F/LOSS for Open Science: Crystallography Open Database

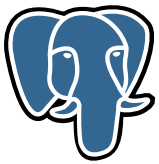
Saulius Gražulis

"Galaktika", 2013

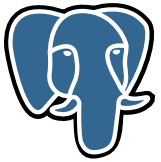
Vilnius University Institute of Biotechnology

Free software (r)evolution

Free software (r)evolution



Free software (r)evolution

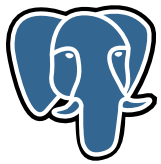


Free software (r)evolution

- A set of tools to achieve our goals
- A set of ideas how to collaborate







Applicability of open source principles to scientific investigations

Science is an open endeavour:

- it is based on open discussion and substantial criticism;
- it is based on publication of the results;
- based on the reuse of the results

Applicability of open source principles to scientific investigations

Science is an open endeavour:

- it is based on open discussion and substantial criticism;
- it is based on **publication** of the results;
- based on the reuse of the results

Applicability of open source principles to scientific investigations

Science is an open endeavour:

- it is based on open discussion and substantial criticism;
- it is based on **publication** of the results;
- based on the **reuse** of the results

The origins of COD

SDPD mailing list:
Open crystallographic database - a role for whom?

```
> When crystallographers will be tired to buy all these fragmented
> databases, maybe they will do something. A role for IUCr ? If not, a
> role for whom ?
```

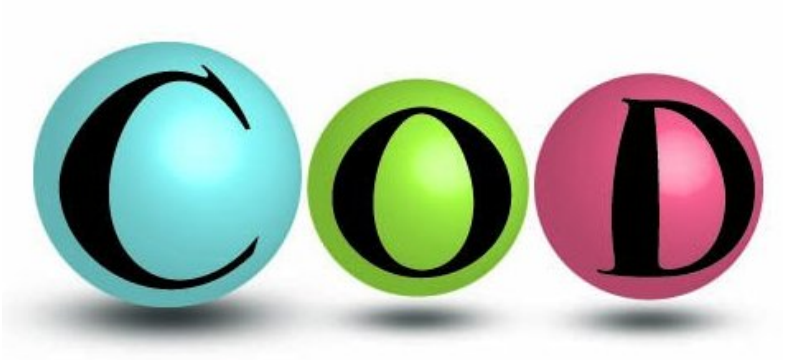
armellebail (Armel Le Bail) Wed Feb 12, 2003 7:24 pm

```
But what if crystallographers work together to establish a public
domain database with all relevant crystallographic data? This would
not only overcome the current situation with 'fragmented' databases,
it would also prevent for becoming dependent from monopolists.

What would be needed?

1. A small team of engaged scientists with some experience in database
and software design to coordinate the project.

2. The authors (i.e. the scientific community = YOU) who provides the
project with database entries (note, that if you have'nt sold your
experimental results exclusively, you are free to distribute the data
to such a database, even if they have already been part of a
publication - and a lot of good data have never been published).

3. Free software a) for maintaining the database, b) for data
evaluation and calculation of derived data (e.g. calculated powder
pattern from crystal structures for search-match purposes), c) for
browsing and retrieval.
```

gemstonede (Dr. Michael BERNDT) Fri Feb 14, 2003 1:26 pm

Crystallography Open Database

A grass-root initiative



http://www.crystallography.net/

- Total ≈**217 000** records
- **179** registered users
- **23** depositors (deposited at least one structure)
- In year 2012:
  - >**56 000** new structures uploaded (**26 000** more than last year)
  - **16** active depositors (who deposited at least one structure in 2012)

Contents and access

Crystallography Open Database

COD Home

Home
What's new?

Accessing COD Data

Browse
Search

Add Your Data

Deposit your data
Manage depositions
Manage/release prepublications

Documentation

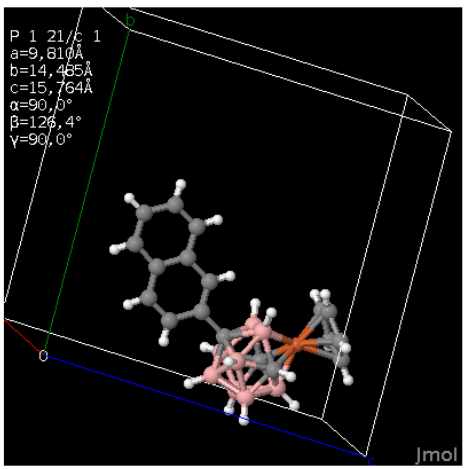
COD Wiki
Obtaining COD
Citing COD
COD Mirrors
Advices to donators

CIF Information Card

Information card for 4079785

4079784 << **4079785** >> 4100000

Preview

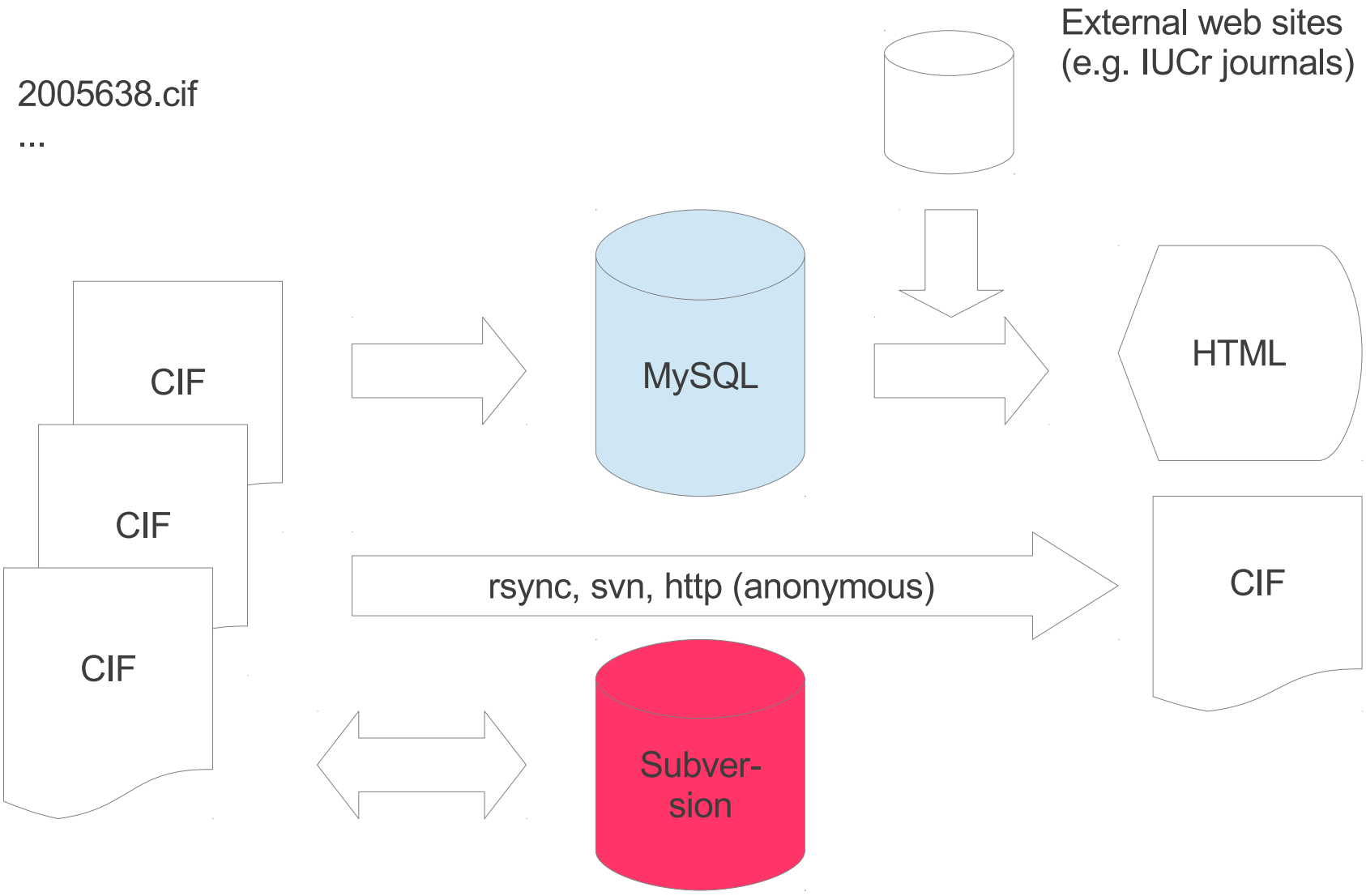


Coordinates 4079785.cif

Structure parameters

Formula	- C18 H22 B8 Fe -
Calculated formula	- C18 H22 B8 Fe -
Title of publication	Three Isomers of Aryl-Substituted
Authors of publication	Bakardjiev, Mario; Štíbr, Bohumil;
Journal of publication	Organometallics
Year of publication	2013
Journal volume	32
Journal issue	2
a	9.81 ± 0.0006 Å
b	14.4851 ± 0.0009 Å

Implementation



Possible uses

- Protein model refinement
- Data validation
- Search for a “black swan”

H = “All swans are white” (the usual case)

Possible uses

- Protein model refinement
- Data validation
- Search for a “black swan”

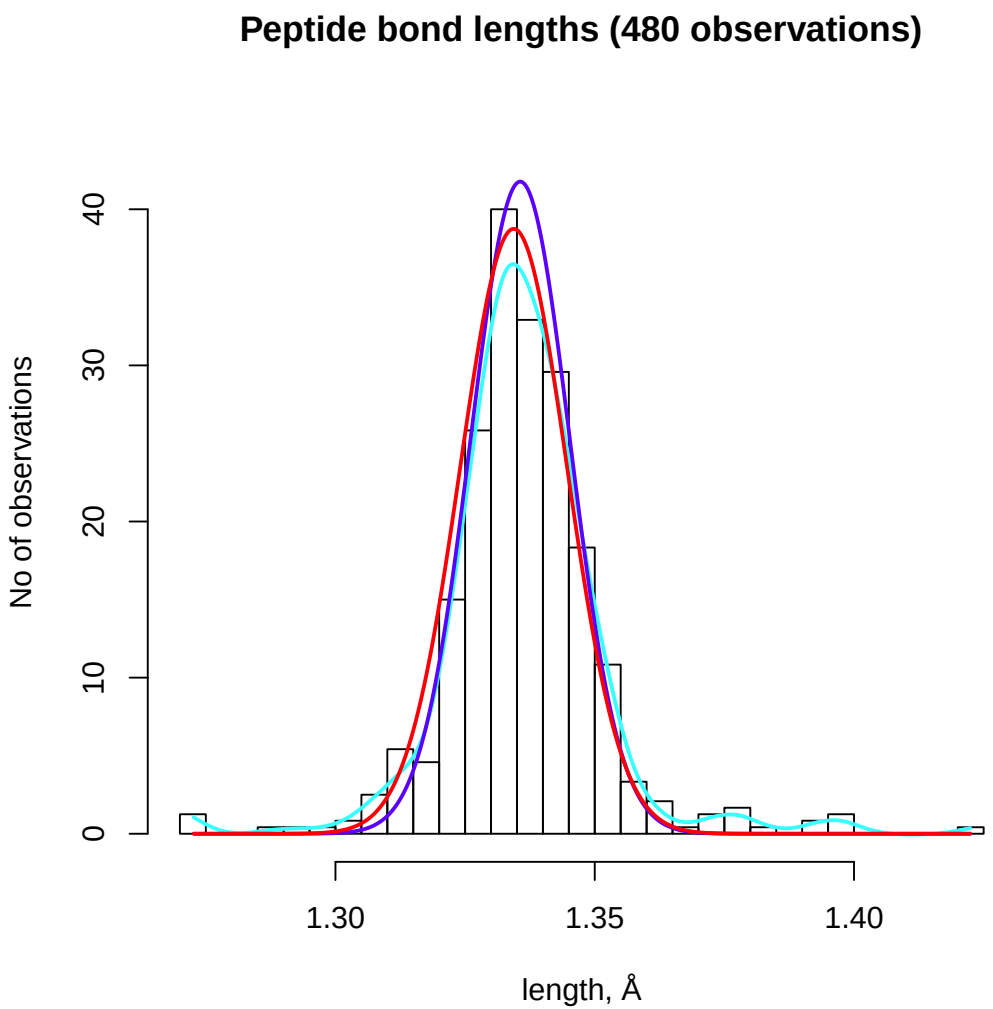
H = “All swans are white” (the usual case)
D = “A black swan has been observed” – an error or an unknown case?

Possible uses

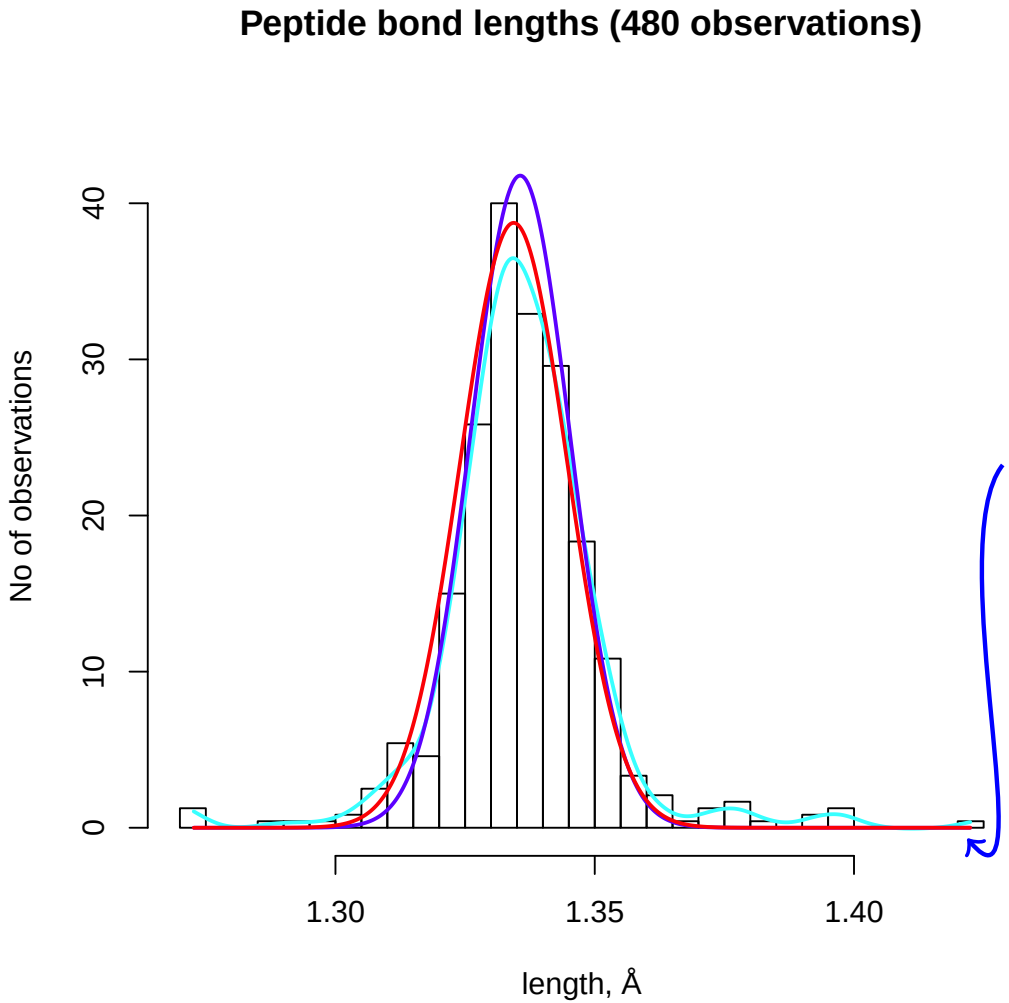
- Protein model refinement
- Data validation
- Search for a “black swan”

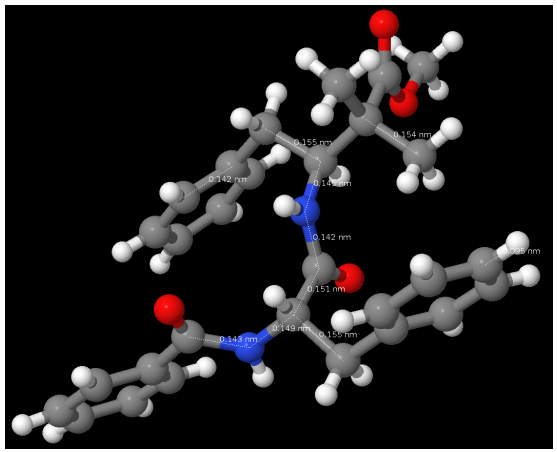
H = “All swans are white” (the usual case)
D = “A black swan has been observed” – an error or an unknown case?

Example:
Peptide bonds



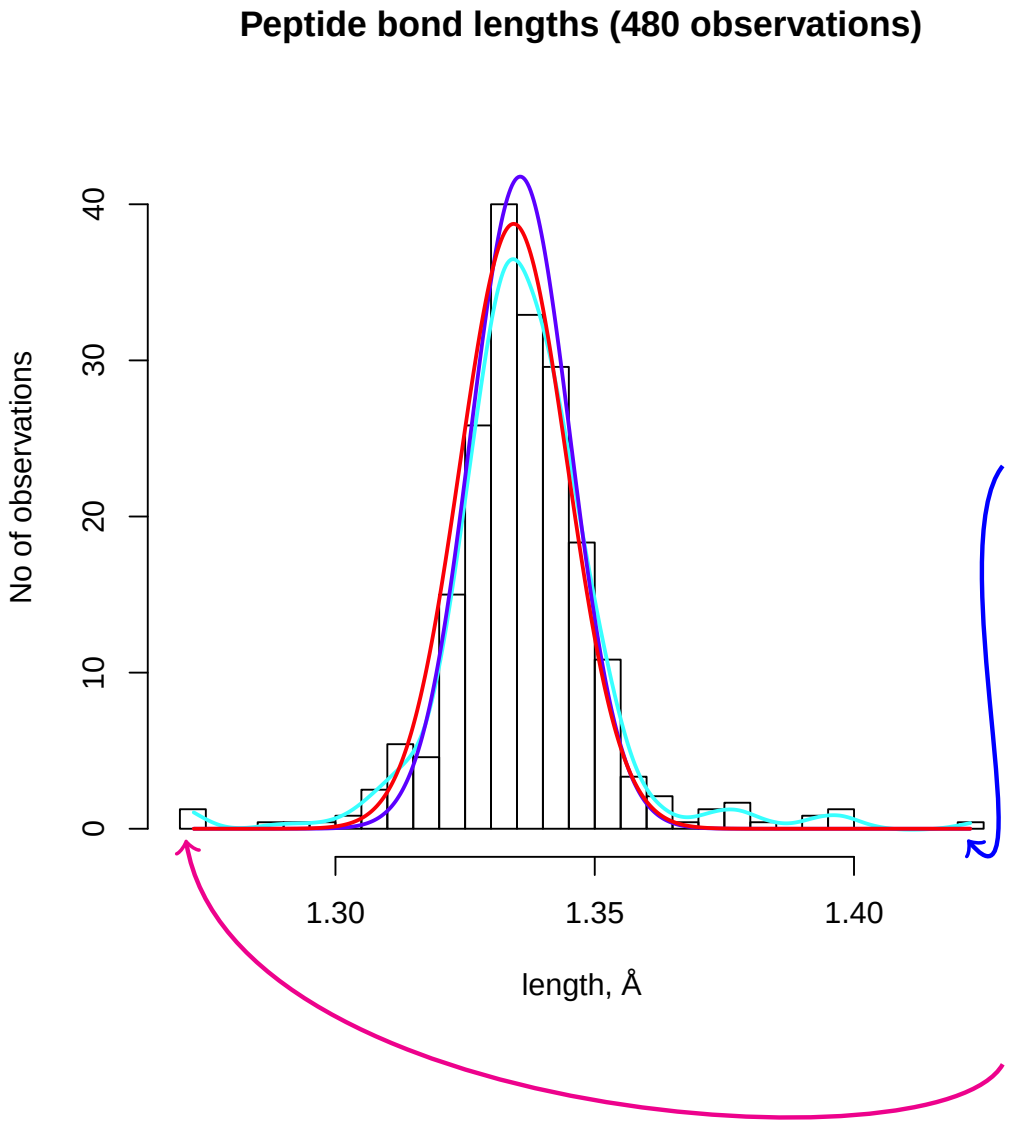
Example:
Peptide bonds

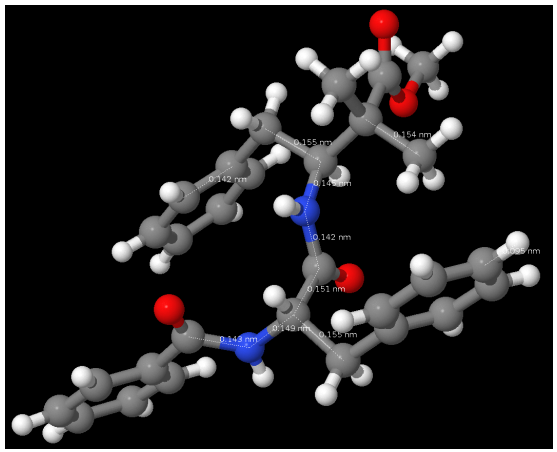




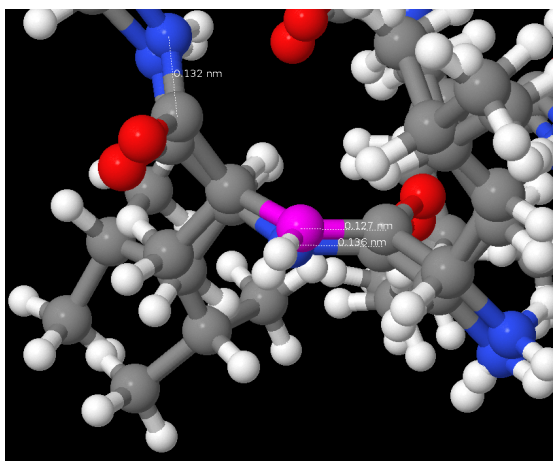
COD 4024607, $l = 1.4224$ Å

Example:
Peptide bonds





COD 4024607, $l = 1.4224$ Å



COD 7203041, $l = 1.2727$ Å

Another example:

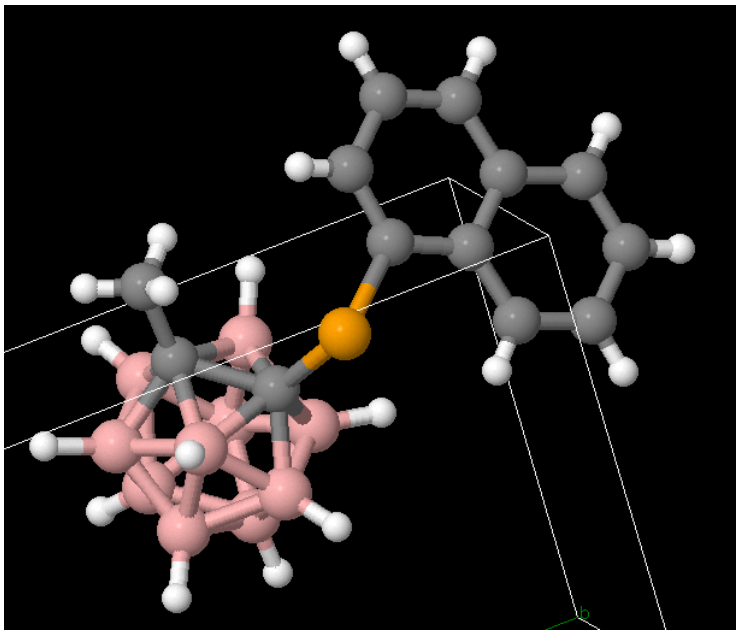
Carbon with valence 6 (!?)

We observe carbons (atoms of type “C”) with **6** neighbours at a covalent distance in our database.(?)

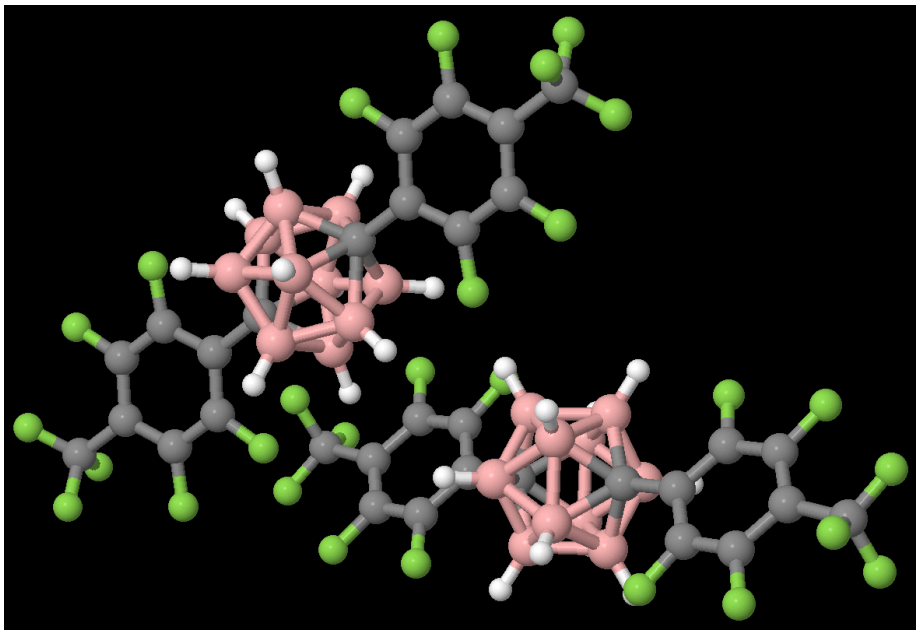
Another example:

Carbon with valence 6 (!?)

We observe carbons (atoms of type “C”) with **6** neighbours at a covalent distance in our database.(!)



COD 7015488



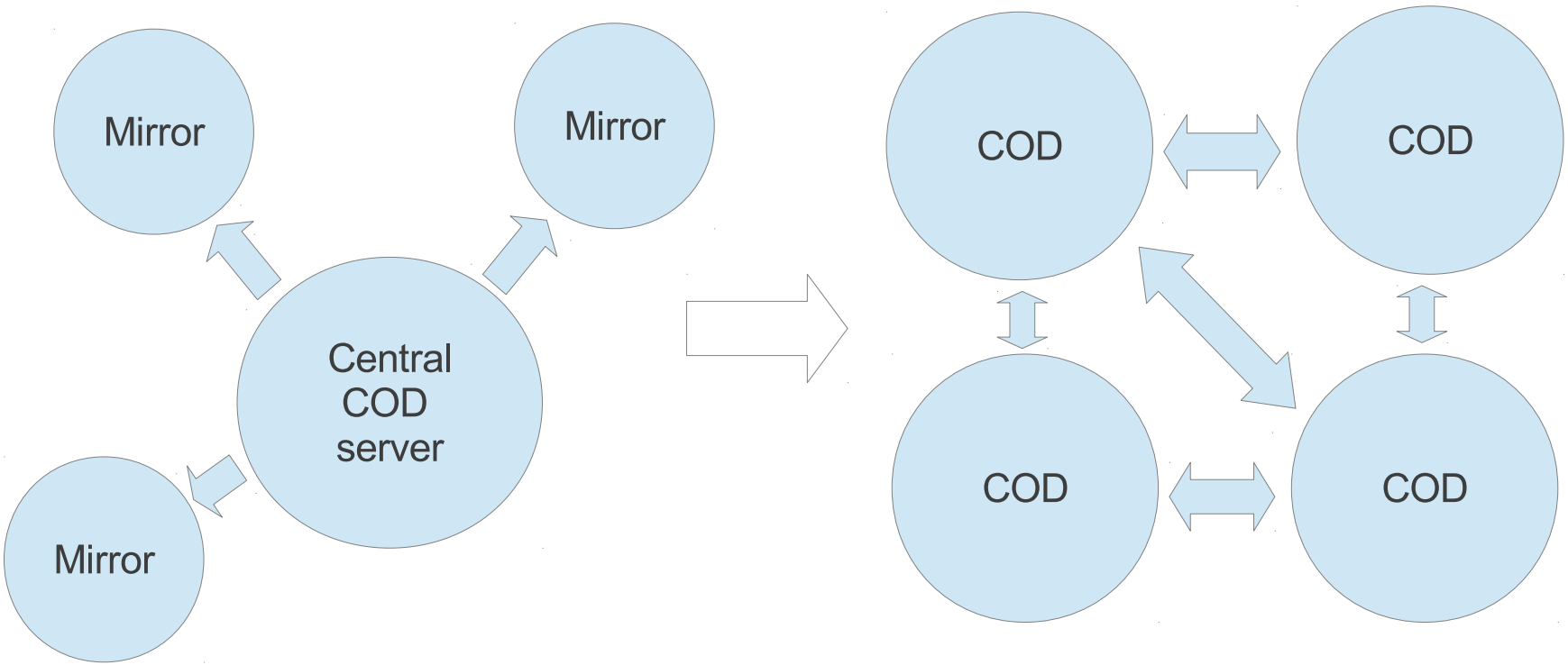
COD 7015654

Applications

New materials (e.g. semiconductors)

New medicines

A workbench for scientists from the whole world



Thank you!