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

Collaborative Research Organisation
New methodologies
Fundamental research

Education and Outreach
Conferences, Workshops,
Bespoke Training, Teaching Materials

Employer of ~60 staff

Scientific editors
Software developers
Applications scientists

Cambridge UK

USA

Originated in 1965

Financially self-supporting

Not-for-profit, UK Registered Charity

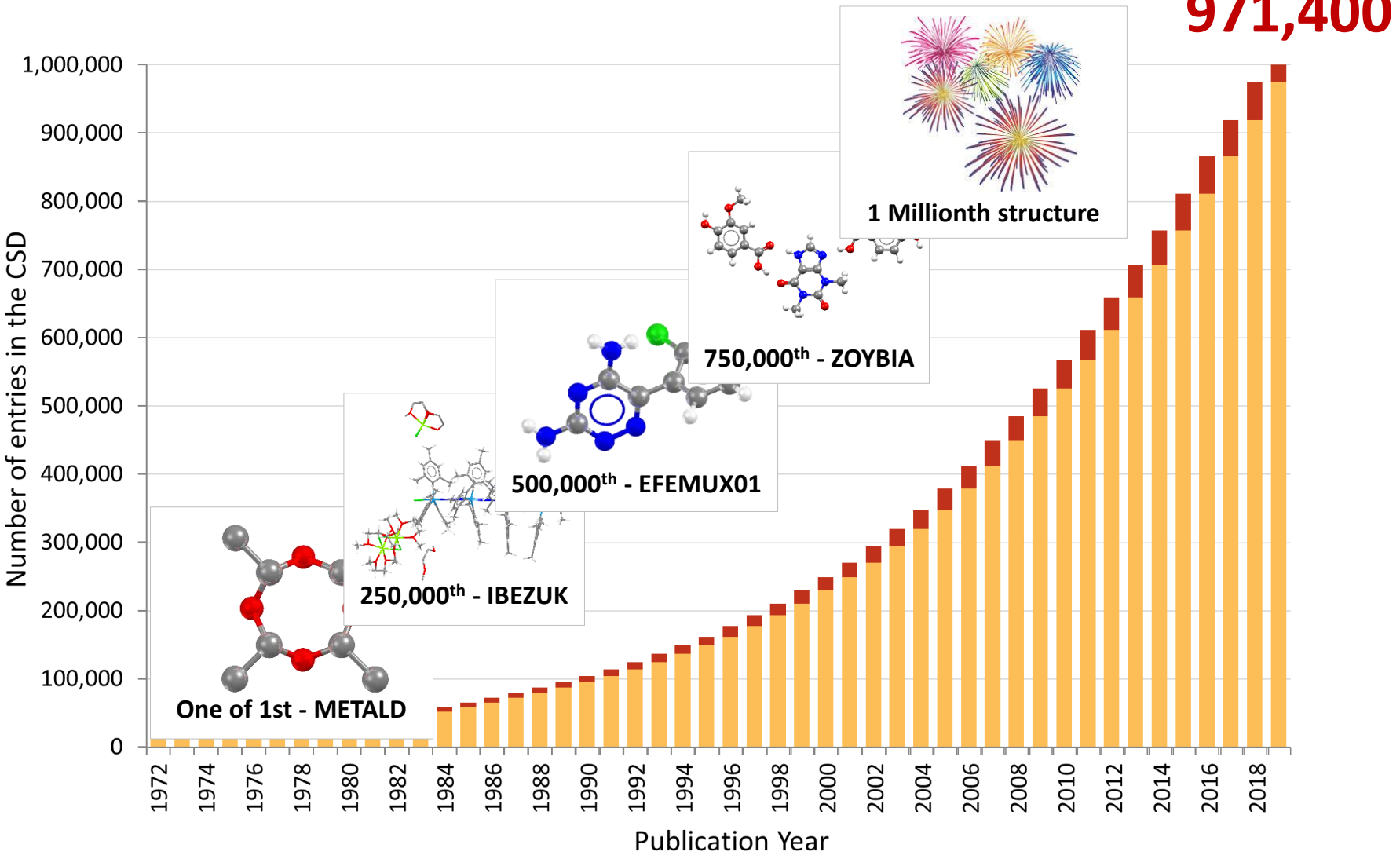
University Partner Institute

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

971,400



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%

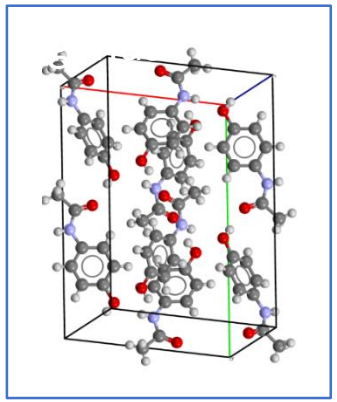
Polymeric: 11%

Single Component
56%

Multi Component
44%

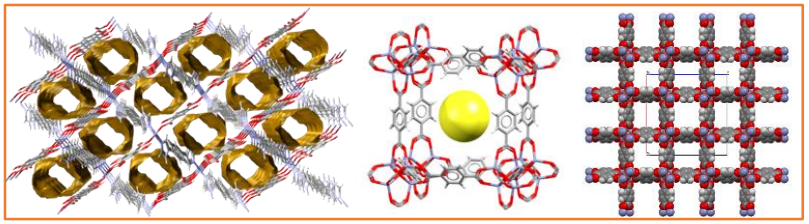
Organic

- Drugs
- Agrochemicals
- Pigments
- Explosives
- Protein ligands

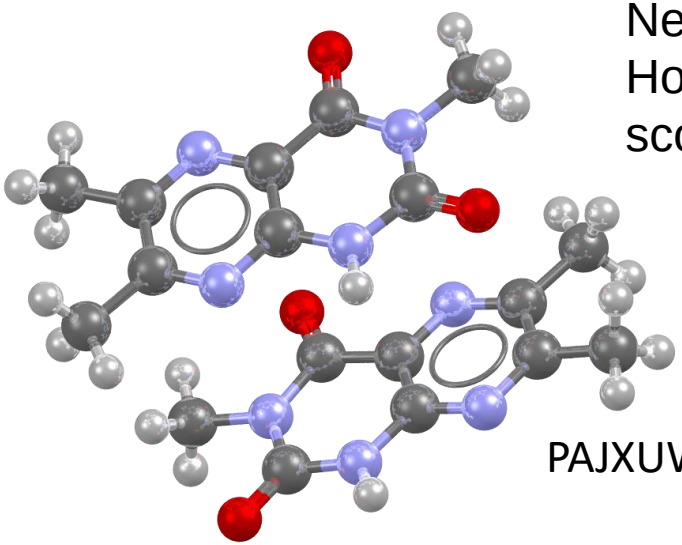


Metal-Organic

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

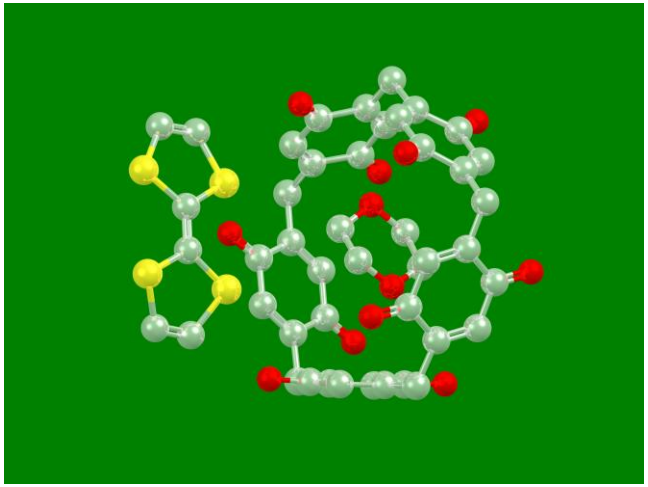


New Zealand



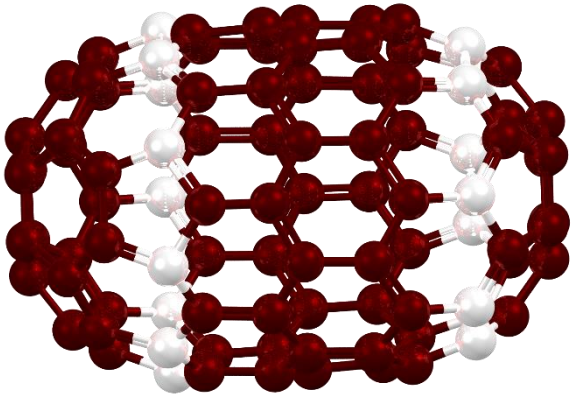
New Zealand Manuka Honey (Leptospermum scoparium)

PAJXUW



WAZHAJ

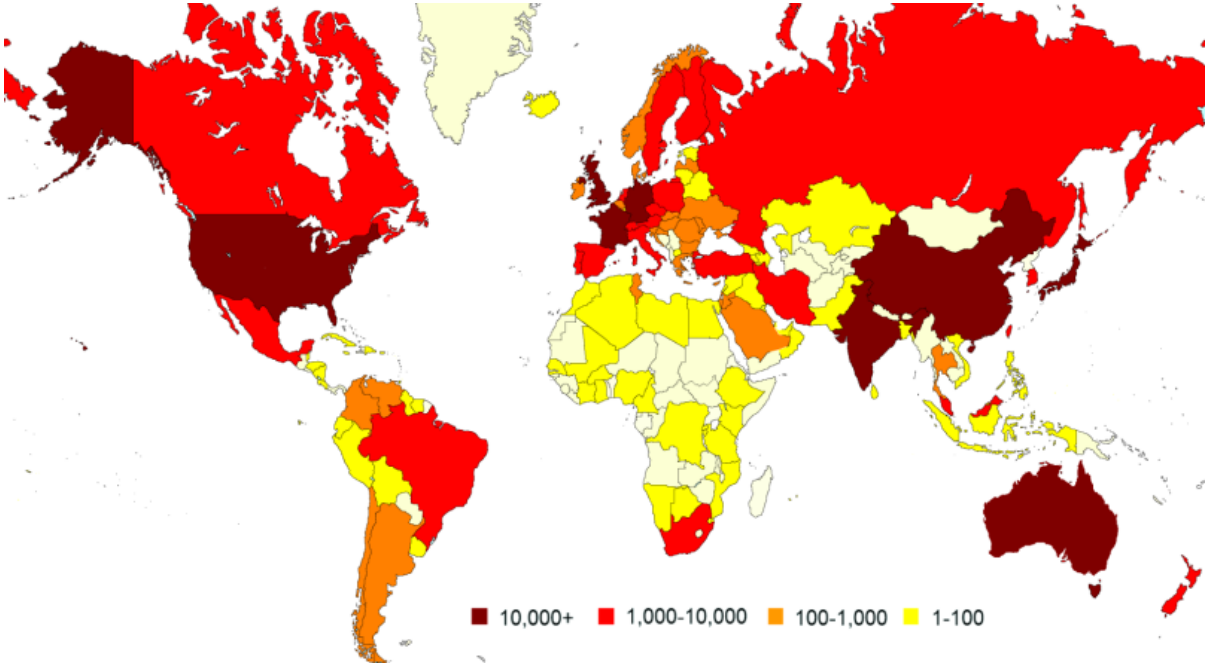
SAVQOX



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

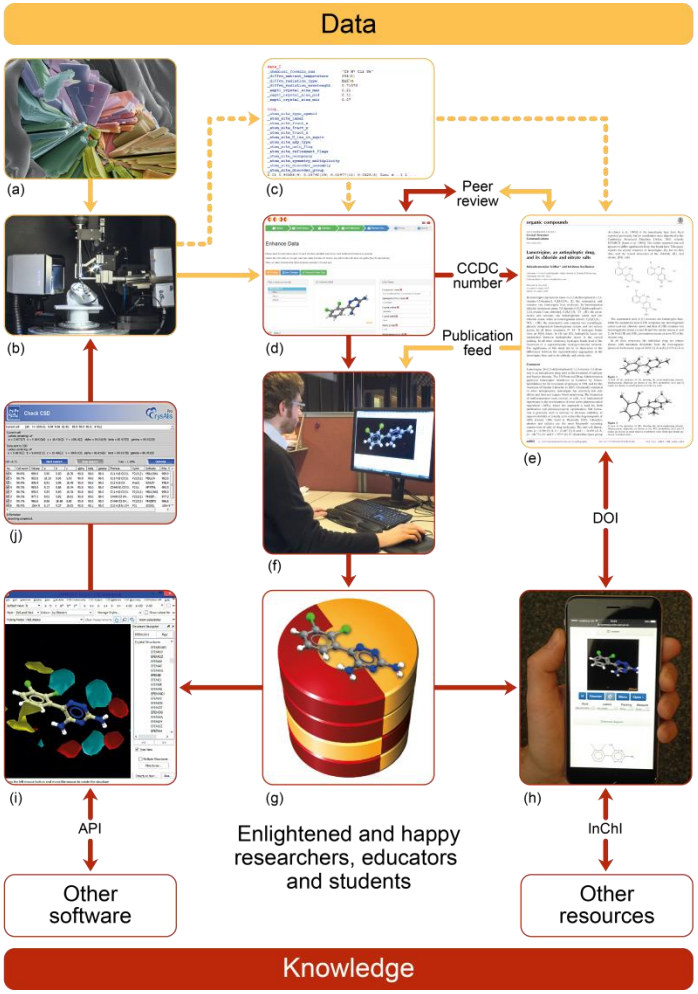
Deposition guidelines

- 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 Cambridge Crystallographic Data Centre (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 ICSD, 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 Cambridge Crystallographic Data Centre

Search | Register | Sign In

Community | Research & Consultancy | Solutions | News & Events | Support & Resources | The CCDC

Home / Community / Deposit a Structure / The CCDC CIF Deposition Guidelines

The CCDC CIF Deposition Guidelines

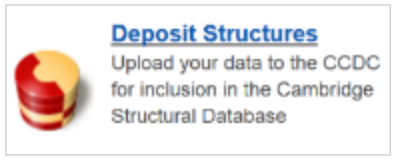
When preparing your CIF for deposition please include as much information as possible and check it carefully. This is especially true for *CSD Communications* where there is no paper to describe the chemistry and experimental details leading to your structure. If you choose to publish your data as a *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 CSD.

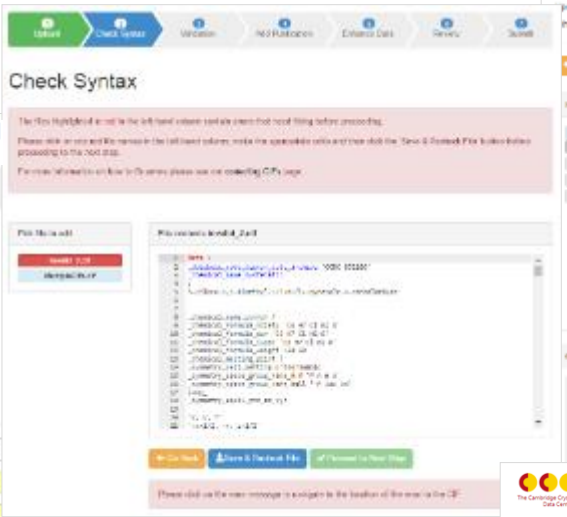
All experimental CIF files (including those from powder diffraction experiments) should contain an R-factor. This should be consistent with the crystallography being performed correctly and to the best ability that would be expected from the material and equipment used. We would like all experimental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

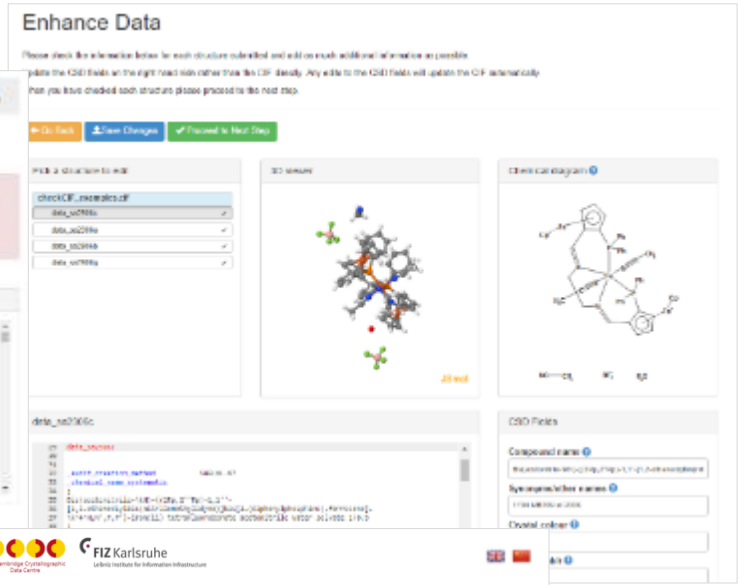
- R-factors (R1, wR2, Rint)

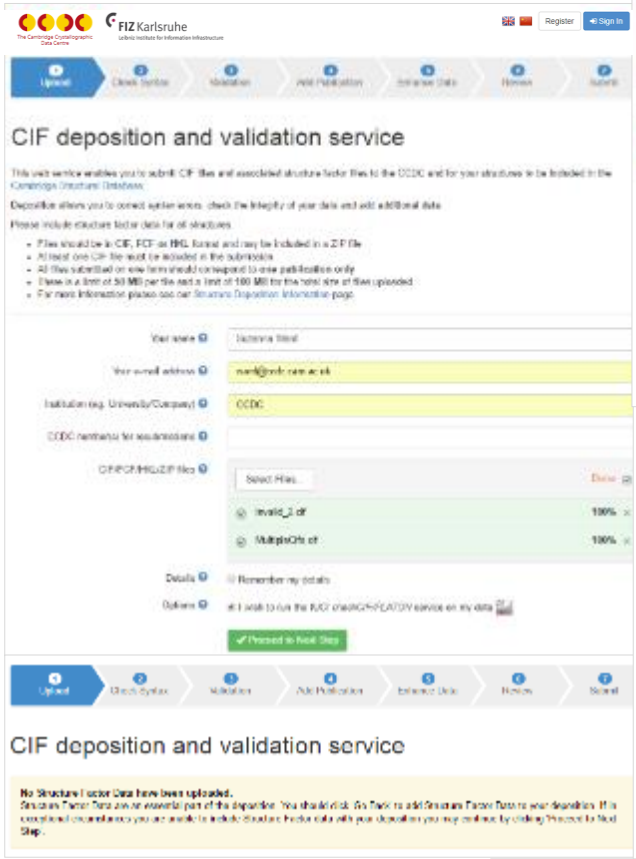
https://www.ccdc.cam.ac.uk/Community/depositastructure/cif-deposition-guidelines/

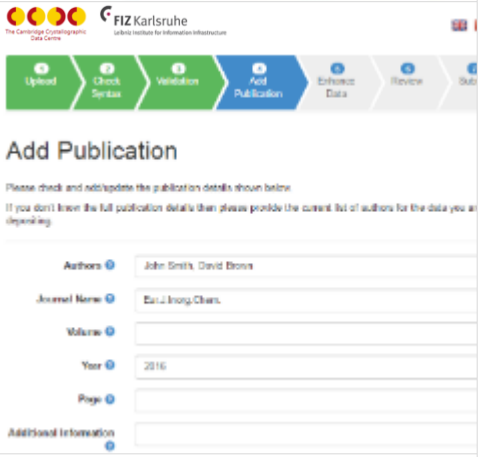
Data Deposition

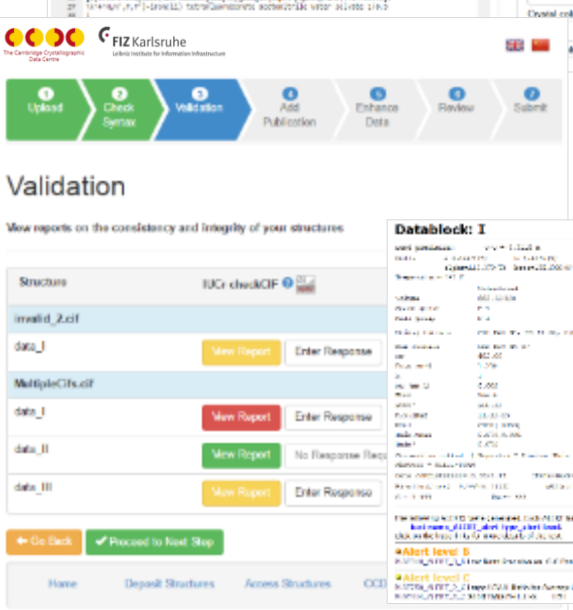






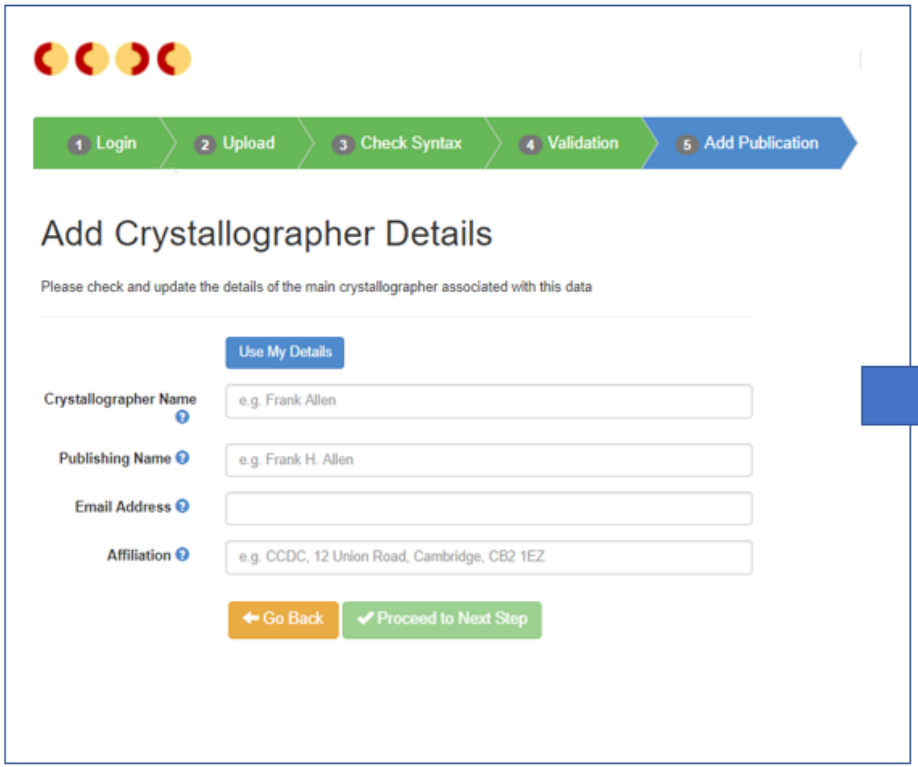


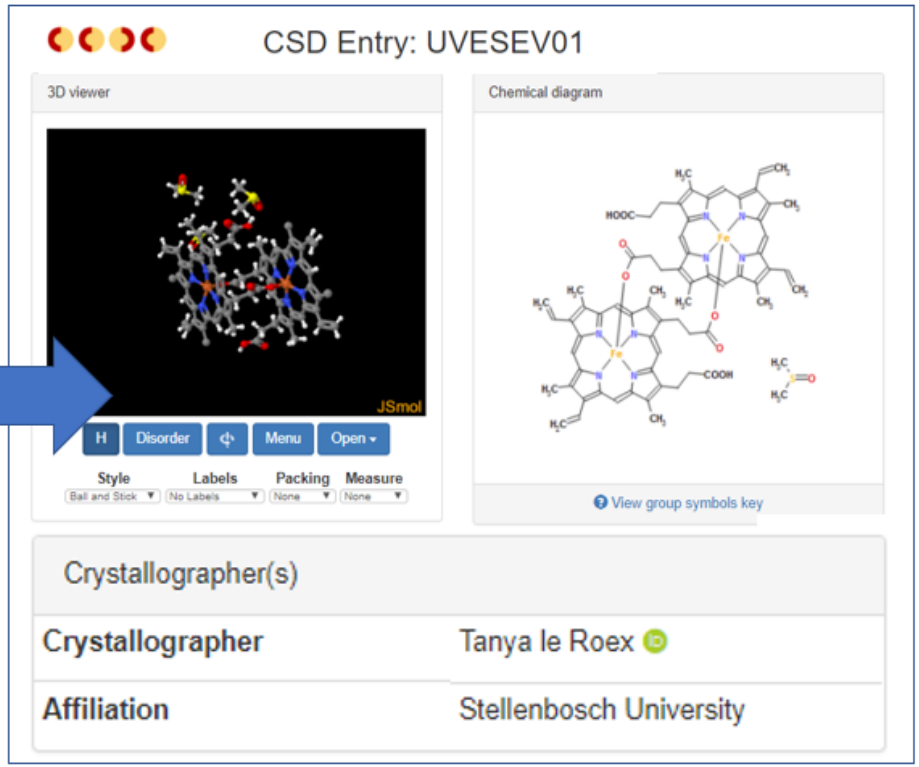




Retrieve Deposited Files

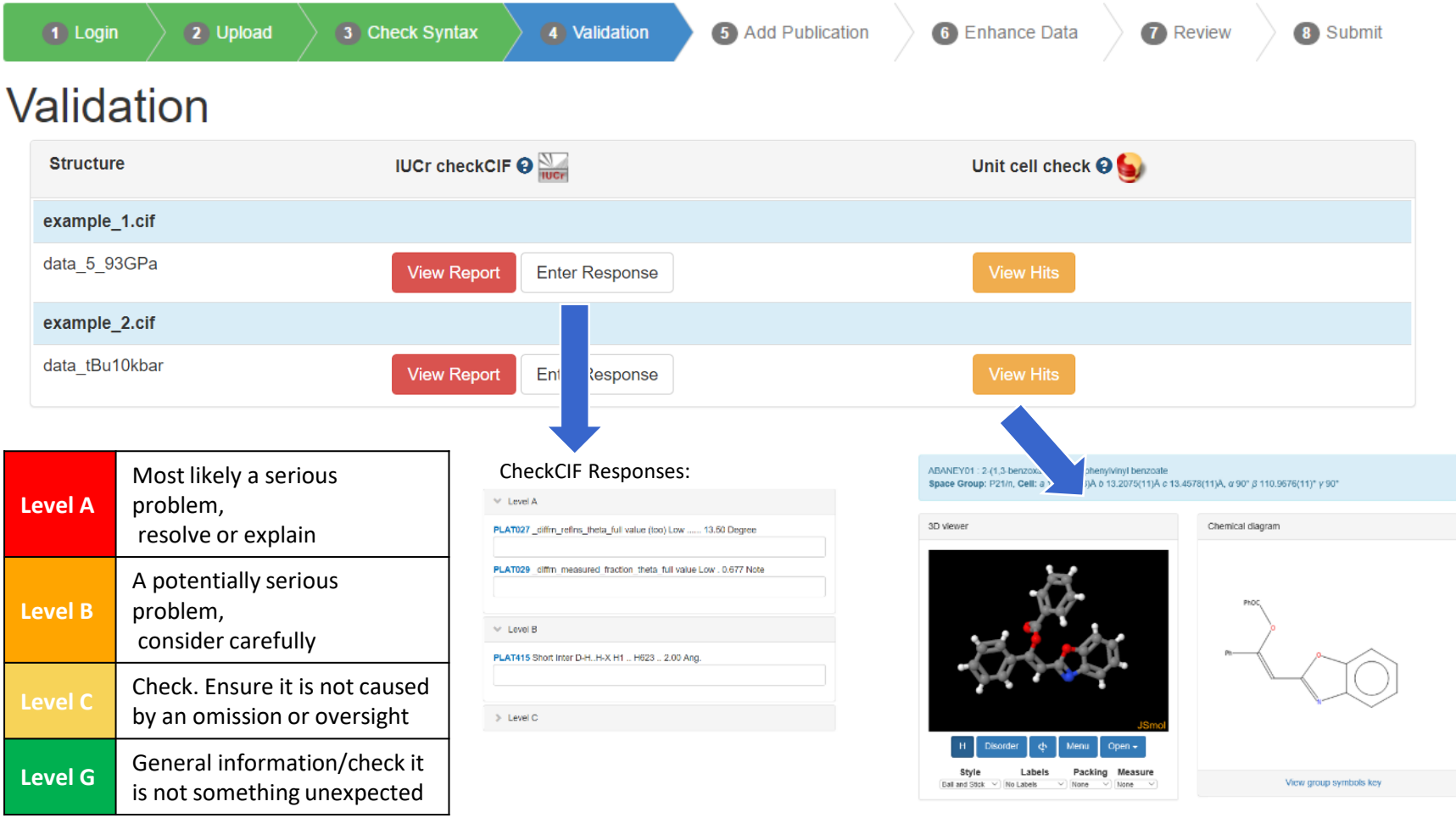
Ability to identify the data producer





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

Integrity and validation checks



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

Data enrichment: By the depositor

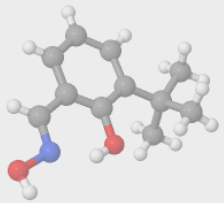
1 Login | 2 Upload | 3 Check Syntax | 4 Validation | 5 Add Publication | 6 Enhance Data | 7 Review | 8 Submit

Pick a structure to edit

example_2.cif
data_tBu10kbar ✓

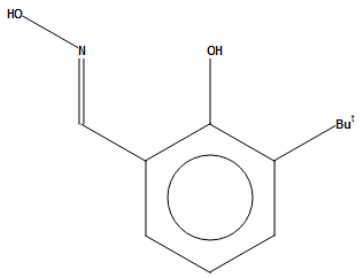
example_1.cif
data_5_93GPa ✓

3D viewer



JSmol

Chemical diagram



data_tBu10kbar

```
data_tBu10kbar


_audit_creation_date              06-12-18
_audit_creation_method            CRYSTALS_ver_12.84

_chemical_name_systematic         3-tButylsalicylaldoxime
_chemical_melting_point           ?
_cell_length_a                    14.811(3)
_cell_length_b                    6.4564(7)
_cell_length_c                    19.759(4)
_cell_angle_alpha                 90
_cell_angle_beta                  94.060(8)
_cell_angle_gamma                 90
_cell_volume                      1884.7(5)

_symmetry_cell_setting            Monoclinic
_symmetry_space_group_name_H-M    'I 2/a'
_symmetry_space_group_name_Hall   '-I 2ya'
loop_
_symmetry_equiv_pos_as_xyz
x,y,z
-x,-y,-z
```

Associated DOIs

Raw data DOI

Data fields

Compound name
3-tButylsalicylaldoxime

Synonyms/other names
3-tButylsalicylaldoxime

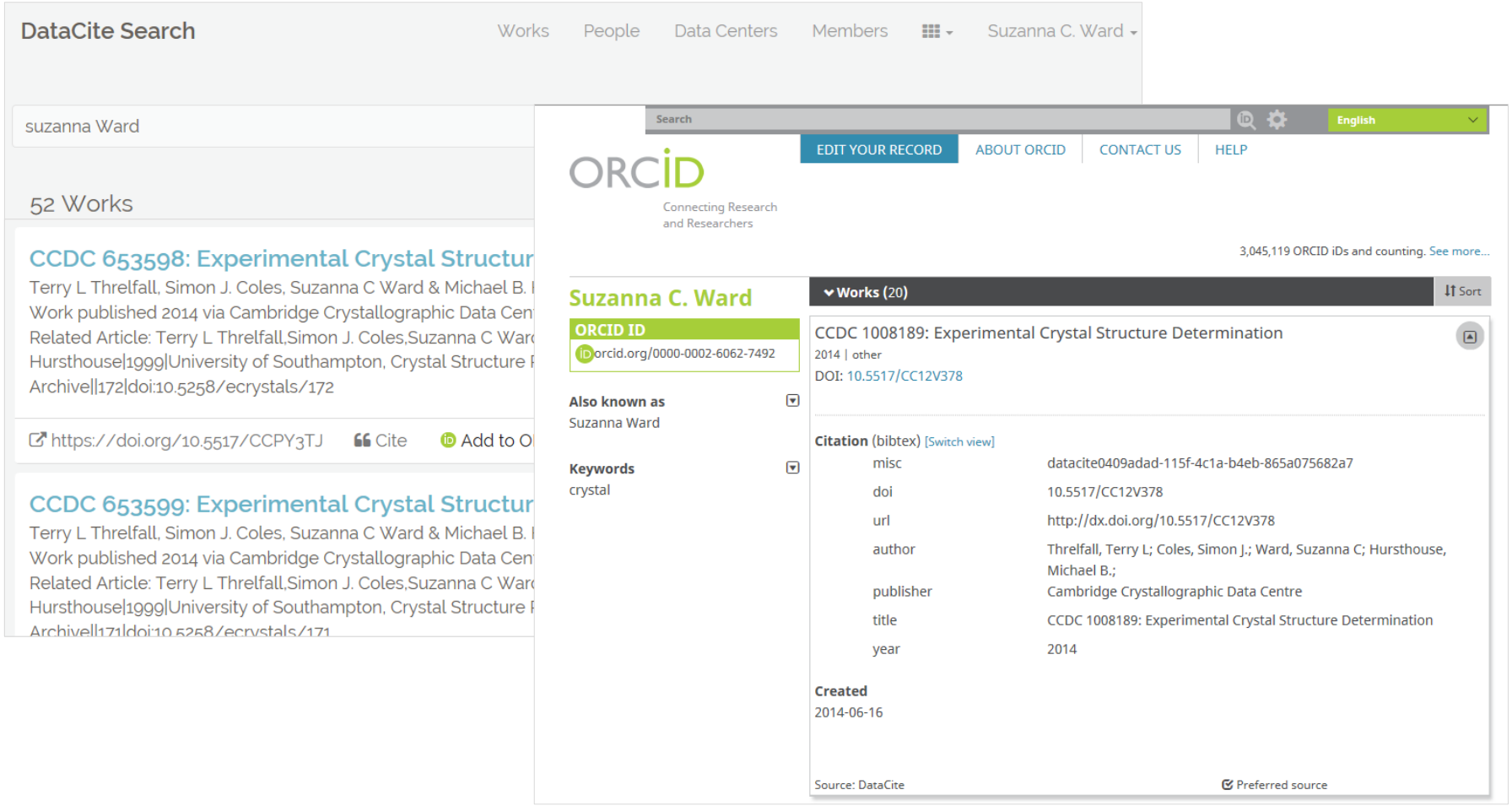
Crystal colour
colourless

Chemical interpretation of 3D data generated by CCDC software. Downloadable after deposition

Opportunity to link to raw data

Opportunity for crystallographer to provide additional domain-specific data items

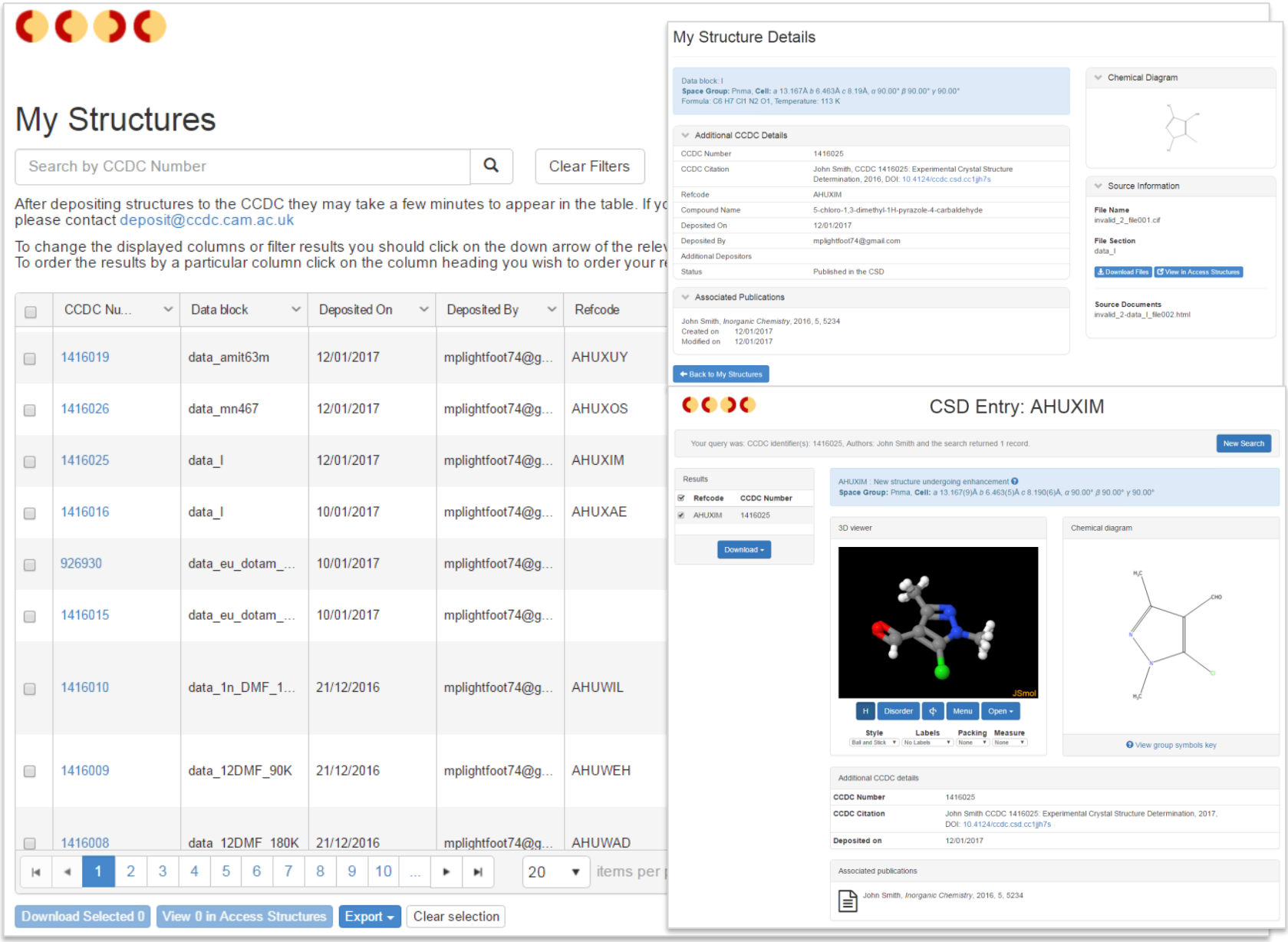
Adding data to your Researcher IDs



Deposition portal

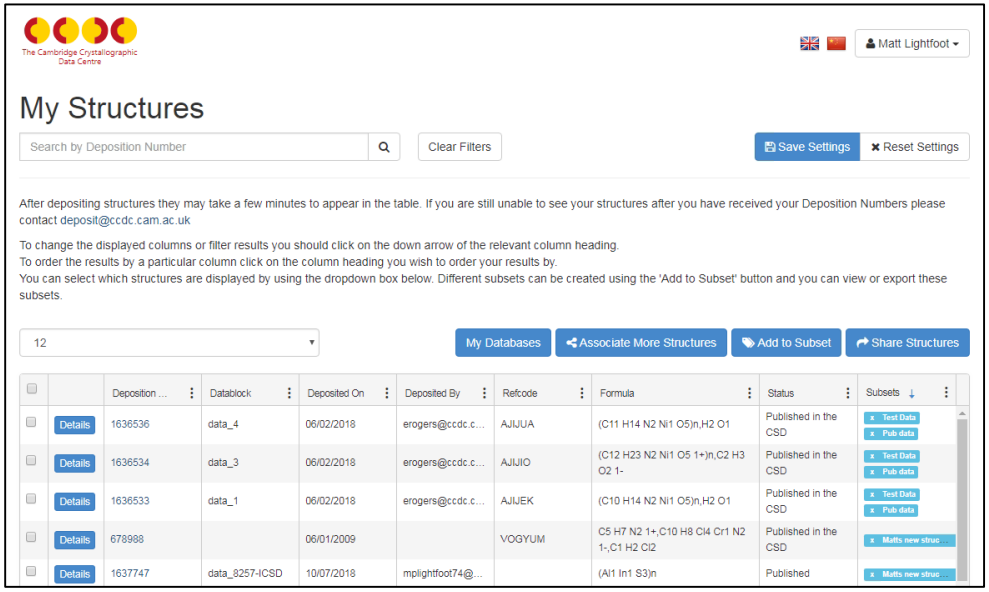
- **Launched December 1st 2016**
  - create a CCDC profile, log on and view and retrieve your depositions
  - deposit new data using your account
- April 2017
  - Edit and update basic information
  - Assign data as *CSD Communications*
  - Extend embargo periods
- July 2017
  - Share data with co-workers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

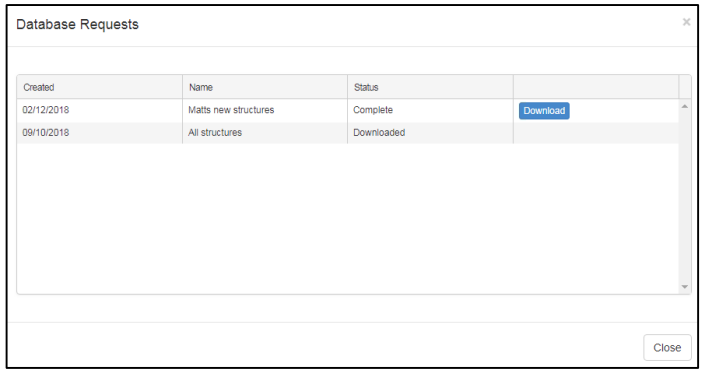
- Currently 29,000 users


My Structures
Search by CCDC Number
Clear Filters
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
items per
Download Selected 0
View 0 in Access Structures
Export
Clear selection
My Structure Details
Chemical Diagram
Additional CCDC Details
CCDC Number 1416025
Refcode AHUXIM
Compound Name 5-chloro-1,3-dimethyl-1H-pyrazole-4-carbaldehyde
Deposited On 12/01/2017
Deposited By mplightfoot74@gmail.com
Status Published in the CSD
Source Information
File Name
File Section
Associated Publications
Source Documents
Back to My Structures
CSD Entry: AHUXIM
New Search
Results
Refcode CCDC Number
AHUXIM 1416025
Download
3D viewer
Chemical diagram
Disorder
Menu
Open
Style
Labels
Packing
Measure
View group symbols key
Additional CCDC details
CCDC Number 1416025
Deposited on 12/01/2017
Associated publications

Coming soon in Deposition/My Structures

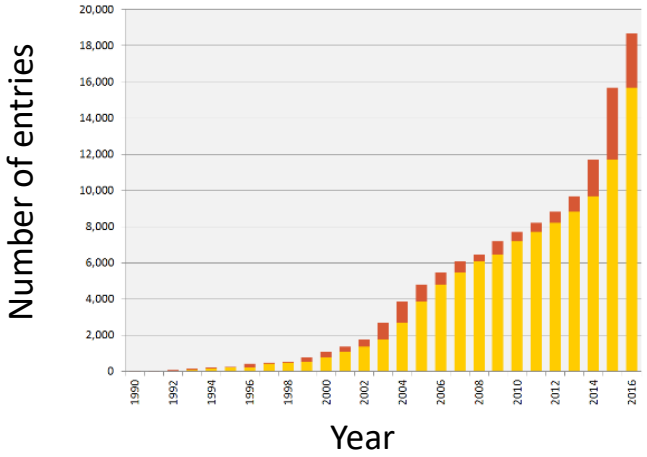
- Ability to create a subset of your structures
- Ability to share a subset of your structures with your colleagues
- Ability to create a CSD format database of your in-house structures to use in CSDS software.

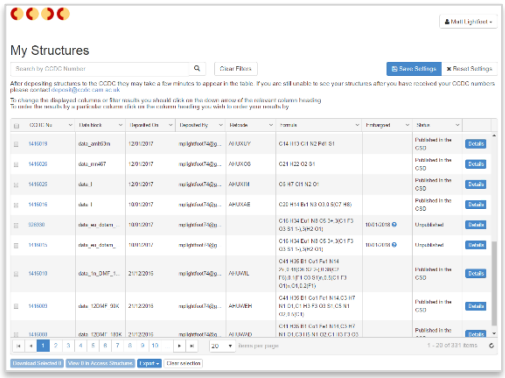


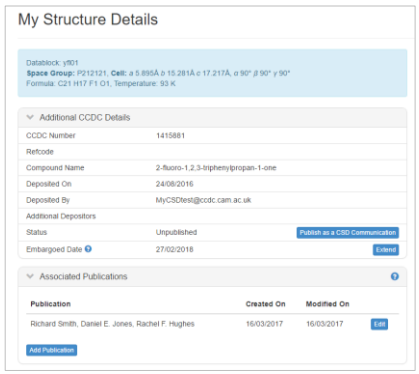


CSD Communications

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







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

It's only data that isn't good enough for publication?

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

Only duplicates are shared through a database?

It isn't the compound I wanted.

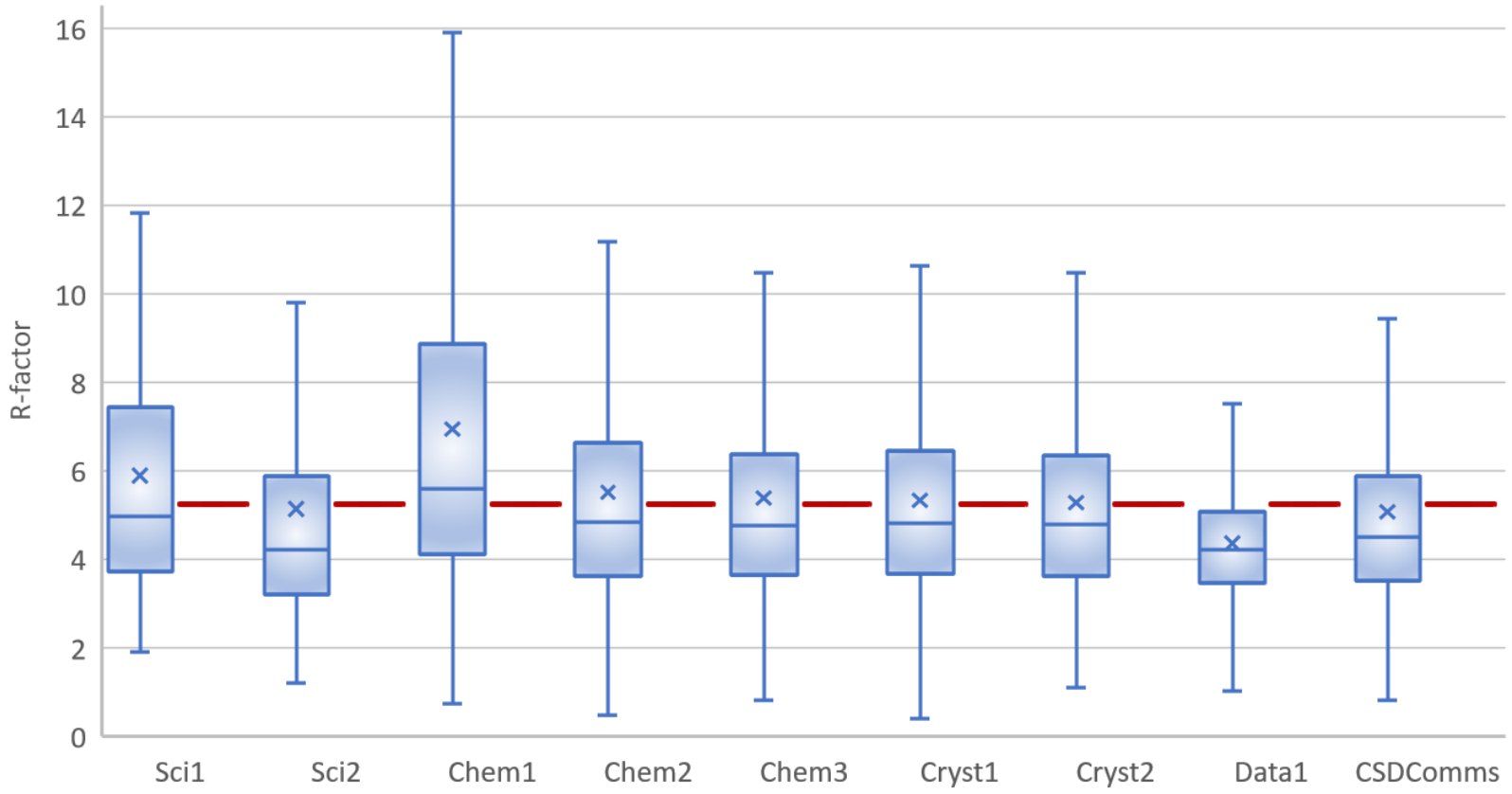
Selection of data sources

Web of Science

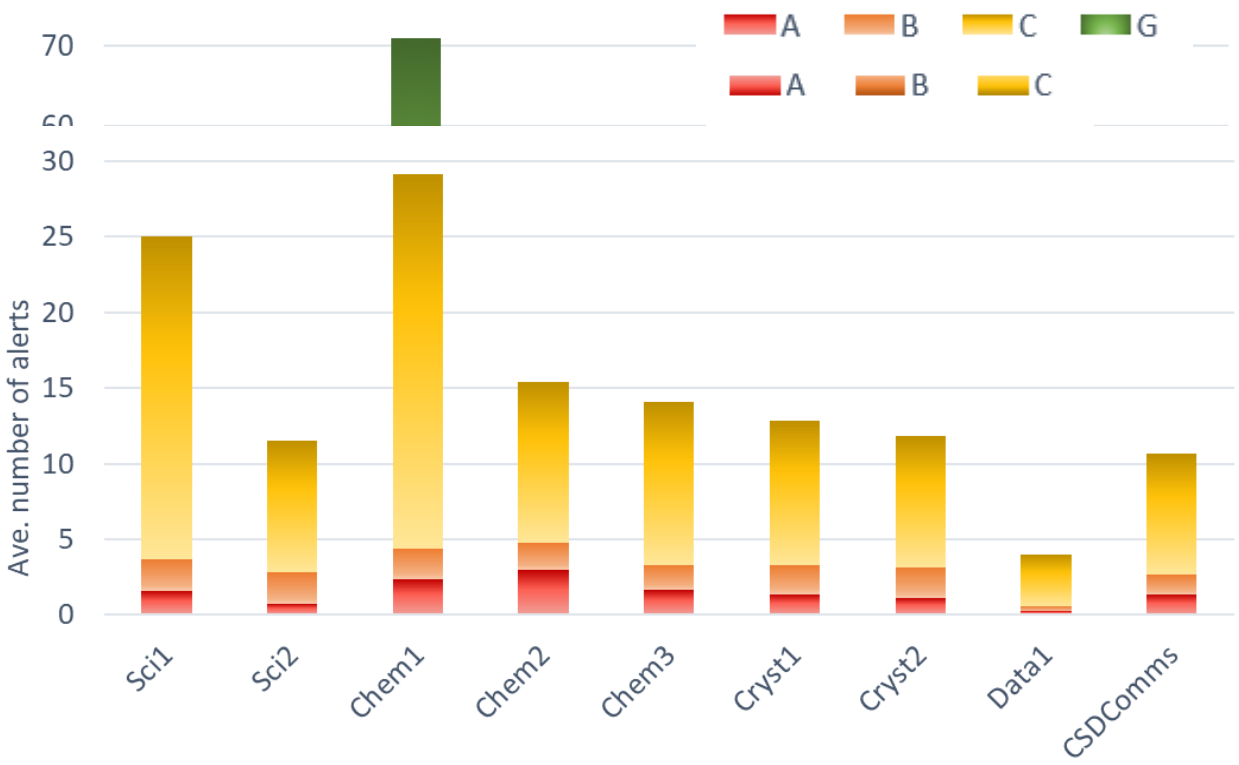
- From Thomson Reuters Web of Knowledge
- High impact factor journals in ISI categories
  - All Journals Sci1, Sci2
  - Chemistry - Multidisciplinary Chem1, Chem2, Chem3
  - Crystallography Cryst1, Cryst2
  - Data based journal Data1
  - Data published through CSD communications CSDComms
  - Average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

R-factor

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

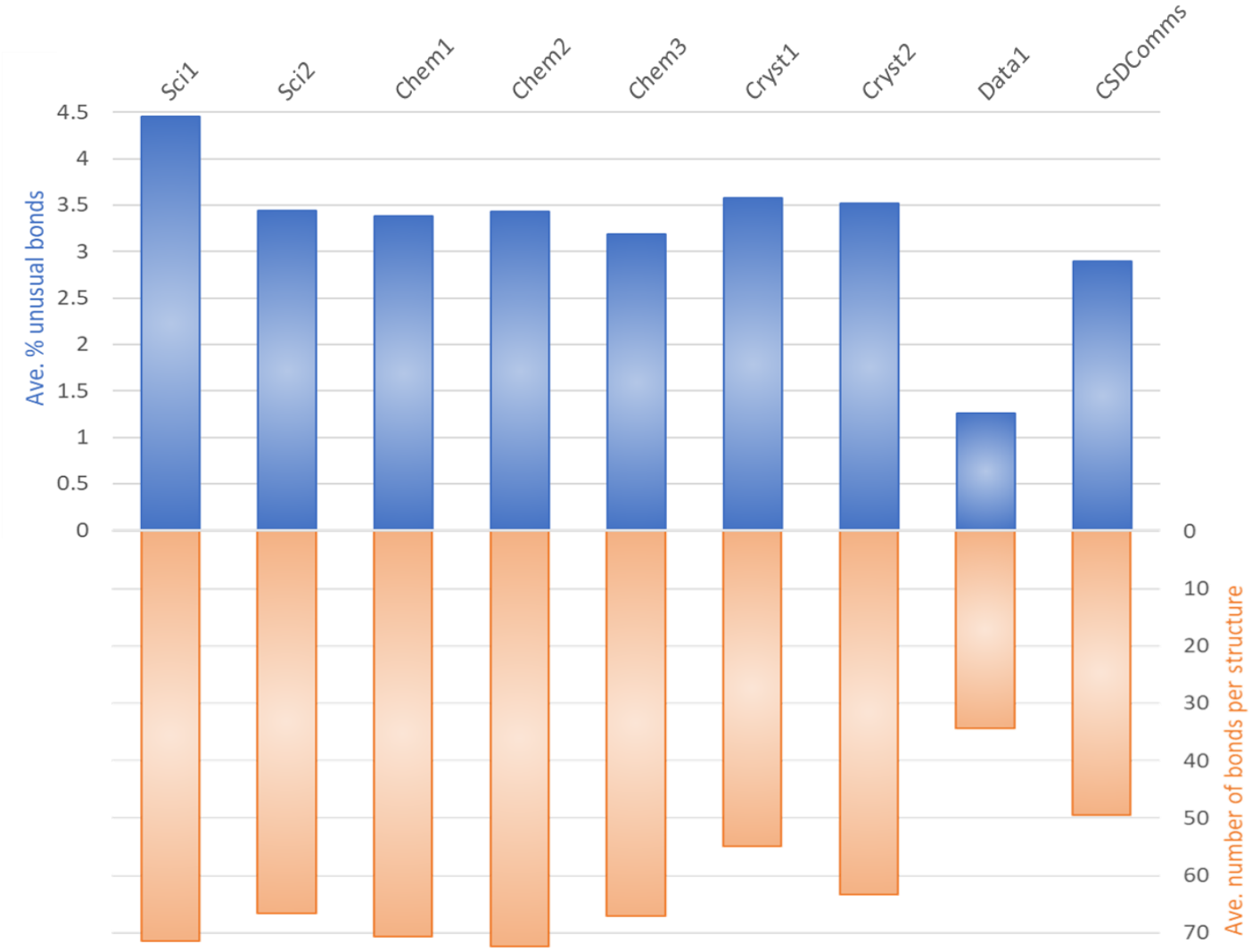


CheckCIF results



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

Molecular complexity

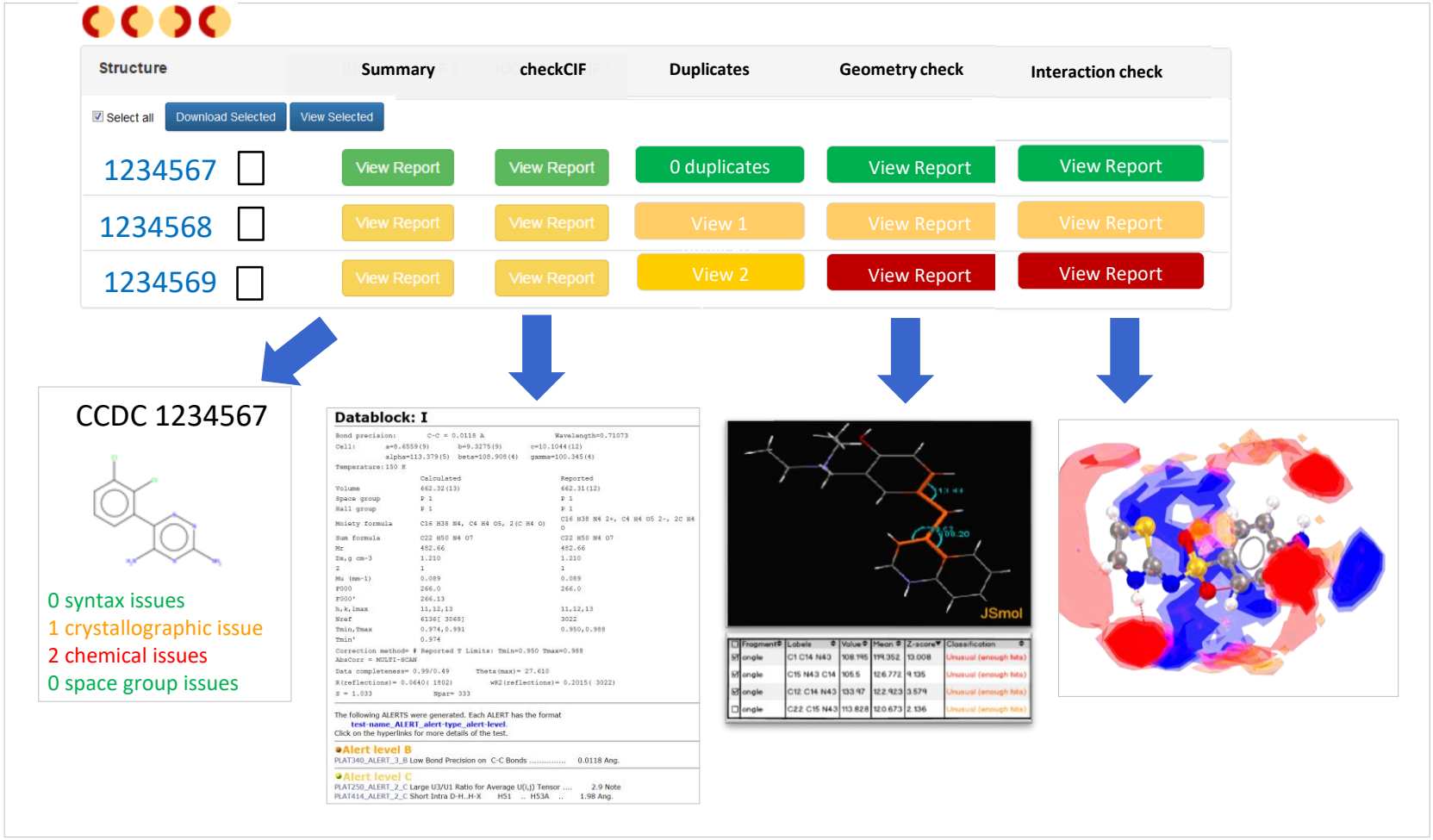


Data integrity and fraud checks

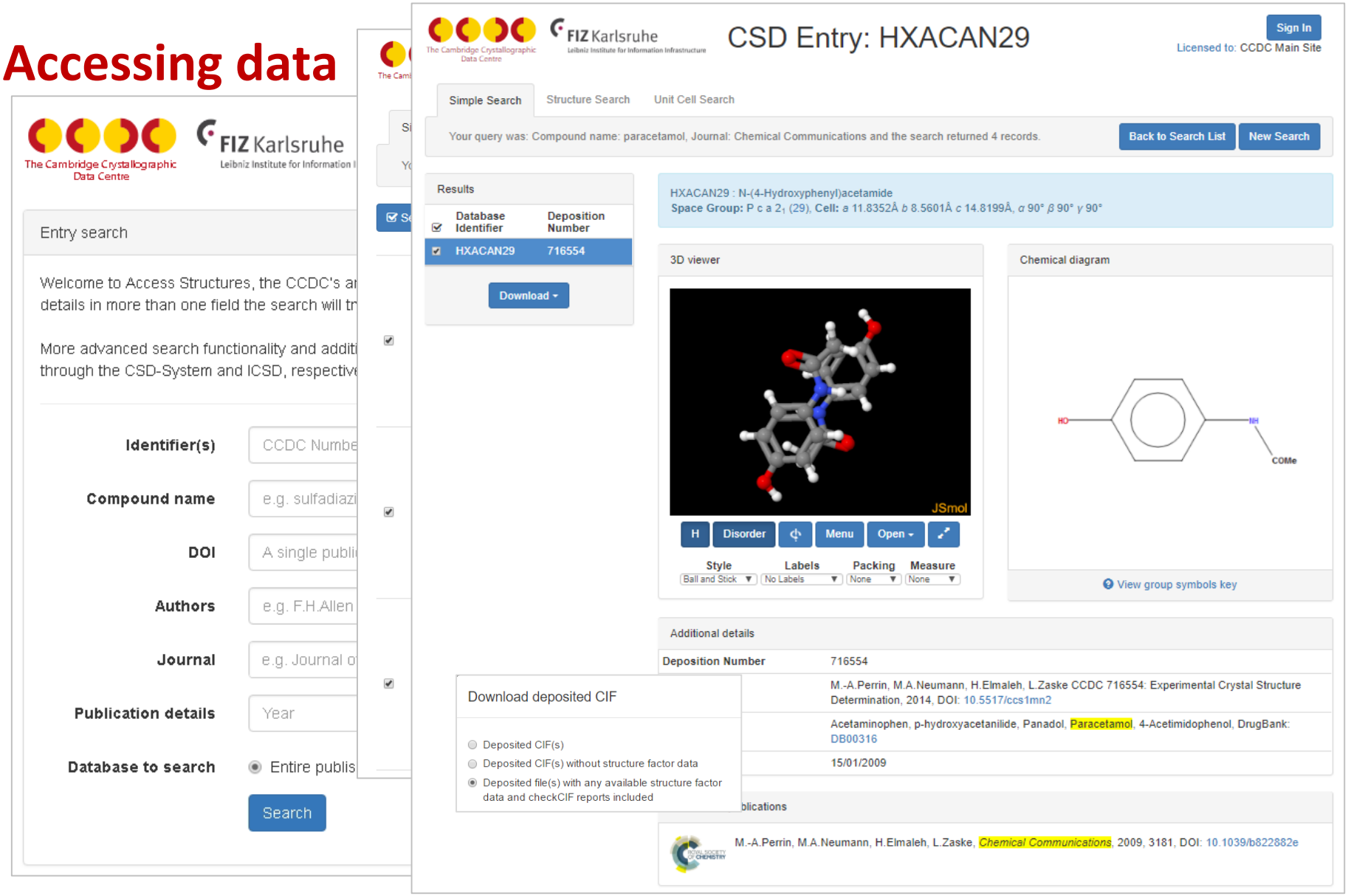
- 1 year CCDC funded post-doc has started

- 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

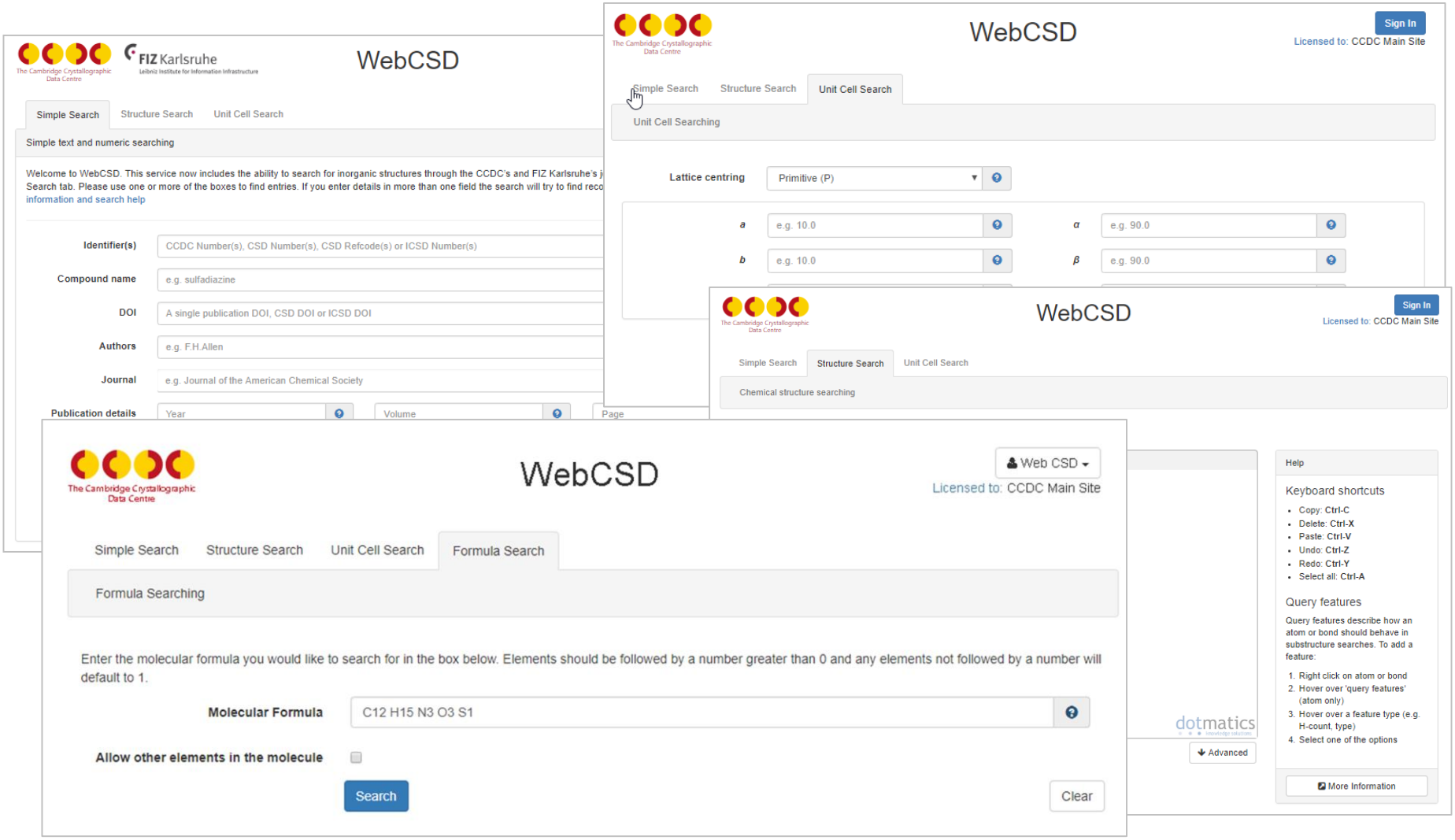
Should we have more checks on all deposited data?



Accessing data



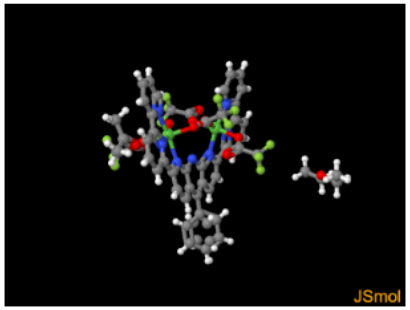
WebCSD – more advanced searching



WebCSD - More information visible to users

EKOSEE : (μ2-2,8-bis(2,2'-Bipyridin-6-yl)-5-phenylpyrido[2,3-b][1,8]naphthyridine)-(μ2-trifluoroacetato)-aqua-tris(trifluoroacetato)-di-nickel(ii) acetone diethyl ether solvate
Space Group: P $2_1/c$ (14), **Cell:** *a* 20.1055(6)Å *b* 18.0229(5)Å *c* 15.0825(5)Å, *α* 90° *β* 90.727(3)° *γ* 90°

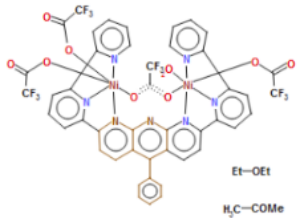
3D viewer



H | Disorder | Menu | Open

Style: Ball and Stick | Labels: No Labels | Packing: None | Measure: None

Chemical diagram



View group symbols key

Additional details

Deposition Number	1450222
Data Citation	Da-wei Huang, Yi-Hung Liu, Shie-Ming Peng, Shiuh-Tzung Liu CCDC 1450222: Experimental Crystal Structure Determination, 2016, DOI: 10.5517/ccdc.csd.cc1kp2cp
Deposited on	29/03/2016

Associated publications

Da-wei Huang, Yi-Hung Liu, Shie-Ming Peng, Shiuh-Tzung Liu, *Dalton Transactions*, 2016, 45, 8265, DOI: 10.1039/C6DT00567E

Chemical details

Formula	$C_{45}H_{25}F_{12}N_7Ni_2O_9.C_4H_{10}O.C_3H_6O$

Crystal details

Space group	P $2_1/c$ (14)
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°
Cell volume	5464.85
Reduced cell	*a* 15.083Å *b* 18.023Å *c* 20.106Å *α* 90.000° *β* 90.727° *γ* 90.000°
Z, Z'	4, 1
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)	150
Density (CCDC)	1.56218

Joint deposition and access for inorganics too

Search | Register | Sign In

Community | Research & Consultancy | Solutions | News & Events | Support & Resources | The CCDC

Home / News & Events / News / Free, unified deposition and access of crystal structure data

Free, unified deposition and access of crystal 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… a from crystallogra… data for all chemistry is… and FIZ Karlsruhe w… h expectation…

The Chair o… : "All information… ocation. Failing that… use a simple inter… tal- organic, an… ography, and make r… of the Chemistry L…

Recent adv… become blurred, for… ts, and fuel additives. This, coupled with the desire from researchers for more integrated databases, has been the driving force behind the development 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.

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

FIZ Karlsruhe

ICSD Entry: 60767

Sign In

Your query was: Identifier(s): 60767 and the search returned 1 record.

New Search

Results

Database Identifier	Deposition Number
ICSD 60767	1553062

Download

ICSD 60767 : ICSD Structure : $(O_5\ V_2)_n$

Space Group: P m m n (59), **Cell:** *a* 11.512Å *b* 3.564Å *c* 4.368Å, *α* 90.0° *β* 90.0° *γ* 90.0°

3D viewer

JSmol

H | Disorder | Menu | Open

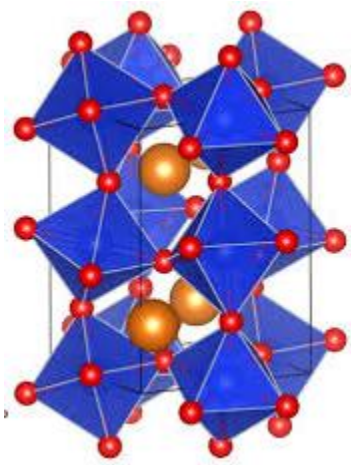
Style: Ball and Stick | Labels: No Labels | Packing: None | Measure: None

Chemical diagram

More Information Available from the ICSD

http://www2.fiz-karlsruhe.de/icsd_home.html

View group symbols key

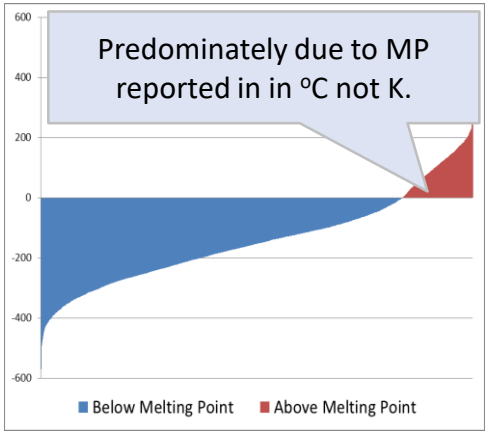


Annual targeted database improvements

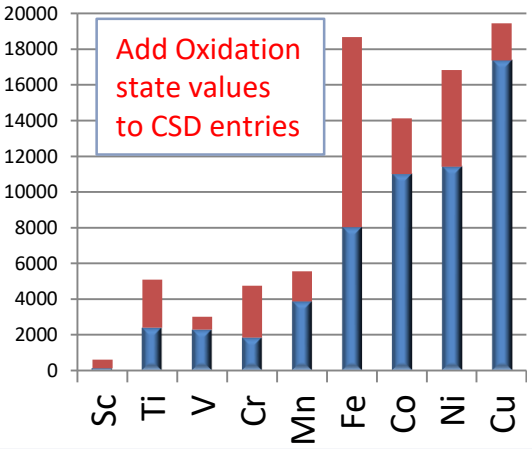
- Greater efficiency allows targeted improvements to further increase integrity, consistency, discoverability and usage of data
  - 50,000 existing entries improved in 2015
  - 74,000 existing entries improved in 2016
  - 62,000 existing entries improved in 2017
  - 80,000 existing entries improved in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

Targeted improvements to drug subset

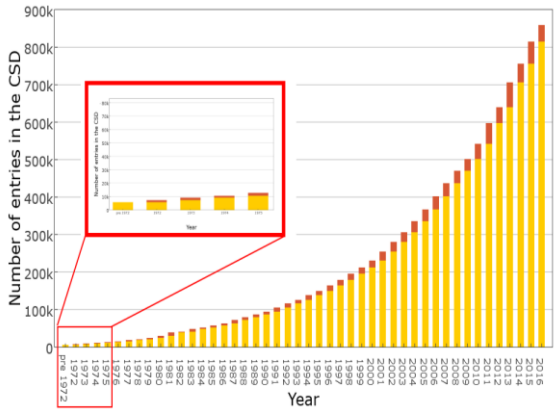
Study Temperature relative to Melting Point



Targeted improvements to metal-organics

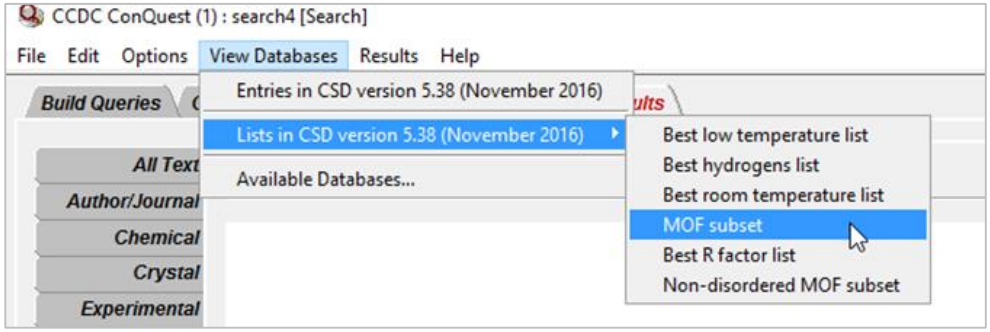


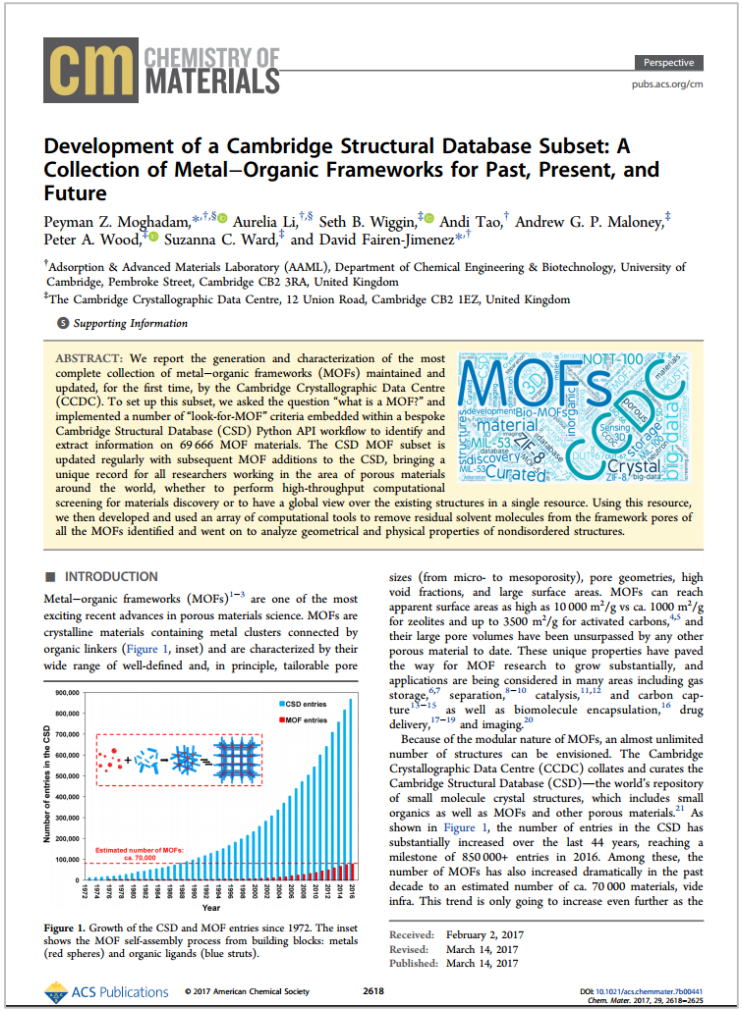
Ensure consistency of early CSD entries



Metal-Organic Frameworks

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





CHEMISTRY OF MATERIALS

Perspective

pubs.acs.org/cm

Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future

Peyman Z. Moghadam,*[†,§] Aurelia Li,[†,§] Seth B. Wiggin,[‡] Andi Tao,[†] Andrew G. P. Maloney,[‡] Peter A. Wood,[‡] Suzanna C. Ward,[‡] and David Fairen-Jimenez*[,†]

[†]Adsorption & Advanced Materials Laboratory (AAML), Department of Chemical Engineering & Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, United Kingdom

[‡]The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom

Supporting Information

ABSTRACT: We report the generation and characterization of the most complete collection of metal–organic frameworks (MOFs) maintained and updated, for the first time, by the Cambridge Crystallographic Data Centre (CCDC). To set up this subset, we asked the question "what is a MOF?" and implemented a number of "look-for-MOF" criteria embedded within a bespoke Cambridge Structural Database (CSD) Python API workflow to identify and extract information on 69 666 MOF materials. The CSD MOF subset is updated regularly with subsequent MOF additions to the CSD, bringing a unique record for all researchers working in the area of porous materials around the world, whether to perform high-throughput computational screening for materials discovery or to have a global view over the existing structures in a single resource. Using this resource, we then developed and used an array of computational tools to remove residual solvent molecules from the framework pores of all the MOFs identified and went on to analyze geometrical and physical properties of nondisordered structures.

■ INTRODUCTION

Metal–organic frameworks (MOFs)[1−3] are one of the most exciting recent advances in porous materials science. MOFs are crystalline materials containing metal clusters connected by organic linkers (Figure 1, inset) and are characterized by their wide range of well-defined and, in principle, tailorable pore sizes (from micro- to mesoporosity), pore geometries, high void fractions, and large surface areas. MOFs can reach apparent surface areas as high as 10 000 m^2/g vs ca. 1000 m^2/g for zeolites and up to 3500 m^2/g for activated carbons,[4,5] and their large pore volumes have been unsurpassed by any other porous material to date. These unique properties have paved the way for MOF research to grow substantially, and applications are being considered in many areas including gas storage,[6,7] separation,[8−10] catalysis,[11,12] and carbon capture[13−15] as well as biomolecule encapsulation,[16] drug delivery,[17−19] and imaging.[20]

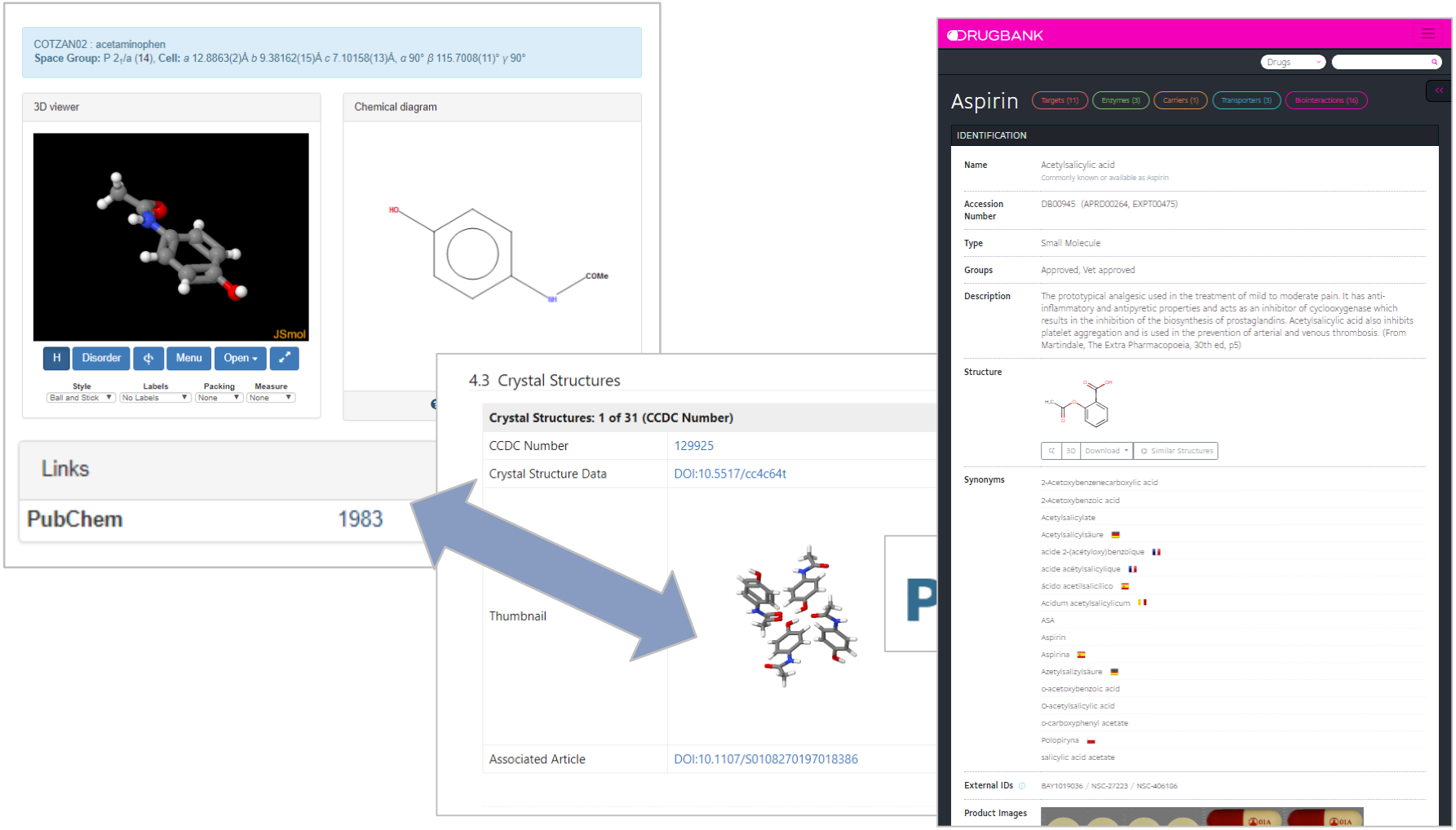
Because of the modular nature of MOFs, an almost unlimited number of structures can be envisioned. The Cambridge Crystallographic Data Centre (CCDC) collates and curates the Cambridge Structural Database (CSD)—the world's repository of small molecule crystal structures, which includes small organics as well as MOFs and other porous materials.[21] As shown in Figure 1, the number of entries in the CSD has substantially increased over the last 44 years, reaching a milestone of 850 000+ entries in 2016. Among these, the number of MOFs has also increased dramatically in the past decade to an estimated number of ca. 70 000 materials, vide infra. This trend is only going to increase even further as the

Figure 1. Growth of the CSD and MOF entries since 1972. The inset shows the MOF self-assembly process from building blocks: metals (red spheres) and organic ligands (blue struts).

Received: February 2, 2017
Revised: March 14, 2017
Published: March 14, 2017

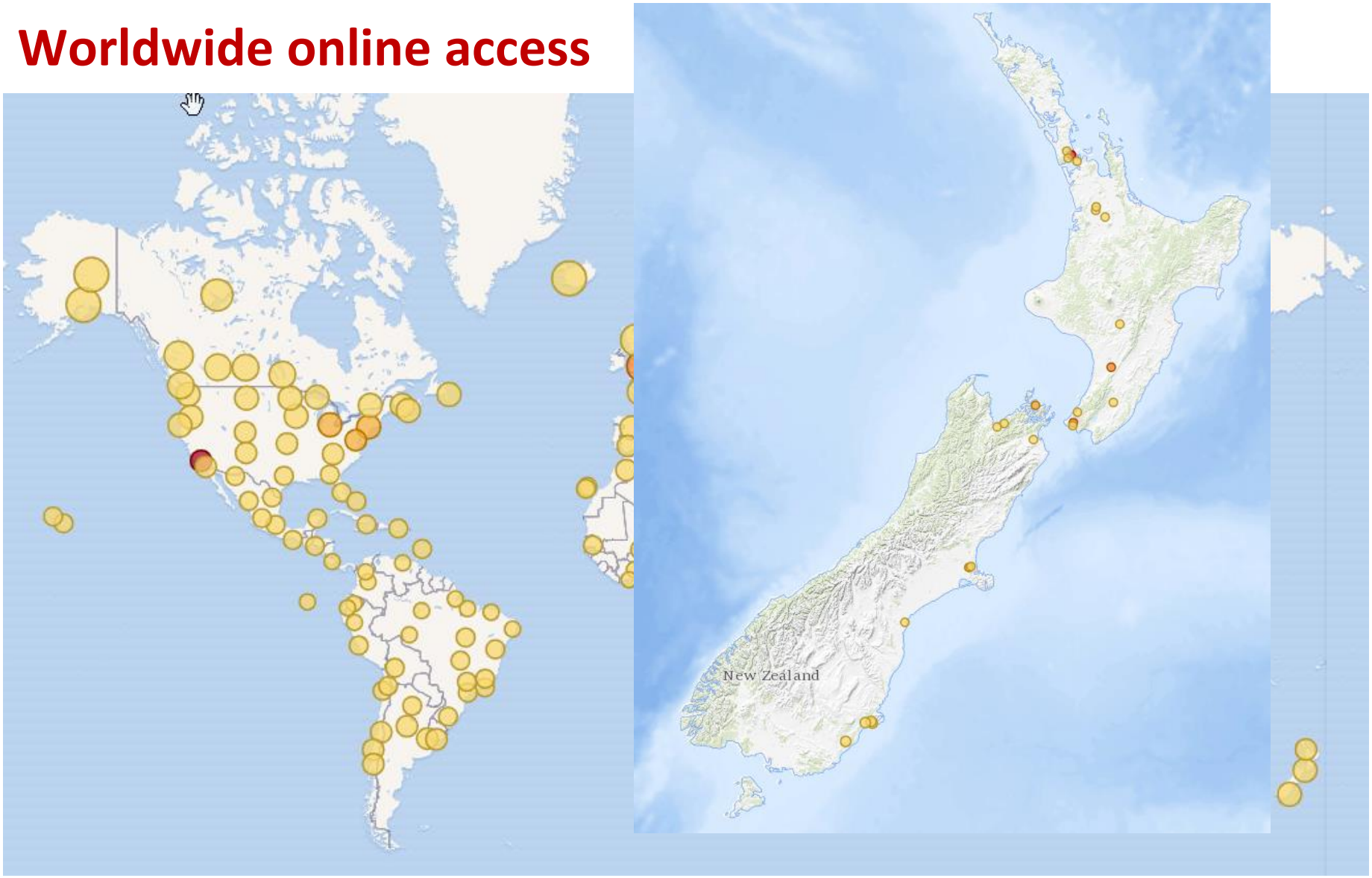
DOI: 10.1021/acs.chemmater.7b00441
Chem. Mater. 2017, 29, 2618−2625

Accessing data from general resources



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

Worldwide online access




AsCA 2018 CRYSTAL 32
December 2-5 2018 | Auckland, New Zealand
Auckland
New Zealand

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
  - lightfoot@ccdc.cam.ac.uk
  - support@ccdc.cam.ac.uk
- Thank you for your attention