

CrystalEye

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

Nick Day

ned24@cam.ac.uk

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

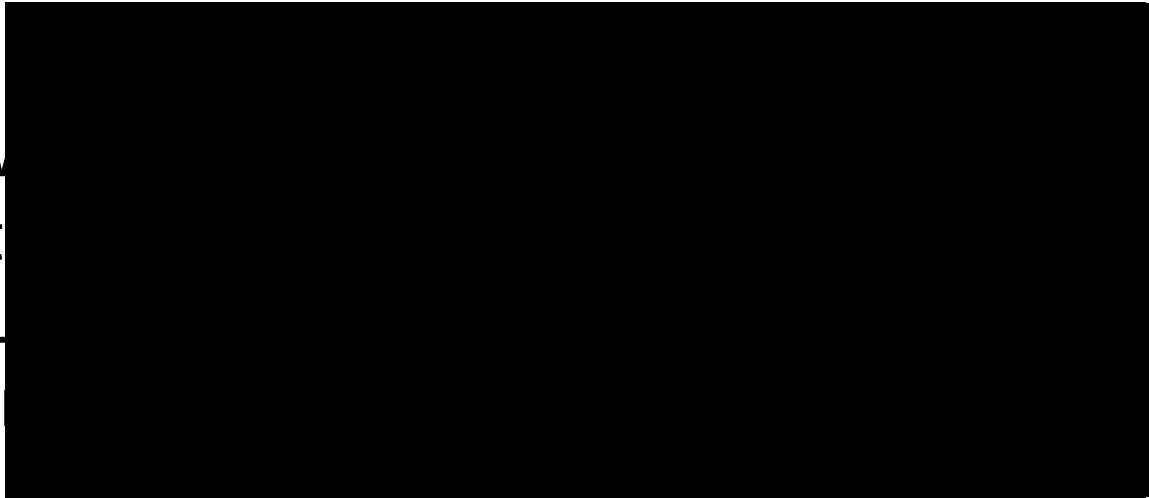
August 23th, 2008

Crystallographic Data as a Research Resource

Can MOPAC predict crystal structures?

Experimental

Calculated

- V
 - L
- 

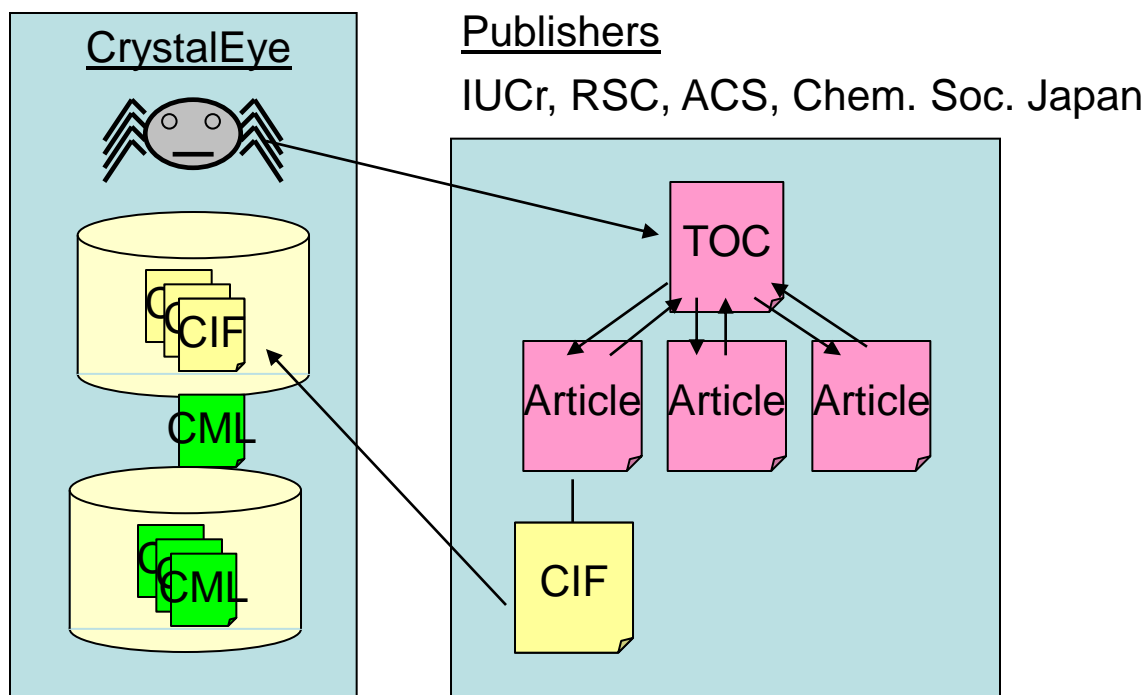
- No easy access to them.
- No connection tables within.

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

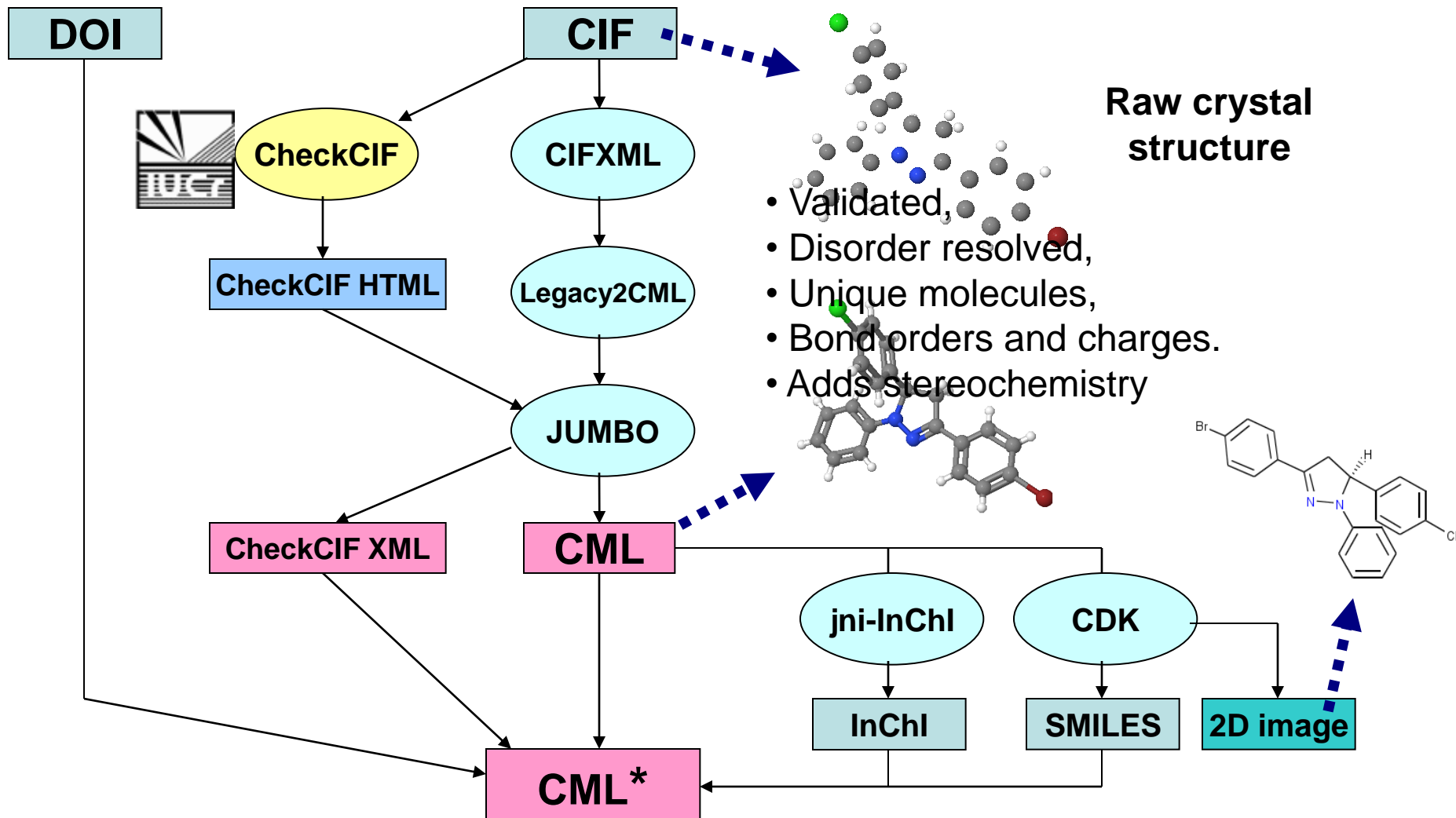
- Departmental Repository for Crystallography, **C3DeR**

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Note: Dates only start from 15/03/01		D	Unpublished, can be released to department					
		G	Unpublished, can be released to group					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:		E	Structure Embargoed					
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

XXX http://wwmm.ch.cam.ac.uk/crystaleye/summary/acta/e/2008/08-00/data/bg2195/bg2195sup1 ☆ Google

Weighted R Factor (All) 0.1256

Available Resources

enter

reset

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://wwmm.ch.cam.ac.uk/crystaleye/summary/acta/c/2008/06-00/data/sq3144/sq3144sup1

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

Data collection parameters

Chemical formula sum	C ₂₆ H ₃₂ FeN ₄ O ₇
Chemical formula moiety	C ₂₆ H ₃₂ FeN ₄ O ₇
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

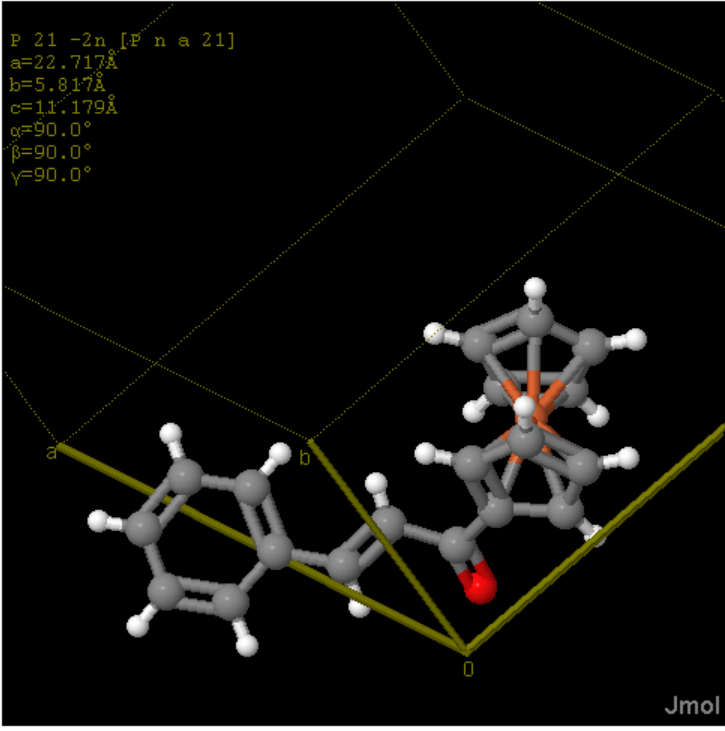
Article (via DOI): [10.1107/S1600536808020059](https://doi.org/10.1107/S1600536808020059)
Compound Class: organometallic
Date Recorded: 2008-06-19

Contact Author: Xiao-Lan Liu
e-mail: xiaolanliu998@yahoo.com.cn

Data collection parameters

Chemical formula sum	C ₁₉ H ₁₆ FeO
Chemical formula moiety	C ₁₉ H ₁₆ FeO
Crystal system	Orthorhombic

```
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



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



n
53

Dissemination

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

Dissemination 5 – Example Services

CrystalEye: Structures containing C-Fe bonds
between 1.93-1.94 Å - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

XXX http://wwmm.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 ₂₇ H ₃₈ FeNi ₂ O ₄	view	view
(C ₂₄ H ₄₈ Fe ₂ N ₂₄ Ni ₃)(H ₂ O) ₃	view	view
C ₂₅ H ₃₀ Fe ₂ O ₄	view	view
C ₂₀ H ₂₁ FeNO ₂	view	view
(C ₈₆ H ₁₁₀ Fe ₂ Li ₂ N ₄ O ₁₂)(C ₄ H ₈ O) ₂	view	view
C ₇ H ₁₃ FeN ₄ O	view	view
C ₁₂ H ₁₉ FeLaN ₉ O ₈	view	view
C ₃₄ H _{37.5} F ₁₂ Fe ₄ N _{1.5} P ₂	view	view

<<< < > >>> Prev 1/1 Next

-P 2m [P 1 21/n 1]
a=8.336Å
b=13.340Å
c=10.429Å
α=90.0°
β=92.0°
γ=90.0°

Jmol

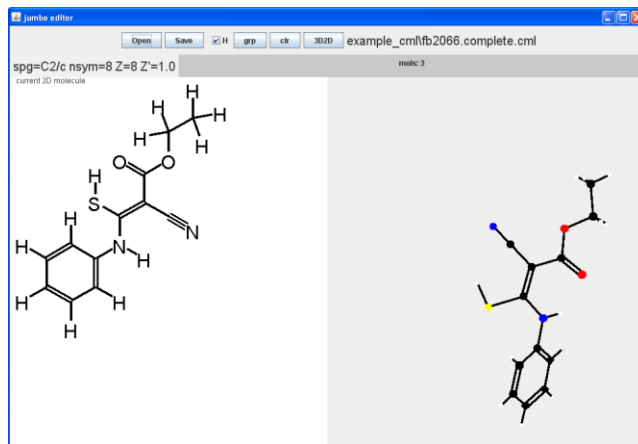
Find: organic Next Previous Highlight all Match case

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