

# CrystalEye

<http://wmm.ch.cam.ac.uk/crystaleye>

**Nick Day**

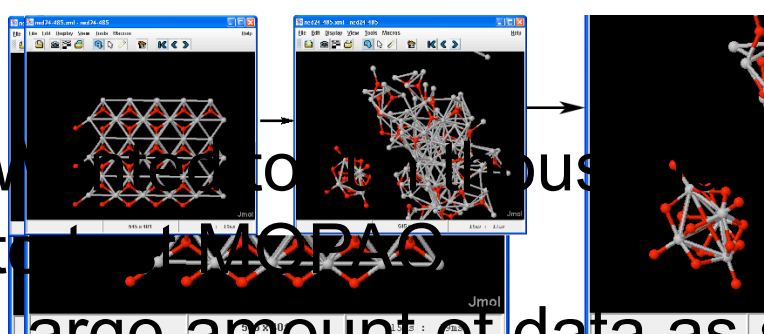
[ned24@cam.ac.uk](mailto:ned24@cam.ac.uk)

Unilever Centre for Molecular Informatics, University of Cambridge, UK.

August 23<sup>th</sup>, 2008

# Crystallographic Data as a Research Resource

Can MOPAC predict crystal structures?

- Experimental                      Calculated
- 
- Very few to thousands of calculations to MOPAC.
  - Large amount of data as supplementary information in CIF form, but...
    - No easy access to them.
    - No connection tables within.

# CrystalEye – a semantic knowledgebase

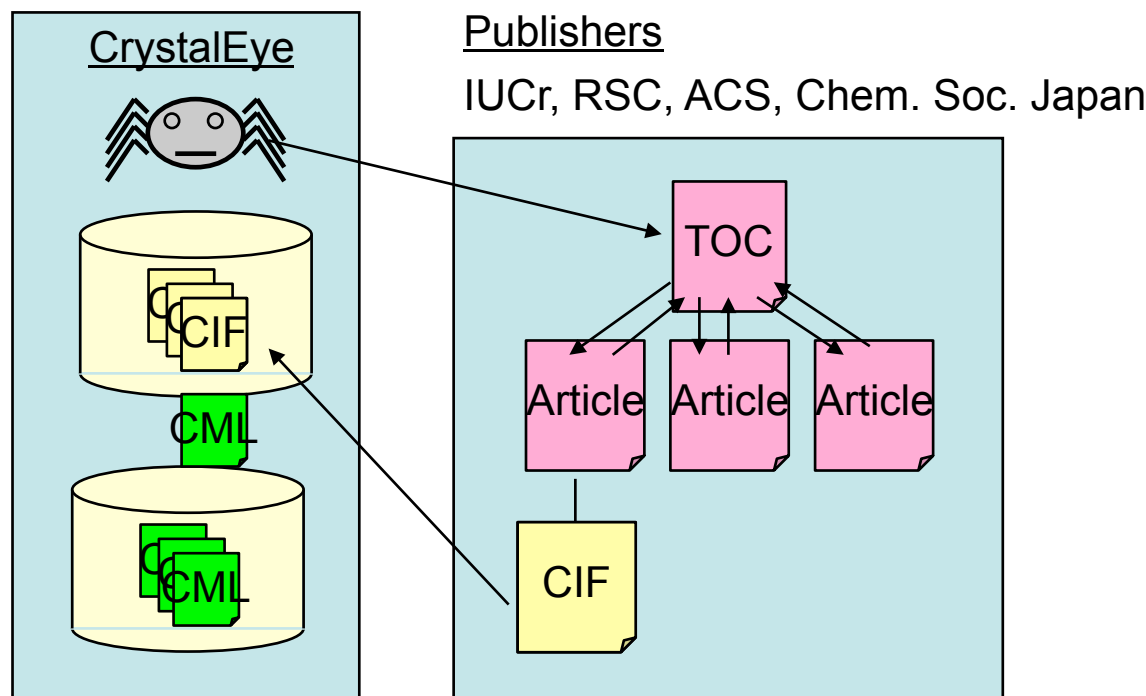
- A completely automated knowledgebase for e-Science that would:

- - - -

-

# Aggregation – Today

- Web spider checks publisher's sites every day.
- Parses HTML pages, extracts links and follows them.
- Currently over 88,000 validated CIFs (>120,000 structures).



- Also aggregate from the Crystallography Open Database.

# Aggregation – Current Developments

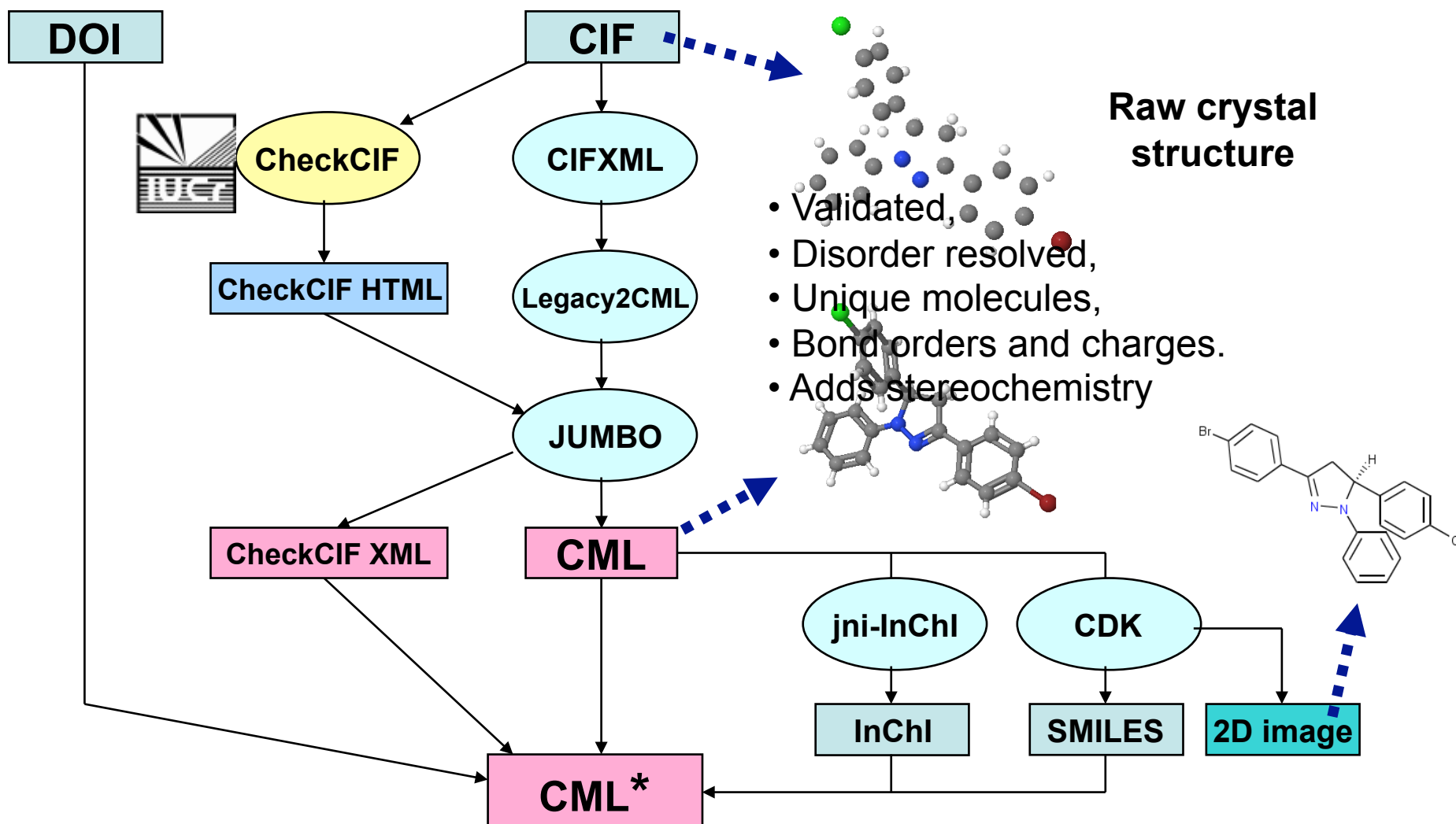
- Departmental Repository for Crystallography, **C3DeR**

Number of Structures:	269		P	Published				
Note: Dates only start from 15/03/01			D	Unpublished, can be released to department				
			G	Unpublished, can be released to group				
Key:		Data on CIF	E	Structure Embargoed				
		Calculated by computer						
		Information from literature						
		Information supplied by group						
Department ID	DOI	Date	Status	Space Group	Sum Formula	Group Leader	Chemist	Crystallographer
wj0440	10.1107/S1600536805008779	20/12/2004	P	P2(1)/c	C10 H12 N2 O7 S	Bill Jones	A. V. Trask	J. Davies
wj0302	10.1107/S1600536804004453	21/05/2003	P	P2(1)/n	C20 H18 Cl N O4	Bill Jones	A. V. Trask	J. Davies
wj0002	10.1107/S1600536803004446	15/03/2001	P	Pca2(1)	C13 H11 N O4	Bill Jones	N. Shan	J. Davies
wj9904	10.1023/A:1015067422886	15/03/2001	P	P b c a	C12 H20 N2 Ni O4 S6	Bill Jones	A. Bond	A. Bond
wj9905	10.1023/A:1015067422886	15/03/2001	P	P b c a	C12 H20 Co N2 O4 S6	Bill Jones	A. Bond	A. Bond

- Part of the eCrystals federation.



# Processing – Workflow



# Dissemination

1. Browse
2. Search
3. RSS feeds
4. Harvesting
5. Example services

# Dissemination

1. Browse
2. Search
3. RSS feeds
4. Harvesting
5. Example services



# Dissemination 1 – Browse

4-Aminophthalimide - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://wwwmm.ch.cam.ac.uk/crystaleye/summary/acta/e/2008/08-00/data/bg2195/bg2195sup1

Weighted R Factor (All) 0.1256

Available Resources

Crystal Components

Moieties

Result files

Raw CML

Complete CML

CIF (cached / original)

Validation

CheckCIF

Images

Ellipsoid

InChI: InChI=1/C8H6N2O2/c9-4-1-2-5-6(3-4)8(12)10-7(5)11/h1-3H,9H2,(H,10,11,12)

SMILES: [H]C1=C([H])C(=C([H])C=2C(=O)N([H])C(=O)C1=2)N([H])[H]

Jmol script terminated

# Dissemination

1. Browse
- 2. Search**
3. RSS feeds
4. Harvesting
5. Example services

# Dissemination 2 – Search

Magnetism and crystal structure of an N3O3-coordinated iron(II) complex - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://wwwmm.ch.cam.ac.uk/crystaleye/summary/acta/c/2008/06-00/data/sq3144/sq3144sup1, Google

## Magnetism and crystal structure of an N3O3-coordinated iron(II) complex

[OPEN DATA](#)

[<< Table of Contents](#)

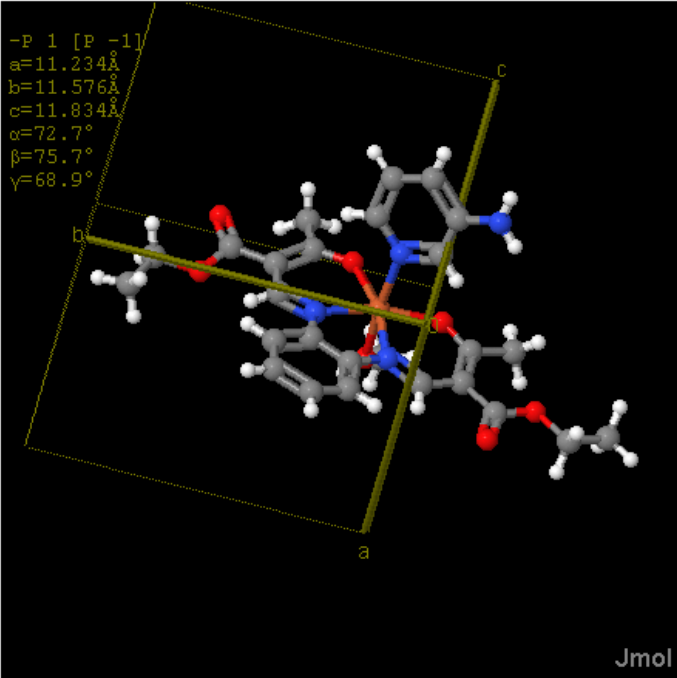
**Publisher:** Acta Crystallographica  
**Journal:** Section C  
**Year/Issue:** 2008/06-00

**Article (via DOI):** [10.1107/S0108270108013498](https://doi.org/10.1107/S0108270108013498)  
**Compound Class:** organometallic  
**Date Recorded:** 2008-04-22

**Contact Author:** Dr Birgit Weber  
**e-mail:** [bwmch@cup.uni-muenchen.de](mailto:bwmch@cup.uni-muenchen.de)

Data collection parameters

Chemical formula sum	C <sub>26</sub> H <sub>32</sub> FeN <sub>4</sub> O <sub>7</sub>
Chemical formula moiety	C <sub>26</sub> H <sub>32</sub> FeN <sub>4</sub> O <sub>7</sub>
Crystal system	triclinic
Space group H-M	P -1
Space group Hall	-P 1



-P 1 [P -1]  
a=11.234 Å  
b=11.576 Å  
c=11.834 Å  
α=72.7°  
β=75.7°  
γ=68.9°

Jmol

Show no. of unit cells along axis:

a:

Jmol script terminated

# Dissemination

1. Browse
2. Search
- 3. RSS feeds**
4. Harvesting
5. Example services

# Dissemination 3 – RSS Feeds

(E)-1-Ferrocenyl-3-phenylprop-2-en-1-one

OPEN DATA

[Table of Contents](#)

**Publisher:** Acta Crystallographica  
**Journal:** Section E  
**Year/Issue:** 2008/08-00

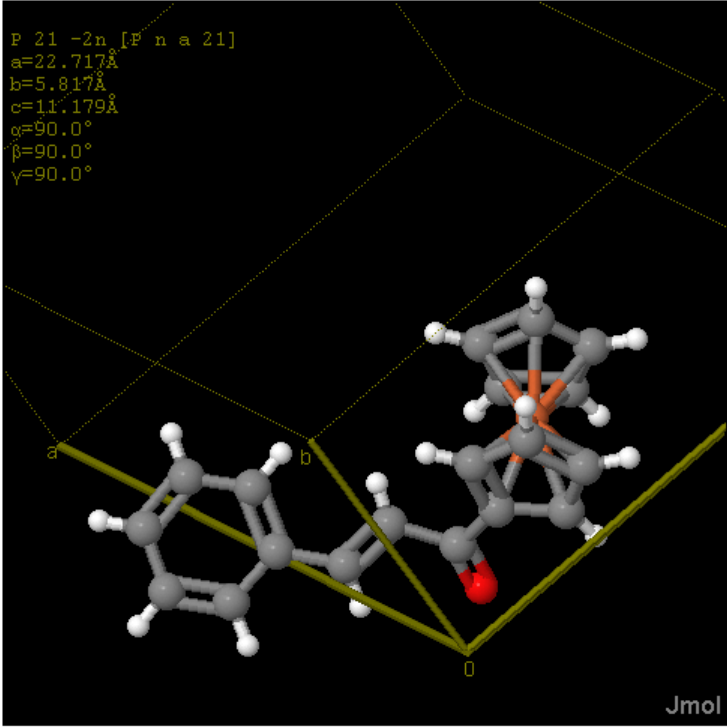
**Article (via DOI):** [10.1107/S1600536808020059](https://doi.org/10.1107/S1600536808020059)  
**Compound Class:** organometallic  
**Date Recorded:** 2008-06-19

**Contact Author:** Xiao-Lan Liu  
**e-mail:** [xiaolanliu998@yahoo.com.cn](mailto:xiaolanliu998@yahoo.com.cn)

Data collection parameters

Chemical formula sum	C <sub>19</sub> H <sub>16</sub> FeO
Chemical formula moiety	C <sub>19</sub> H <sub>16</sub> FeO
Crystal system	Orthorhombic

Crystallographic data:  
P 21 -2n [P n a 21]  
a=22.717Å  
b=5.817Å  
c=11.179Å  
α=90.0°  
β=90.0°  
γ=90.0°



Jmol

Jmol script terminated



UNIVERSITY OF  
CAMBRIDGE

# Dissemination

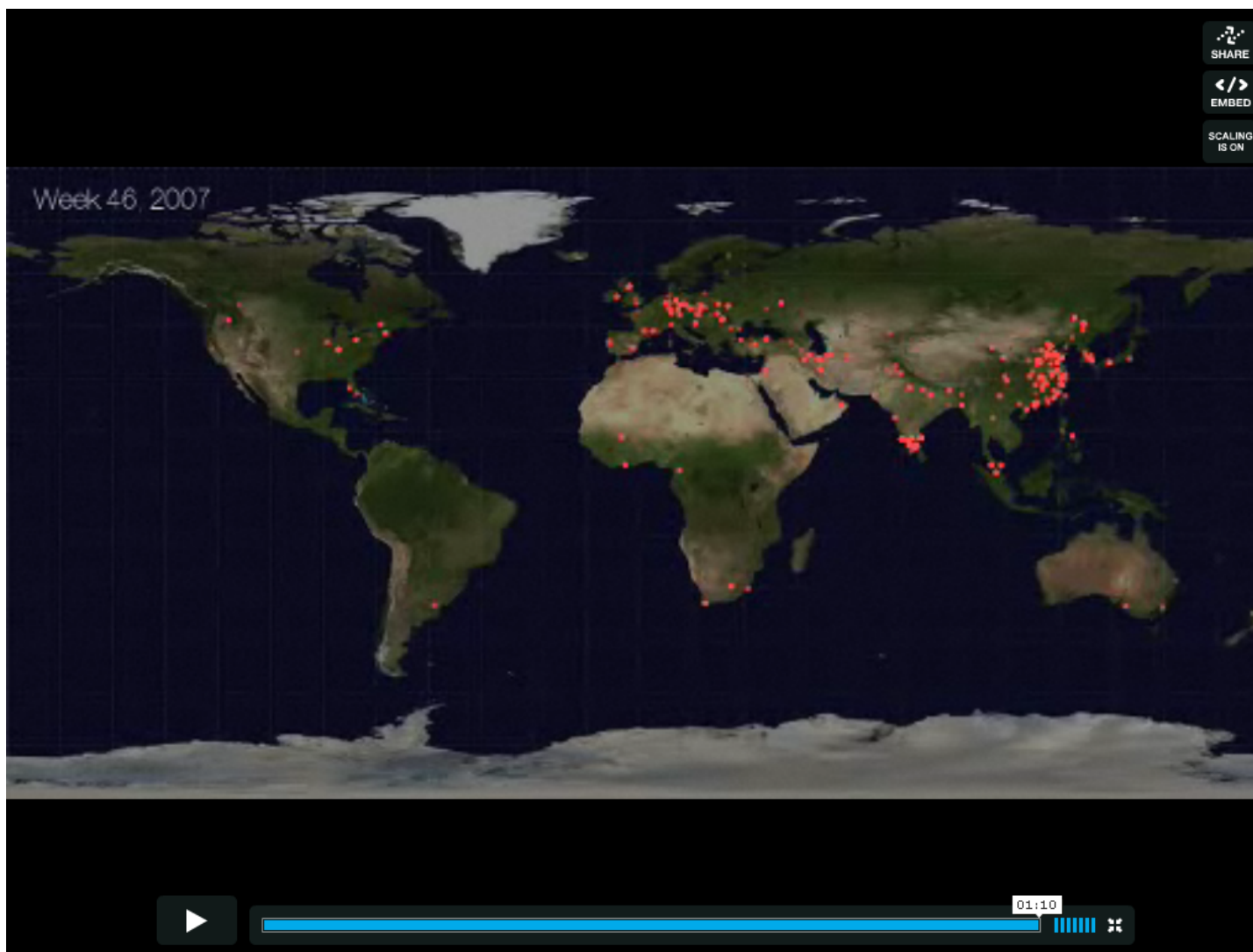
1. Browse
2. Search
3. RSS feeds
- 4. Harvesting**
5. Example services

# Dissemination 4 – Harvesting

- Users wanted to harvest the entire CML collection using RSS feeds.
- However, no standard way for harvesters to discover or recover from missed items using standard RSS.
- Atom Archive feeds (RFC5005) extend Atom to fix these problems.



# Reuse of Harvested CrystalEye data



UNIVERSITY OF  
CAMBRIDGE

Unilever  
Cambridge  
Centre for Molecular Informatics



# Dissemination

1. Browse
2. Search
3. RSS feeds
4. Harvesting
- 5. Example services**

# Dissemination 5 – Example Services

CrystalEye: Structures containing C-Fe bonds<br />between 1.93-1.94 Å - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://wwwmm.ch.cam.ac.uk/crystaleye/bondlengths/C-Fe/1.93-1.94.html

## STRUCTURES CONTAINING C-Fe BONDS BETWEEN 1.93-1.94 Å

C <sub>27</sub> H <sub>38</sub> FeN <sub>12</sub> O <sub>4</sub>	<a href="#">view</a>	<a href="#">view</a>
(C <sub>24</sub> H <sub>48</sub> Fe <sub>2</sub> N <sub>24</sub> Ni <sub>3</sub> )(H <sub>2</sub> O) <sub>3</sub>	<a href="#">view</a>	<a href="#">view</a>
C <sub>25</sub> H <sub>30</sub> Fe <sub>2</sub> O <sub>4</sub>	<a href="#">view</a>	<a href="#">view</a>
C <sub>20</sub> H <sub>21</sub> FeNO <sub>2</sub>	<a href="#">view</a>	<a href="#">view</a>
(C <sub>86</sub> H <sub>110</sub> Fe <sub>2</sub> Li <sub>2</sub> N <sub>4</sub> O <sub>12</sub> )(C <sub>4</sub> H <sub>8</sub> O) <sub>2</sub>	<a href="#">view</a>	<a href="#">view</a>
C <sub>7</sub> H <sub>13</sub> FeN <sub>4</sub> O	<a href="#">view</a>	<a href="#">view</a>
C <sub>12</sub> H <sub>19</sub> FeLaN <sub>9</sub> O <sub>8</sub>	<a href="#">view</a>	<a href="#">view</a>
C <sub>34</sub> H <sub>37.5</sub> F <sub>12</sub> Fe <sub>4</sub> N <sub>1.5</sub> P <sub>2</sub>	<a href="#">view</a>	<a href="#">view</a>

3D visualization of a complex organometallic structure (Jmol). The structure shows a central iron atom coordinated by a complex ligand system. The unit cell parameters are displayed: a=8.336 Å, b=13.340 Å, c=10.429 Å, α=90.0°, β=92.0°, γ=90.0°.

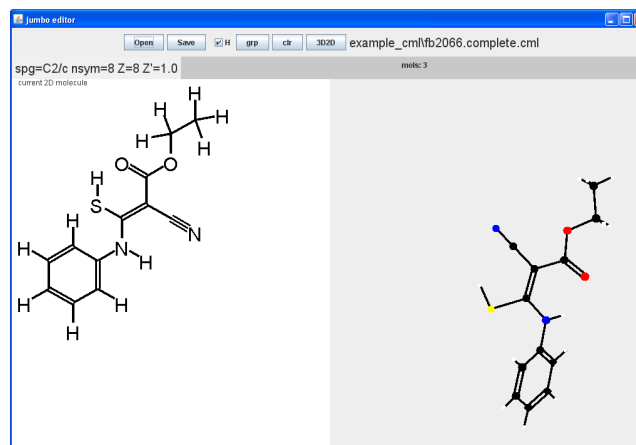
Chemical structure diagram of a complex organometallic compound, showing a central iron atom coordinated by a complex ligand system, including a 1,5-diazacyclooctane ring and several nitrile groups.

Find:  [Next](#) [Previous](#) [Highlight all](#) ☐ Match case

Jmol script terminated

# Related Work – C3DE

- 2D Structure/connection table editor – sponsored by IUCr .
- Near-automatic authoring of 2D structures from CIF.



# Thanks

- UCC – Peter Murray-Rust, Jim Downing, Andrew Walkingshaw
- Summer students – Mark Holt, Dan Hagon, Lee Harper, David Bebb
- Collaborators – IUCr, University of Southampton
- Sponsors – IUCr, JISC