2018	erogers@ccdc.c	AJIJIO	(C12 H23 N2 Ni1 O5 1+)n,C2 H3 O2 1-	Published in the CSD	x Test Data x Pub data						
Details	1636533	data_1	06/02/2018	erogers@ccdc.c	AJIJEK	(C10 H14 N2 Ni1 O5)n,H2 O1	Published in the CSD	x Test Data x Pub data						
Details	678988		06/01/2009		VOGYUM	C5 H7 N2 1+,C10 H8 Cl4 Cr1 N2 1-,C1 H2 Cl2	Published in the CSD	× Matts new stru						Clos
Details	1637747	data 8257-ICSD	10/07/2018	mplightfoot74@		(Al1 In1 S3)n	Published	x Matts new stru						



CSD Communications



0000

14 4 1

My Structures

Q Citrar Fiber

* * 20 *

ns. Lipst-

O 16 H 34 EUL NB C6 34 3(01 F3

twifalls_ALUW/D_NLOLG1151104154115100

- CSD Communications allow you to publish data directly through the CSD
- Currently 25,766 CSD Communications



A Mattheatine

Extense

Details

ted in the

Year

My Structure Details

	operatione. So is						
V Additional CCDC Deta	is						
CCDC Number	1415881						
Refcode							
Compound Name	2-fluoro-1,2,3-trip	nenylpropan-1-one					
Deposited On	24/08/2016	24/08/2016					
Deposited By	MyCSDtest@ccdd	MyCSDtest@ccdc.cam.ac.uk					
Additional Depositors							
Status	Unpublished		Publish as a CSD Co	mmunicato			
Embargoed Date 😏	27/02/2018			Extern			
✓ Associated Publication	5						
Publication		Created On	Modified On				
	an Dashal F. Musher	10000017	10/02/2017				



Impact on quality of the CSD?

- *CSD Communications* vs. peer-reviewed structures
- Does the quality of the data depend on the source?
- Common question/challenges we hear

How can I rely on data if it hasn't been peer-reviewed?

It isn't the compound I wanted. It's only data that isn't good enough for publication?

Only duplicates are shared through a database?



Selection of data sources

Web of Science

- From Thomson Reuters Web of Knowledge
- High impact factor journals in ISI categories
 - All Journals Sci1, Sci2
 - Chemistry Multidisciplinary Chem1, Chem2, Chem3
 - Crystallography Cryst1, Cryst2
 - Data based journal Data1
 - Data published through CSD communications CSDComms
 - Average all data CSD



R-factor

Standard reliability metric – measure of agreement between model and experimental data





CheckCIF results





Molecular complexity



22



Data integrity and fraud checks

- 1 year CCDC funded post-doc has started
- Outputs/aims include:
 - Identification of potentially fraudulent CIF data at deposition and in archive
 - Investigation into data integrity in the CSD
 - Additional checks, flags and options added to Deposition, Referee Service and available through CSD-System



Should we have more checks on all deposited data?









WebCSD – more advanced searching

The Cambridge Crystallographic Data Centre	Z Karlsruhe Li issistete for Information Infrastructure	WebCSD	The Cambridge Crystallographic Data Centre	rch Unit Cell Search	VebCSD		Sign In Licensed to: CCDC Main Site
Simple Search Structur	re Search Unit Cell Search		Unit Cell Searching				
Welcome to WebCSD. This se Search tab. Please use one or information and search help	ening ervice now includes the ability to search for in r more of the boxes to find entries. If you ente	organic structures through the CCDC's and FiZ Karlsruhe's r details in more than one field the search will try to find rec	j Lattice centring	Primitive (P)	0		
Identifier(s) Compound name	CCDC Number(s), CSD Number(s), CSD I	Refcode(s) or ICSD Number(s)	a e	.g. 10.0	φ α e.g. 90 φ β e.g. 90	.0	0
DOI Authors	A single publication DOI, CSD DOI or ICSI	סס סס	The Cambridge Cysts Data Center Simple Se	lographic arch Structure Search Unit Cell Search	WebCSD		Sign In Licensed to: CCDC Main Site
Journal Publication details	e.g. Journal of the American Chemical Soc Year	Volume O	Page	tructure searching			
The Cambridge Cryst Data Centre Simple Se	allographic s	it Cell Search	CSD	Licensed	& Web CSD → d to: CCDC Main Site		Help Keyboard shortcuts - Copy: Ctrl-C Delete: Ctrl-X Paste: Ctrl-X - Undo: Ctrl-Z - Redo: Ctrl-Y
Formula S Enter the mo default to 1.	Searching	earch for in the box below. Elements should	be followed by a number greate	r than 0 and any elements not follo	wed by a number will		Select all: CtrI-A Query features Query features describe how an atom or bond should behave in substructure searches. To add a feature: Right Click on atom or bond Z. Hover over 'query features'
Allow oth	Molecular Formula	C12 H15 N3 O3 S1			Θ	dotmatics	(atom only) 3. Hover over a feature type (e.g. H-count, type) 4. Select one of the options
		Search			Clear		More Information



WebCSD - More information visible to users

EKOSEE : (µ2-2,8-bis(2,2'-Bi nickel(ii) acetone diethyl ether Space Group: P 2 ₁ /c (14), Ce	pyridin-6-yl)-5-phenylpyrido[2,3-b][1,8 r solvate ell: a 20.1055(6)Å b 18.0229(5)Å c 15]naphthyridine)-(µ2-trifluoroacetato)-aqua-tris(trifluoroacetato)-di 5.0825(5)Å, α 90° β 90.727(3)° γ 90°	-	
3D viewer		Chemical diagram		
×.			Associated publications	
			Da-wei Huan 10.1039/C6D	ig, YI-Hung Liu, Shie-Ming Peng, Shiuh-Tzung Liu, <i>Dalton Transactions</i> , 2016, 45, 8265, DOI:)T00567E
The second se		Et—OEt	Chemical details	
	JSmol	н,с-соме	Formula	C ₄₅ H ₂₅ F ₁₂ N ₇ Ni ₂ O ₉ ,C ₄ H ₁₀ O,C ₃ H ₆ O
H Disorder 🔶 Menu Open -			Crystal details	
Style Labels	s Packing Measure	O View events werkelte here	Space group	P 2 ₁ /c (14)
		View group symbols key	Unit cell	a 20.1055(6)Å b 18.0229(5)Å c 15.0825(5)Å α 90° β 90.727(3)° γ 90°
Additional details			Cell volume	5464.85
Deposition Number	1450222		Reduced cell	a 15.083Å b 18.023Å c 20.106Å α 90.000° β 90.727° γ 90.000°
Data Citation	Da-wei Huang, Yi-Hung Liu, Shie- Crystal Structure Determination, 2	Ming Peng, Shiuh-Tzung Liu CCDC 1450222: Experimental 016, DOI: 10.5517/ccdc.csd.cc1kp2cp	Z, Z'	4, 1
Deposited on	29/03/2016		Habit	plate
			Disorder	C33,C34,C36,C37 and C33',C34',C36',C37' disordered over two sites with occupancies 0.931:0.069
			Colour	yellow
			Experimental details	
			R-factor (%)	5.29

Temperature (K)

Density (CCDC)

150

1.56218



Joint deposition and access for inorganics too

a from

"All

ocation.

ise a

of the

become

ts, and

talography,

data for all

and FIZ

The Cambridge Crystallographic Data Centre		Search	'n	Q Register	€ Sign In	
Community -	Research & Consultancy -	Solutions -	News & Events 👻	Support & Resources	• The CO	CDC -
Home / News & Events / N	lews / Free, unified deposition and	access of crystal structure	e data			

Free. unified deposition and access of crvstal structure data

The Cambridge Crystallographic Data Centre (CCDC) and FIZ Karlsruhe – Leibniz Institute for Information Infrastructure (FIZ Karlsruhe) today announced the launch of their joint deposition and access services for crystallographic data across all chemistry. These services will enable researchers to share data through a single deposition portal and explore all chemical structures for free worldwide.

"With this jo crystallogra chemistry is Karlsruhe w expectation

The Chair o information Failing that simple inter organic, and and make re Chemistry L

Recent adv blurred, for

Over 180,000 entries from the Inorganic Crystal Structure Database (ICSD)

fuel additives. This, coopied with the desire non-researchers for more integrated databases, has been the divelopment of these joint services.

As a result, researchers and educators worldwide, working across all fields of chemistry, are able to explore over one million crystallographic structures through a joint Access Structures service enabling them to view and retrieve deposited datasets associated with structures in the Cambridge Structural Database (CSD) and the Inorganic Crystal Structure Database (ICSD).

Crystallographers can deposit organic, inorganic and metal-organic structures through a unified deposition service. This features a streamlined online portal for easy submission and integrates a variety of checks to alert researchers about the validity, integrity and originality of their data. Additional features include the rapid assignment of deposition numbers and the ability for depositors to choose to share their data immediately through an appropriate database. Alternatively, data destined for inclusion in a scientific article is automatically shared at the point of publication through workflows with most major publishers. Anyone looking for structures previously stored in the FIZ Karlsruhe depot can still find them using the published depot number.







Annual targeted database improvements

- **nprovements** dispersion python by the pytho
- Greater efficiency allows targeted improvements to further increase integrity, consistency, discoverability and usage of data
 - 50,000 existing entries improved in 2015
 - 74,000 existing entries improved in 2016
 - 62,000 existing entries improved in 2017
 - 80,000 existing entries improved in 2018

Study Temperature relative to Melting Point



Targeted improvements to metal-organics





Targeted improvements to drug subset

Ensure consistency of early CSD entries







Metal-Organic Frameworks

- Two CSD subsets identifying all metal-organic frameworks within the database
 - MOF list >80,000 entries intended to be as widely-applicable as possible
 - 2. Non-disordered list >65,000 entries for highthroughput calculations
 - Both subsets updated quarterly
- Python script available to download





The Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present and Future. P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward, D. Fairen-Jimenez, Chem. Mater., 2017, 29 2618-2625 DOI: 10.1021/acs.chemmater.7b00441



Accessing data from general resources



https://pubchem.ncbi.nlm.nih.gov/compound/1983#section=Crystal-Structures



Worldwide online access









Summary

- The CSD is a worldwide community resource
- It is important that we develop services that help you
- What are your priorities for deposition and access?

- Please let us know
 - <u>lightfoot@ccdc.cam.ac.uk</u>
 - <u>support@ccdc.cam.ac.uk</u>
- Thank you for your attention