Efficient long-term open-access data archiving in mining industries

Saulius Gražulis & the SOLSA consortium

Amsterdam, RTM Conference, 2017

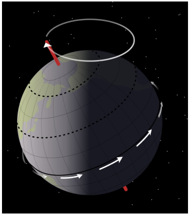
Vilnius University Institute of Biotechnology

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Data importance

Hipparchus (c. 190 – c. 120 BCE)

- measured the longitude of Spica and Regulus and other bright stars
- compared his measurements with data from his predecessors, Timocharis and Aristillus, who lived ≈**100** years before him,
- discovered what is now called *the precession of the equinoxes*



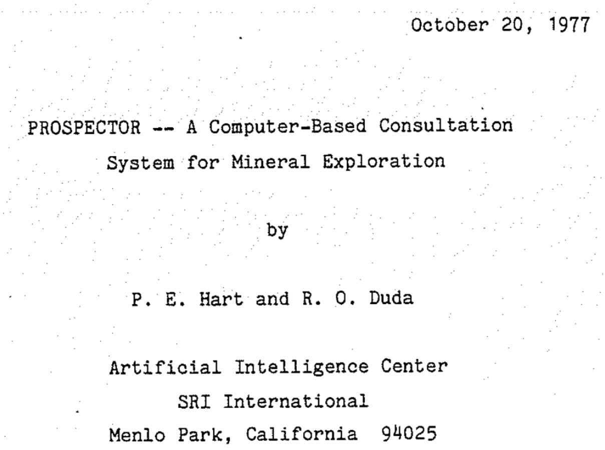
By NASA, Public Domain

(Wikipedia, see also articles on Timocharis and Aristyllus)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Data and AI systems for geology

[Hart and Duda, 1977]



October 20, 1977

PROSPECTOR -- A Computer-Based Consultation
System for Mineral Exploration

by

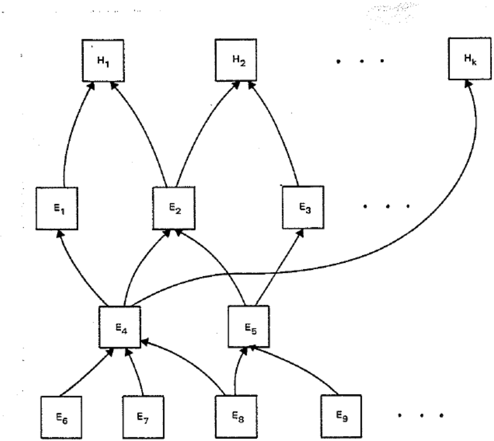
P. E. Hart and R. O. Duda

Artificial Intelligence Center
SRI International
Menlo Park, California 94025

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

The PROSPECTOR network of inference

[Hart and Duda, 1977]



This project has received funding from the the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Data kinds in the SOLSA project

Sonic drilling in Ni laterites

courtesy of Eijkelkamp SonicSampDrill

Discover SOLSA

http://solsa-mining.eu/

- Crystal structures (COD)
- Raman spectra (ROD)
- Hyperspectral spectra (HOD)

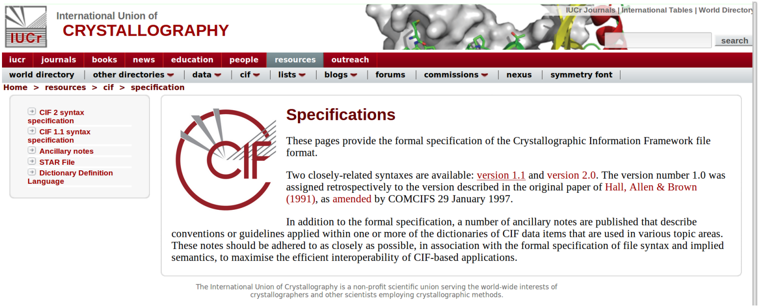
This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Requirements for long-term data archiving and reuse

- Platform independence
  - Text-based formats (ASCII, UTF-8)
- Software independence
- Network-transparency
  - Standard, open protocols (W3C http)
  - Standard, open data carrier formats (JSON, XML, CIF).
  - RESTful servers
- Machine-readable semantics
  - Dictionaries, schemas
- Durability
  - Persistent identifiers
  - Open data principles
  - FAIR principles

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Data exchange in crystallography



International Union of
CRYSTALLOGRAPHY
IUCr Journals | International Tables | World Directory
search
iucr journals books news education people resources outreach
world directory other directories data cif lists blogs forums commissions nexus symmetry font

Home > resources > cif > specification

- CIF 2 syntax specification
- CIF 1.1 syntax specification
- Ancillary notes
- STAR File
- Dictionary Definition Language

Specifications

These pages provide the formal specification of the Crystallographic Information Framework file format.

Two closely-related syntaxes are available: version 1.1 and version 2.0. The version number 1.0 was assigned retrospectively to the version described in the original paper of Hall, Allen & Brown (1991), as amended by COMCIFS 29 January 1997.

In addition to the formal specification, a number of ancillary notes are published that describe conventions or guidelines applied within one or more of the dictionaries of CIF data items that are used in various topic areas. These notes should be adhered to as closely as possible, in association with the formal specification of file syntax and implied semantics, to maximise the efficient interoperability of CIF-based applications.

The International Union of Crystallography is a non-profit scientific union serving the world-wide interests of crystallographers and other scientists employing crystallographic methods.

[Hall et al., 1991]

The Crystallographic Interchange File/Framework (CIF):

- Provides standard means for data publishing and exchange;
- Is suitable for archiving;
- Is maintained by the IUCr;

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

CIF for scientific data

examples/data/2100858-head.cif:

```
data_2100858
loop_
_publ_author_name
'Buttner, R. H.'
'Maslen, E. N.'
_publ_section_title
;
 Structural parameters and electron difference density in BaTiO~3~
;
_journal_issue                    6
_journal_name_full                'Acta Crystallographica Section B'
_journal_page_first               764
_journal_page_last                769
_journal_volume                   48
_journal_year                     1992
_chemical_compound_source         'synthetic, from a mixture of KF:KMoO4:BaTiO3'
_chemical_formula_sum             'Ba O3 Ti'
_chemical_formula_weight          233.24
_symmetry_cell_setting            tetragonal
_symmetry_space_group_name_Hall   'P 4 -2'
_symmetry_space_group_name_H-M    'P 4 m m'
_cell_angle_alpha                 90.0
_cell_angle_beta                  90.0
_cell_angle_gamma                 90.0
_cell_formula_units_Z             1
_cell_length_a                    3.9998(8)
_cell_length_b                    3.9998(8)
_cell_length_c                    4.0180(8)
```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Controlled vocabularies

examples/dictionaries/cif-core-example.cif:

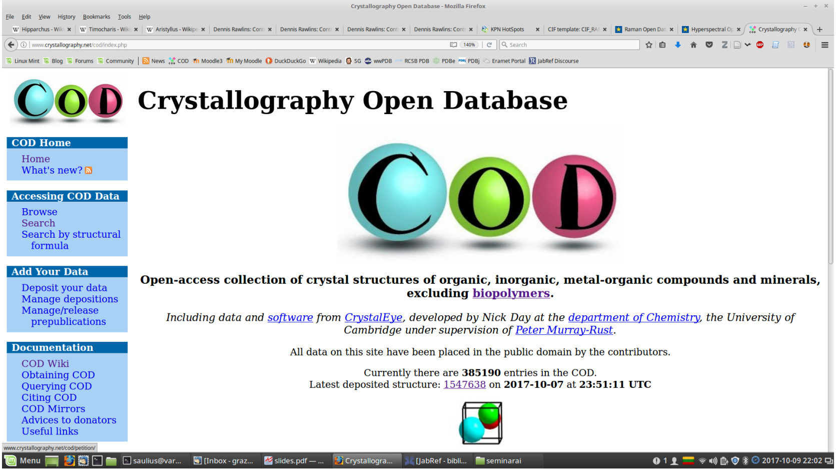
```
data_cell_length_
    loop_ _name                ’_cell_length_a’
                               ’_cell_length_b’
                               ’_cell_length_c’
    _category                    cell
    _type                        numb
    _type_conditions             esd
    _enumeration_range           0.0:
    _units                       A
    _units_detail              ’angstroms’
    _definition
;              Unit-cell lengths in angstroms corresponding to the structure
               reported. The values of _refln_index_h, *_k, *_l must
               correspond to the cell defined by these values and _cell_angle_
               values. The values of _diffrn_refln_index_h, *_k, *_l may not
               correspond to these values if a cell transformation took place
               following the measurement of the diffraction intensities. See
               also _diffrn_reflns_transf_matrix_.
;
```

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 689868.

Crystallographic data

The Crystallography Open Database

http://www.crystallography.net/cod



Crystallography Open Database

COD Home

Home
What's new?

Accessing COD Data

Browse
Search
Search by structural formula

Add Your Data

Deposit your data
Manage depositions
Manage/release prepublications

Documentation

COD Wiki
Obtaining COD
Querying COD
Citing COD
COD Mirrors
Advices to donators
Useful links

Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers.

Including data and software from CrystalEye, developed by Nick Day at the department of Chemistry, the University of Cambridge under supervision of Peter Murray-Rust.

All data on this site have been placed in the public domain by the contributors.

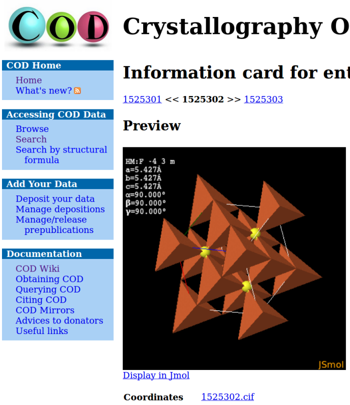
Currently there are **385190** entries in the COD.
Latest deposited structure: 1547638 on **2017-10-07** at **23:51:11 UTC**

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

A COD crystal structure page example

Sphalerite

http://www.crystallography.net/cod/1525302.html



Crystallography Open Database

COD Home

- Home
- What's new?

Accessing COD Data

- Browse
- Search
- Search by structural formula

Add Your Data

- Deposit your data
- Manage depositions
- Manage/release prepublications

Documentation

- COD Wiki
- Obtaining COD
- Querying COD
- Citing COD
- COD Mirrors
- Advices to donators
- Useful links

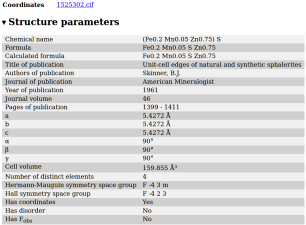
Information card for entry 1525302

1525301 << **1525302** >> 1525303

Preview

Display in Jmol

Coordinates 1525302.cif



Coordinates 1525302.cif

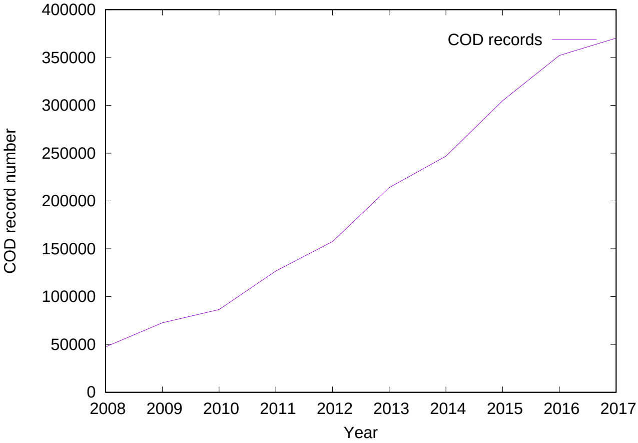
▾ Structure parameters

Chemical name	(Fe0.2 Mn0.05 Zn0.75) S
Formula	Fe0.2 Mn0.05 S Zn0.75
Calculated formula	Fe0.2 Mn0.05 S Zn0.75
Title of publication	Unit-cell edges of natural and synthetic sphalerites
Authors of publication	Skinner, B.J.
Journal of publication	American Mineralogist
Year of publication	1961
Journal volume	46
Pages of publication	1399 - 1411
a	5.4272 Å
b	5.4272 Å
c	5.4272 Å
α	90°
β	90°
γ	90°
Cell volume	159.855 Å^3
Number of distinct elements	4
Hermann-Mauguin symmetry space group	F -4 3 m
Hall symmetry space group	F -4 2 3
Has coordinates	Yes
Has disorder	No
Has F_{obs}	No

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD persistence

COD is on-line for 13 years, increased 7-fold over the last 8 years; currently contains over 385 000 records (October 2017):

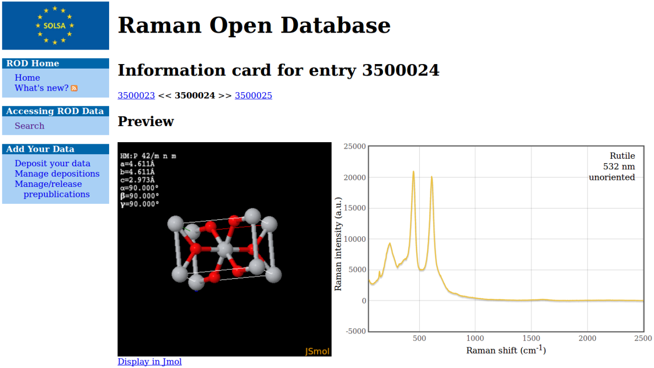


This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Raman spectroscopy data

The Raman Open Database

http://solsa.crystallography.net/rod



Raman Open Database

ROD Home
- Home
- What's new?

Accessing ROD Data
- Search

Add Your Data
- Deposit your data
- Manage depositions
- Manage/release prepublications

Information card for entry 3500024

3500023 << **3500024** >> 3500025

Preview

```
HM:P 42/m n m
a=4.611Å
b=4.611Å
c=2.973Å
α=90.000°
β=90.000°
γ=90.000°
```

Display in Jmol

Data records contributed to the ROD by Yassine El Mendili

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

ROD data files

ROD uses CIF syntax

examples/data/3500024-head.rod:

```
#------------------------------------------------------------------------------
#$Date: 2017-10-05 18:15:36 +0300 (Thu, 05 Oct 2017) $
#$Revision: 219 $
#$URL: svn://172.16.1.102/rod/cif/3/50/00/3500024.rod $
#------------------------------------------------------------------------------
#
# This file is available in the Raman Open Database (ROD),
# http://solsa.crystallography.net/rod/
#
# All data on this site have been placed in the public domain by the
# contributors.
#
data_3500024
loop_
_publ_author_name
'El Mendili, Y'
_publ_section_title
;
 SOLSA communication to ROD
;
_journal_name_full                'Personal communication to ROD'
_journal_year                     2017
_chemical_compound_source         'commercial powder Prolabo pur'
_chemical_formula_structural      'O2 Ti'
```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

The ROD dictionary

ROD uses controlled vocabulary in CIF DDLm dictionaries

http://solsa.crystallography.net/rod/cif/dictionaries/cif_raman_0.1.1.dic
http://solsa.crystallography.net/rod/cif/dictionaries/cif_rod_0.1.0.dic

`examples/dictionaries/raman-example.dic`:

```
save__raman_measurement_device.direction_polarization
    _definition.id               ’_raman_measurement_device.direction_polarization’
# ... some text omited for brevity ...
    _definition.update           2017-04-10
    _description.text
;
    The direction polarization of the measurement device.
;
# ...
    loop_
    _enumeration_set.state
    _enumeration_set.detail
    unoriented
;
 Unoriented.
;
    Z(XX)Z
;
 Laser polarized parallel to the x axis; analyzer set to pass the x axis
 polarized light.
;
```

ROD dictionaries coded by Antanas Vaitkus

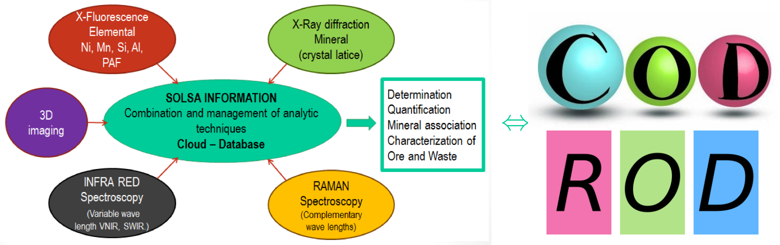
This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 689868.

Semantic versioning of the ROD dictionaries

- ROD dictionaries undergo semantic versioning:
    - Bug-fix releases (1.2.x) are compatible backwards and forward;
    - Minor releases (1.x) are backwards compatible;
    - Incompatible changes will be marked by major releases (1.x → 2.x);

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

SOLSA project, COD and ROD



X-Fluorescence
Elemental
Ni, Mn, Si, Al,
PAF

X-Ray diffraction
Mineral
(crystal latice)

3D
imaging

SOLSA INFORMATION
Combination and management of analytic techniques
Cloud – Database

Determination
Quantification
Mineral association
Characterization of
Ore and Waste

⇔

INFRA RED
Spectroscopy
(Variable wave length VNIR, SWIR.)

RAMAN
Spectroscopy
(Complementary wave lengths)

COD will be used in SOLSA for:

- mineral identification;
- subsequent data dissemination.

SOLSA data flow diagram courtesy Monique Le Guen, ERAMET.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

The fun of REST

RESTful queries [Fielding, 2000]:

- Programming language, transfer protocol **independent**
- GET queries should be null-potent (do not change anything; always provide the same result for the same query);
- POST/PUT queries should be idempotent (the same query executed several times should have the same result as just one query).

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD query examples

Web, REST, SQL

- Via the WWW interface – go for “search” in:
    - http://www.crystallography.net/cod
    - http://solsa.crystallography.net/rod
    - http://solsa.crystallography.net/hod
- Via the **stable** URLs (REST):
    - http://www.crystallography.net/cod/2000000.cif
    - http://solsa.crystallography.net/rod/3500021.rod
    - http://solsa.crystallography.net/rod/3500021.html
    - http://www.crystallography.net/cod/result?text=perovskite
- Via the **views** of the SQL database:

    -
      ```
      mysql -u cod_reader cod -h www.crystallography.net\
          -e ’select file, a, b, c, vol, formula
              from data where
                  year between 2013 and
                               2014 and
                  formula regexp " C[0-9]* "
                  order by vol desc limit 10’
      ```

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 689868.

Acknowledgements

VU Institute of Biotechnology

Virginijus Siksnys
(*head of the dept.*)

Andrius Merkys
Antanas Vaitkus
Erikas Raginis

The SOLSA team

Monique Le Guen
Beate Orberger
Daniel Chateigner
Henry Pilliere
and all the team working on the project!

COD Advisory board

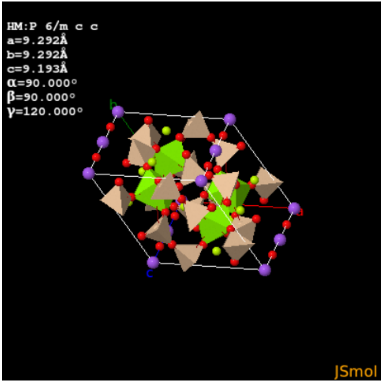
Daniel Chateigner
Robert T. Downs
Werner Kaminsky
Armel Le Bail
Luca Lutterotti
Peter Moeck
Peter Murray-Rust
Miguel Quirós

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Thank you!

http://en.wikipedia.org/wiki/Emerald



http://www.crystallography.net/5000095.html

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

References I

Fielding, R. T. (2000).
Architectural Styles and the Design of Network-based Software Architectures.
PhD thesis, University of California, Irvine.

Hall, S. R., Allen, F. H., and Brown, I. D. (1991).
The crystallographic information file (CIF): a new standard archive file for crystallography.
Acta Crystallographica Section A, 47:655–685.

Hart, P. E. and Duda, R. O. (1977).
Prospector – a computer-based consultation system for mineral exploration.
techreport, Artificial Intelligence Center, SRI International, Menlo Park, California 94025.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

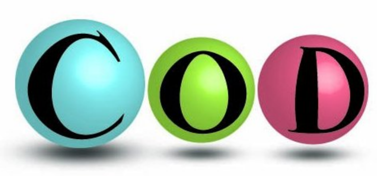
References II

Selimi, M. and Freitag, F. (2014).
Tahoe-lafs distributed storage service in community network clouds.
2014 IEEE Fourth International Conference on Big Data and Cloud Computing.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

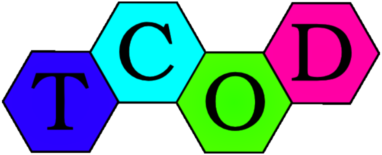
Open Crystallographic Databases

COD, TCOD, PCOD, MPOD, ...



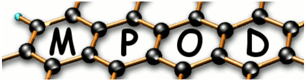
http://www.crystallography.net/cod

$>$ 385 000 entries (ready to grow $> 10^6$?)



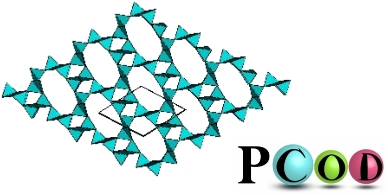
http://www.crystallography.net/tcod

$>$ 2000 entries (ready to grow to $>$ 350 000?)



http://mpod.cimav.edu.mx/

$>$ 300 entries



http://www.crystallography.net/pcod

$> 10^6$ entries (ready to grow to $> 10^8$?)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD accessibility

COD is a **fully open-access database**. All records are available under public domain designation.

Provided access methods are:

- Web search
- URLs constructed from stable identifiers
- RESTful interfaces
- Full data download

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Hyperspectral image database (HOD)

http://solsa.crystallography.net/hod

A “hybrid” approach necessary due to large size of raster data:

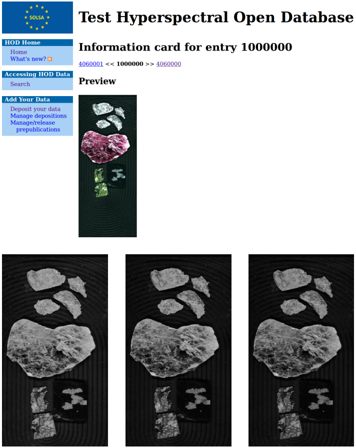
- Metadata and image headers stored in CIF;
- Raster data stored as “raw” binaries;

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

HOD record example

examples/hod/1000000-head.cif:

```
data_1000000
loop_
_[local]_description
'ENVI File'
'Created [Wed Jun 08 12:34:07 2016]'
_[local]_wavelength_units         Nanometers
loop_
_hyper_bands.default
220
227
253
_hyper_bands.lines                 937
_hyper_bands.number                288
_hyper_bands.samples               384
_hyper_file.byte_order             0
_hyper_file.data_type              4
_hyper_file.type                   ENVI_Standard
_hyper_header.offset               0
_hyper_header_file.contents
;ENVI
description = {
  ENVI File, Created [Wed Jun 08 12:34:07 2016]}
samples = 384
lines   = 937
```



Test Hyperspectral Open Database

Information card for entry 1000000

4060001 << 1000000 >> 4060000

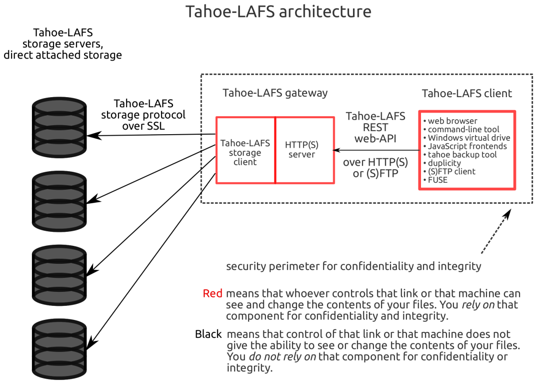
Preview

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

SOLSA Large File Store

Suitable, e.g., for images

Uses Tahoe-LAFS (https://tahoe-lafs.org) as a back-end [Selimi and Freitag, 2014]:



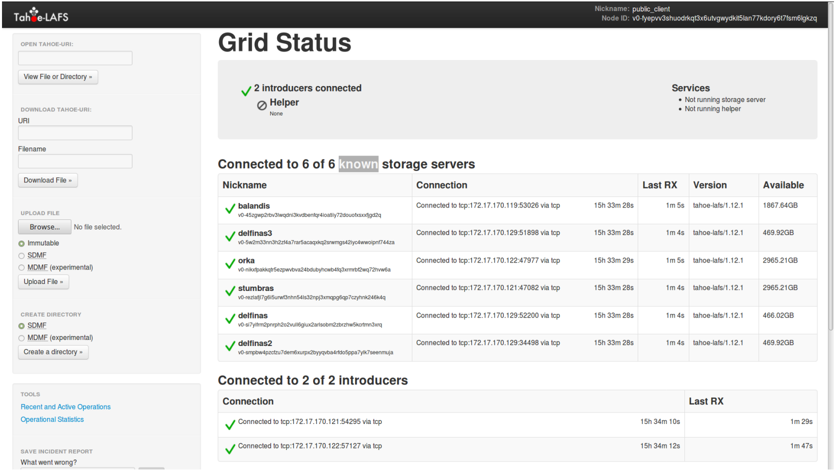
Red means that whoever controls that link or that machine can see and change the contents of your files. You *rely on* that component for confidentiality and integrity.

Black means that control of that link or that machine does not give the ability to see or change the contents of your files. You *do not rely on* that component for confidentiality or integrity.

Quoted from https://tahoe-lafs.org/trac/tahoe-lafs

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Tahoe LAFS Grid for SOLSA



Tahoe-LAFS

Nickname: public_client
Node ID: v0-fyepvv3shuodrkqt3x6utvgwydkit5lan77kdory6t7fsm6lgkzq

OPEN TAHOE-URI:

View File or Directory »

DOWNLOAD TAHOE-URI:

URI

Filename

Download File »

UPLOAD FILE

Browse... No file selected.

● Immutable
○ SDMF
○ MDMF (experimental)

Upload File »

CREATE DIRECTORY

● SDMF
○ MDMF (experimental)

Create a directory »

TOOLS

Recent and Active Operations
Operational Statistics

SAVE INCIDENT REPORT

What went wrong?

Grid Status

✓ 2 introducers connected

⊘ Helper
None

Services

- Not running storage server
- Not running helper

Connected to 6 of 6 known storage servers

Nickname	Connection		Last RX	Version	Available
✓ **balandis** v0-45zgwp2rbv3lwqdni3kvdbenfqr4ioatiiy72douofxsxxfjgd2q	Connected to tcp:172.17.170.119:53026 via tcp	15h 33m 28s	1m 5s	tahoe-lafs/1.12.1	1867.64GB
✓ **delfinas3** v0-5w2m33nn3h2zf4a7rar5acaqxkq2srwmgs42iyc4wwoipnf744za	Connected to tcp:172.17.170.129:51898 via tcp	15h 33m 28s	1m 4s	tahoe-lafs/1.12.1	469.92GB
✓ **orka** v0-nikxfpakkqtr5ezpwvbva24bdubyhcwb4fq3xrmrbf2wq72hvw6a	Connected to tcp:172.17.170.122:47977 via tcp	15h 33m 29s	1m 5s	tahoe-lafs/1.12.1	2965.21GB
✓ **stumbras** v0-rezlafjl7g6i5urwf3nhn54ls32npj3xmqpg6qp7czyhnk246k4q	Connected to tcp:172.17.170.121:47082 via tcp	15h 33m 28s	1m 4s	tahoe-lafs/1.12.1	2965.21GB
✓ **delfinas** v0-si7yifrm2pnrph2o2vull6giux2arlsobm2zbrzhw5kcrtmn3xrq	Connected to tcp:172.17.170.129:52200 via tcp	15h 33m 28s	1m 4s	tahoe-lafs/1.12.1	466.02GB
✓ **delfinas2** v0-smpbw4pzcfzu7dem6xurpx2byyqvba4rfdo5ppa7ylk7seenmuja	Connected to tcp:172.17.170.129:34498 via tcp	15h 33m 28s	1m 4s	tahoe-lafs/1.12.1	469.92GB

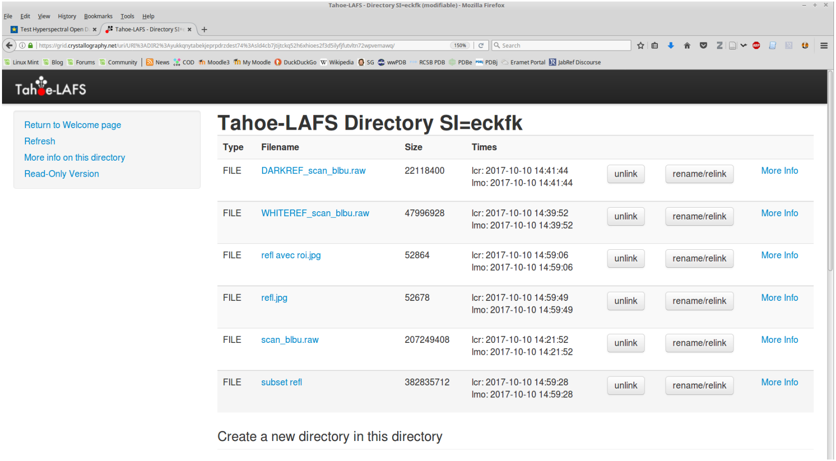
Connected to 2 of 2 introducers

Connection		Last RX
✓ Connected to tcp:172.17.170.121:54295 via tcp	15h 34m 10s	1m 29s
✓ Connected to tcp:172.17.170.122:57127 via tcp	15h 34m 12s	1m 47s

Tahoe-LAFS for SOLSA set up by Erikas Raginis

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

HOD files on the Tahoe LAFS grid



https://grid.crystallography.net/uri/URI%3ADIR2%3Ayukkqnytabekjeprpdrzdest74%3Asld4cb7jtijtckq52h6xhioes2f3d5ilyfjfutvltn72wpvemawq/

Return to Welcome page
Refresh
More info on this directory
Read-Only Version

Tahoe-LAFS Directory SI=eckfk

Type	Filename	Size	Times			
FILE	DARKREF_scan_blbu.raw	22118400	lcr: 2017-10-10 14:41:44 lmo: 2017-10-10 14:41:44	unlink	rename/relink	More Info
FILE	WHITEREF_scan_blbu.raw	47996928	lcr: 2017-10-10 14:39:52 lmo: 2017-10-10 14:39:52	unlink	rename/relink	More Info
FILE	refl avec roi.jpg	52864	lcr: 2017-10-10 14:59:06 lmo: 2017-10-10 14:59:06	unlink	rename/relink	More Info
FILE	refl.jpg	52678	lcr: 2017-10-10 14:59:49 lmo: 2017-10-10 14:59:49	unlink	rename/relink	More Info
FILE	scan_blbu.raw	207249408	lcr: 2017-10-10 14:21:52 lmo: 2017-10-10 14:21:52	unlink	rename/relink	More Info
FILE	subset refl	382835712	lcr: 2017-10-10 14:59:28 lmo: 2017-10-10 14:59:28	unlink	rename/relink	More Info

Create a new directory in this directory

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

HOD (large) data retention policy

A managed data phase-out policy possible:

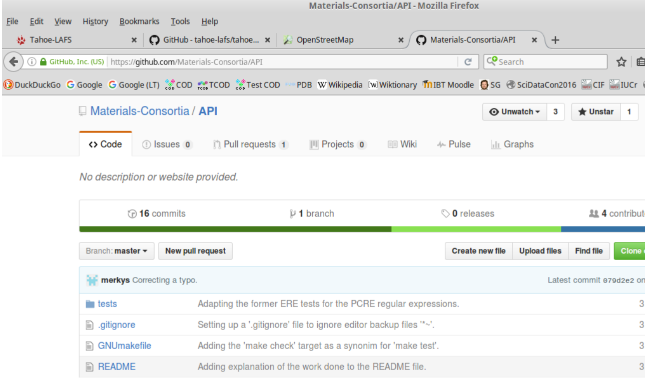
- Keep data that are:
  - The first of their kind;
  - The best of their kind;
  - The most often used/cited;
  - A small but representative test set (for software);
- Apply lossy compression to older records (×20 fold possible)
- Discard data for other records, leave just (aggregated) metadata;

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

SOLSA

Common REST API

- Agreed upon in the 2016 Leiden CECAM workshop;
- Suitable for all structural and QM databases.


Materials-Consortia/API - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Tahoe-LAFS
GitHub - tahoe-lafs/tahoe...
OpenStreetMap
Materials-Consortia/API
GitHub, Inc. (US) https://github.com/Materials-Consortia/API
Search
DuckDuckGo Google Google (LT) COD TCOD Test COD PDB Wikipedia Wiktionary IBT Moodle SG SciDataCon2016 CIF IUCr
Materials-Consortia / API
Unwatch 3
Unstar 1
Code
Issues 0
Pull requests 1
Projects 0
Wiki
Pulse
Graphs
16 commits
1 branch
0 releases
4 contribut
Branch: master
New pull request
Create new file
Upload files
Find file
Clone

No description or website provided.

merkys Correcting a typo. Latest commit 079d2e2 on

tests	Adapting the former ERE tests for the PCRE regular expressions.	3
.gitignore	Setting up a '.gitignore' file to ignore editor backup files '*~'.	3
GNUmakefile	Adding the 'make check' target as a synonim for 'make test'.	3
README	Adding explanation of the work done to the README file.	3

https://github.com/Materials-Consortia/API

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Definitions of input and output

```
(* The top-level ’filter’ rule: *)
Filter = Keyword, Expression;
(* Keywords *)
Keyword = "filter=" ;
(* Values *)
Value = Identifier | Number | String ;
(* ... some token definitions skipped for brevity ... *)
(* Expressions *)
Expression = Term, [Spaces], [ OR, [Spaces], Expression ] ;
Term = Comparison, [Spaces], [ AND, [Spaces], Term ] ;
(* Operator Comparison operator tokens: *)
Operator = ’<’, [ ’=’ ] | ’>’, [ ’=’ ] | ’=’ | ’!’, ’=’ ;
Comparison = Value, [Spaces], Operator, [Spaces], Value |
             NOT, [Spaces], Comparison |
             ’(’, [Spaces], Expression, [Spaces], ’)’ ;
```

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 689868.

Schemas for return data

Schemas allow to:

- formally agree on what is right and wrong;
- validate program outputs and documents automatically.

```
"query": {
    "type": "object",
    "properties": {
        "representation": { "type": "string" },
        "api_version": { "type": "string" },
        "time_stamp": { "type": "string" },
        "data_returned": { "type": "integer" },
        "data_available": { "type": "integer" },
        "last_id": { "type": "string" }
    },
    "required": [ "representation", "api_version",
                  "time_stamp" ]
},
```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

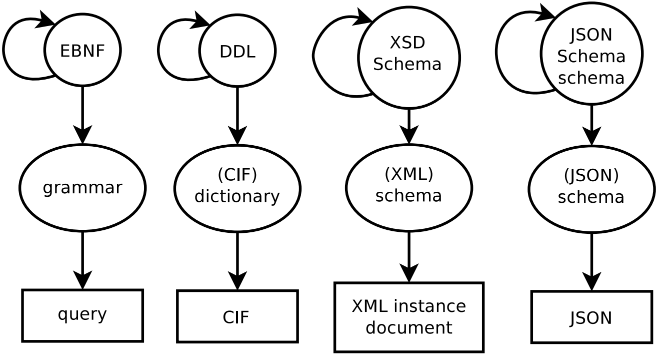
API query examples

http://crystallography.net/cod/optimade/structures?filter=elements="Si,O"ANDnelements=2&limit=1

```
{
  "resource": {
    "base_url": "http://www.crystallography.net/cod/optimade/v1.0.0-alpha.1/"
  },
  "query": {
    "api_version": "v1.0.0-alpha.1",
    "data_returned": 1,
    "representation": "/structures?filter=elements=\"Si,O\"ANDnelements=2&limit=1",
    "last_id": "1010921",
    "time_stamp": "2017-04-06T05:46:50Z",
    "implementation": {
      "maintainer": {
        "email": "cod-bugs@ibt.lt"
      },
      "title": "Crystallography␣Open␣Database",
      "version": "v1.0.0-alpha.11",
      "source_url": "svn://crystallography.net/cod/trunk/cod/cgi-bin/optimade.pl@194653"
    },
    "data_available": 344
  },
  "data": [
    {
      "last_modified": "2017-02-28T05:33:56Z",
      "properties": {
        "formula": "O2␣Si"
      },
      "url": "http://www.crystallography.net/cod/1010921.cif",
      "immutable_id": "http://www.crystallography.net/cod/1010921.cif@130149",
```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Common pattern of self-describing data definitions



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.