

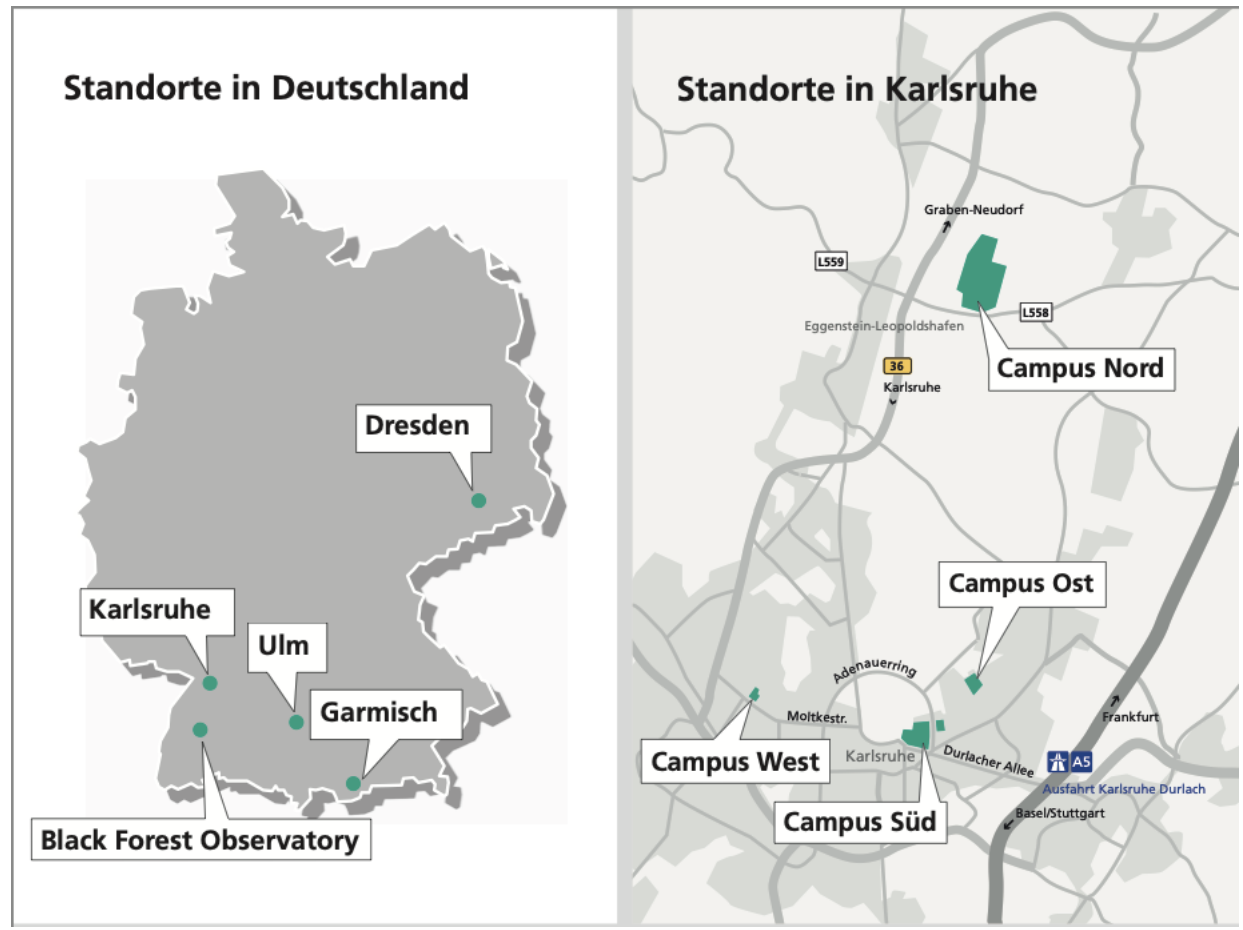
A photograph of a server room with rows of server racks. The racks are filled with server units, many of which have glowing green lights. The room is dimly lit, with a bright light source visible on the left side, creating a strong contrast and casting long shadows. The floor is covered with a green grid pattern.

Research Data Management at Karlsruhe Institute of Technology (KIT)

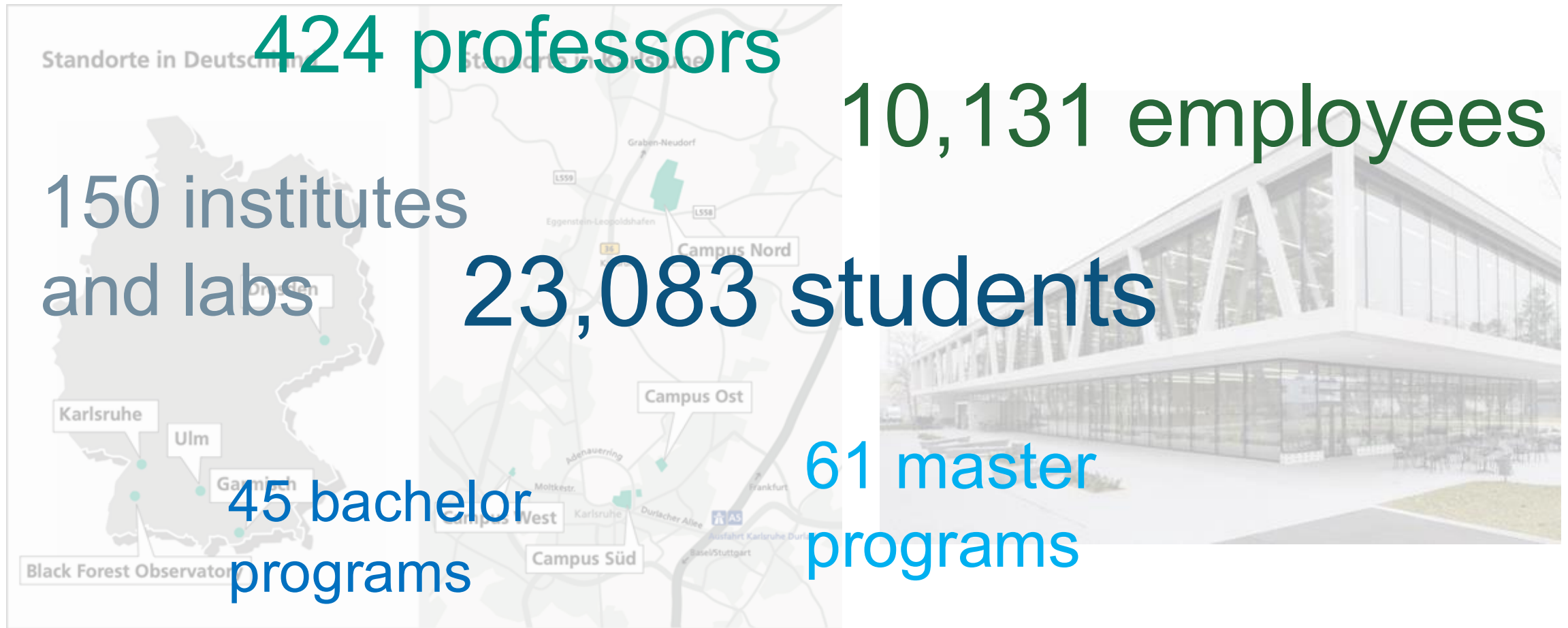
Dr. Rossella Aversa

OAVdA, 05.06.2026

Karlsruhe Institute of Technology (KIT)

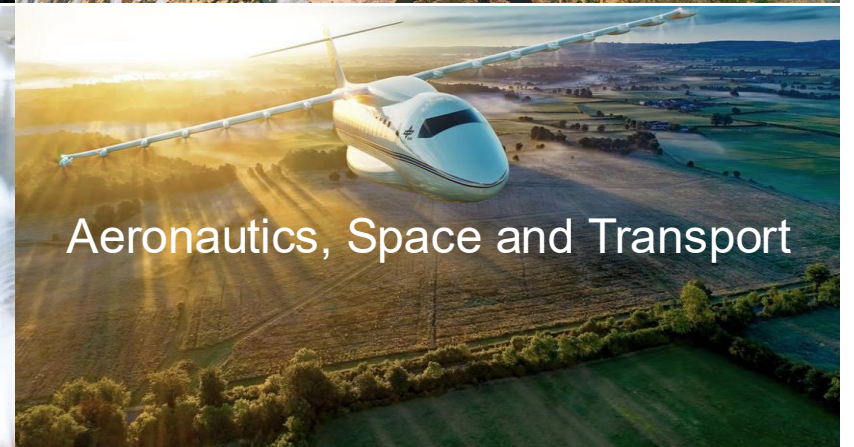
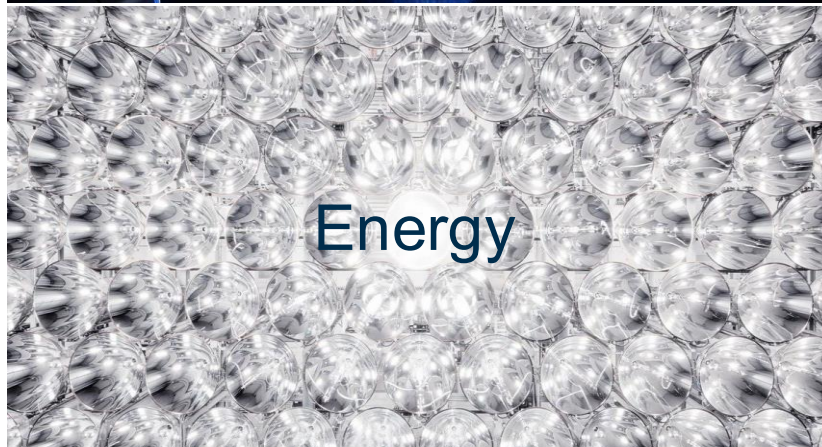
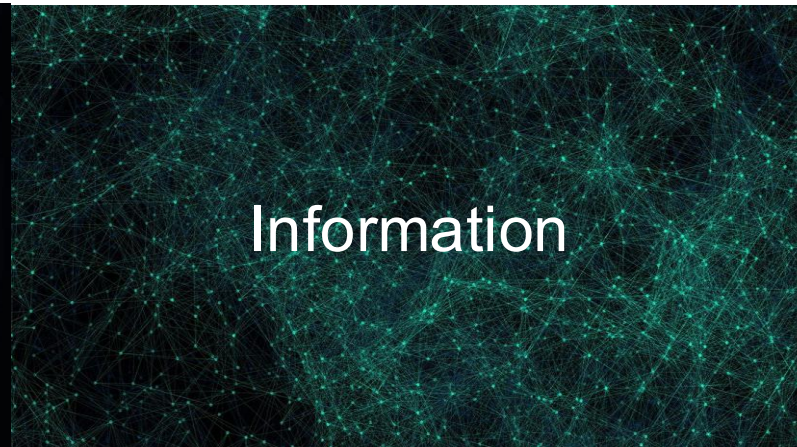
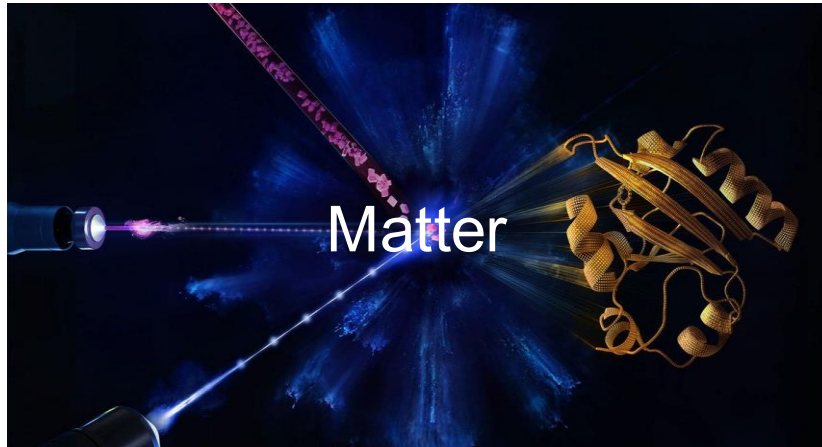


Karlsruhe Institute of Technology (KIT)



Source: [KIT website](#)

Helmholtz Association



Source: [Helmholtz Association website](#)

KIT: Scientific Computing Center (SCC)



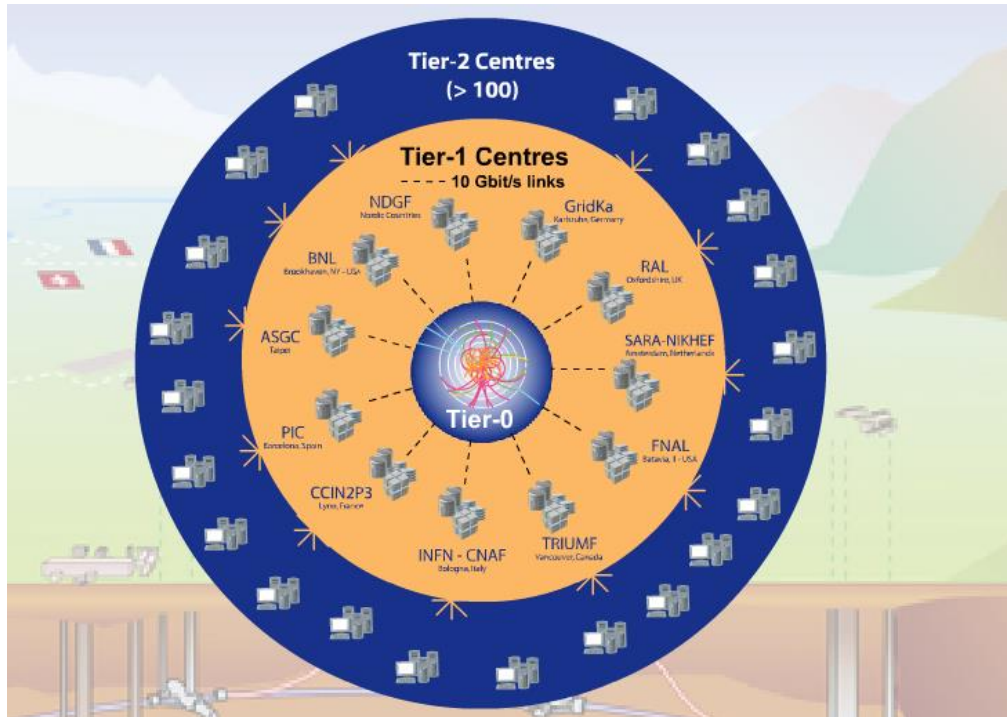
GridKa

bwUniCluster

Source: [SCC website](#)

KIT involvement: Worldwide LHC Computing Grid (WLCG)

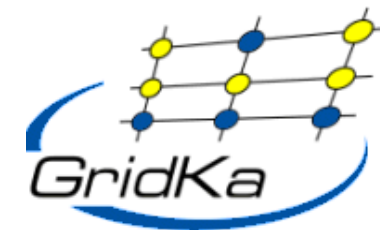
- Infrastructure for LHC data storage and analysis
- More than 170 institutions and 34 countries



Courtesy of Bob Jones (CERN)

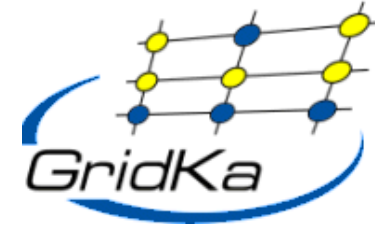


- **Tier-0** CERN
- **Tier-1** ~ 13 sites
- **Tier-2** ~ 150 sites



KIT involvement: Worldwide LHC Computing Grid (WLCG)

- Solar panels installed on GridKa data center and office buildings
- Unique in WLCG
- 500 kWp → GridKa could be powered 100% by solar energy on a sunny day

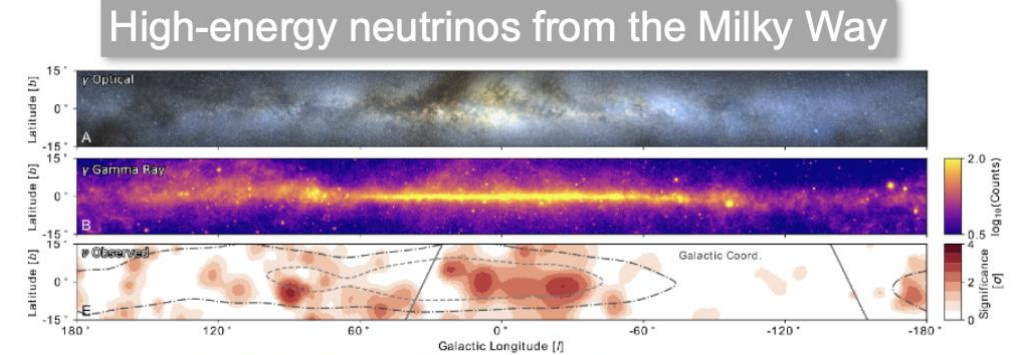


Courtesy of Achim Streit (KIT-SCC)

KIT involvement: IceCube

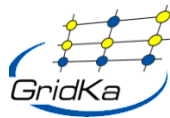


- Multi-purpose neutrino observatory at South Pole
- Indirect study of dark matter
- Cosmic ray physics
- Validation of hadronic interaction models



IceCube Collab., Science 2023

- Instrumentation (hardware, construction, installation)
- Software
- Data Analysis
- HPC Tier-1



Courtesy of Markus Roth (KIT) and Kathrin Valerius (KIT)

KIT involvement: high-energy universe

Einstein Telescope
Gravitational waves



KATRIN @ KIT
Neutrino mass



Courtesy of Andreas Haungs (KIT) and Kathrin Valerius (KIT)

Data Exploitation Methods (DEM)

Semantics

PRIMA

Alphabetical | Hierarchy

- Consumable
- Correlative Characterization
- Data Collaboration Platform
- Data Repository
- Equipment
 - Instrument
 - Sample Carrier
 - Sample Holder
- Institution
- Metadata
 - Metadata Repository
 - Persistent Identifier
 - Project
 - Publication Data
 - Dataset
 - Research Data
 - Analyzed Data
 - Processed Data
 - Raw Data
 - Reference Data
 - Research Software
 - Scientific Publication
 - Settings
 - Study
 - System
 - Technique

[PRIMA Thesaurus](#)

Schemas

Metadata Details (2b649aca-2d52-47f8-81d7-c8a1d9348bbe)

```

1- {
2-   "labelProperties": [
3-     {
4-       "labelDescription": "Biological SEM
5-       "images": ,
6-       "labelTerm": {
7-         "term": "Biology",
8-         "termDefinition": "https://vocabularie
9-         .unesco.org/browser/rest/v1/thesauru
10-        /data?uri=http%3A%2F%2Fvocabularies
11-        .unesco
12-        .org%2Fthesaurus%2Fconcept%2198Format
13-        =application%2Fjson",
14-        "termValueType": "string"
15-       }
16-     },
17-     {
18-       "imageLabelType": "tag",
19-       "correspondingResource": "sha512sum:
20-       c111982f51dd90a09afee757abb77747597db1
21-       b5f95cc4e7b69ab7f25e9d7d96ca76e40a778
22-       7fc43ad5efc4a0f7e06c9b991eaf6c8a3f3e39
23-       e9b7e4acc87"
24-     }
25-   ]
26- }
    
```

Metadata

Schema: midata_basic_schema v1

Hash: sha171a127387e390d11918a2b364b6790c42442

Created at: 2023-09-18 13:49:19

Last Modified: 2023-09-18 13:49:19

Downloads

JSON

[ML data basic schema](#)

Graphs

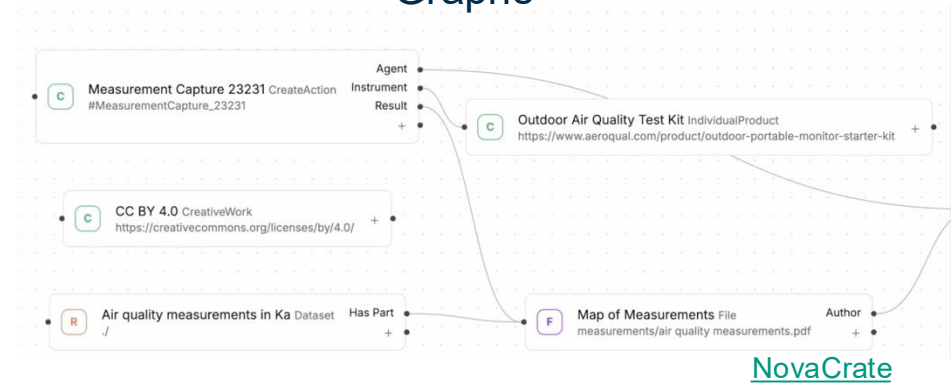
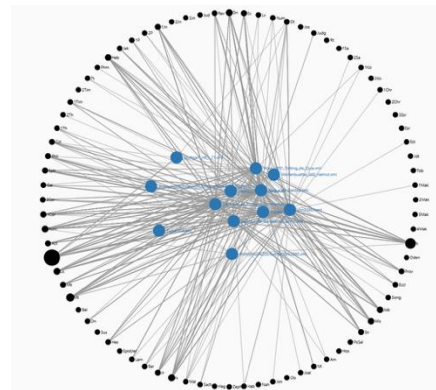


Image segmentation



[ToRoll](#)

Data Analysis



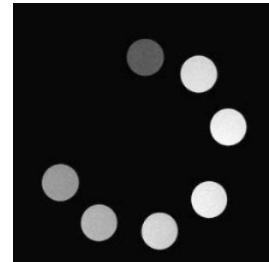
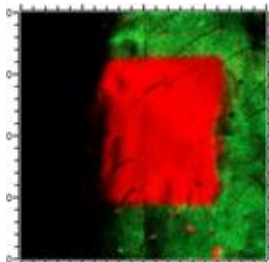
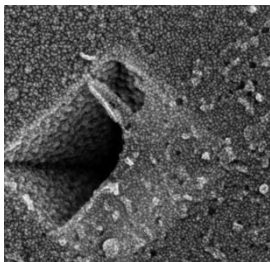
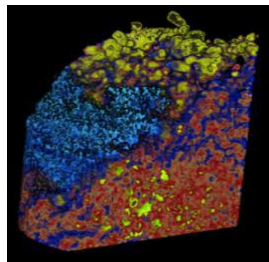
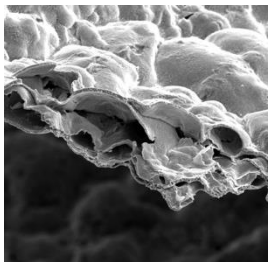
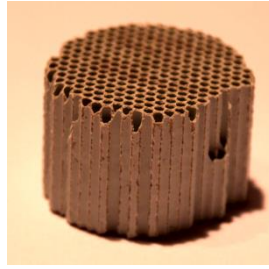
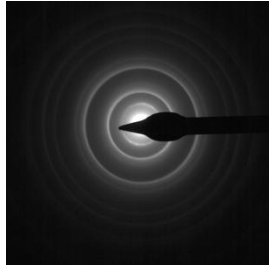
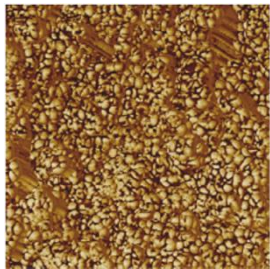
Preliminary

RDM software

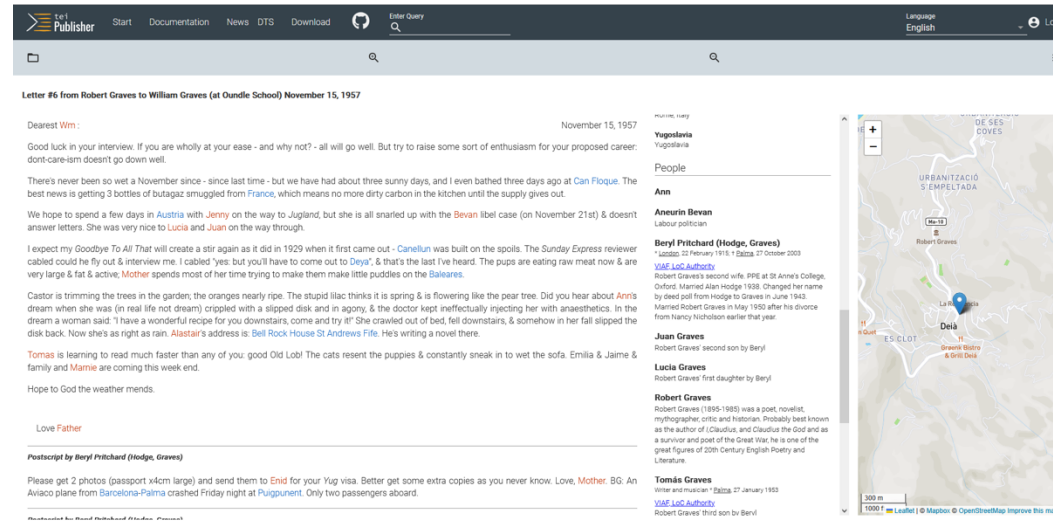
[Typed PID Maker](#)

Current DEM Applications

Materials Science and Nanoscience



Digital Humanities



Letter #6 from Robert Graves to William Graves (at Oundle School) November 15, 1957

November 15, 1957

Dearest Wm:

Good luck in your interview. If you are wholly at your ease - and why not? - all will go well. But try to raise some sort of enthusiasm for your proposed career: dont-care-ism doesn't go down well.

There's never been so wet a November since - since last time - but we have had about three sunny days, and I even bathed three days ago at [Can Floque](#). The best news is getting 3 bottles of butagaz smuggled from France, which means no more dirty carbon in the kitchen until the supply gives out.

We hope to spend a few days in Austria with Jerry on the way to Jugland, but she is all snarled up with the Bevan libel case (on November 21st) & doesn't answer letters. She was very nice to Lucia and Juan on the way through.

I expect my *Goodbye To All That* will create a stir again as it did in 1929 when it first came out - *Canellun* was built on the spoils. The Sunday Express reviewer cabled could he fly out & interview me. I cabled 'yes, but you'll have to come out to Deyri', & that's the last I've heard. The pups are eating raw meat now & are very large & fat & active. Mother spends most of her time trying to make them make little puddles on the *Baleares*.

Castor is trimming the trees in the garden, the oranges nearly ripe. The stupid lilac thinks it is spring & is flowering like the pear tree. Did you hear about Ann's dream when she was (in real life not dream) crippled with a slipped disk and in agony, & the doctor kept ineffectually injecting her with anaesthetics. In the dream a woman said: 'I have a wonderful recipe for you downstairs, come and try it!' She crawled out of bed, fell downstairs, & somehow in her fall slipped the disk back. Now she's as right as rain. Alastair's address is: Bell Rock House St Andrews File. He's writing a novel there.

Tomas is learning to read much faster than any of you: good Old Lolo! The cats resent the puppies & constantly sneak in to wet the sofa. Emilia & Jaime & family and Mamie are coming this week end.

Hope to God the weather mends.

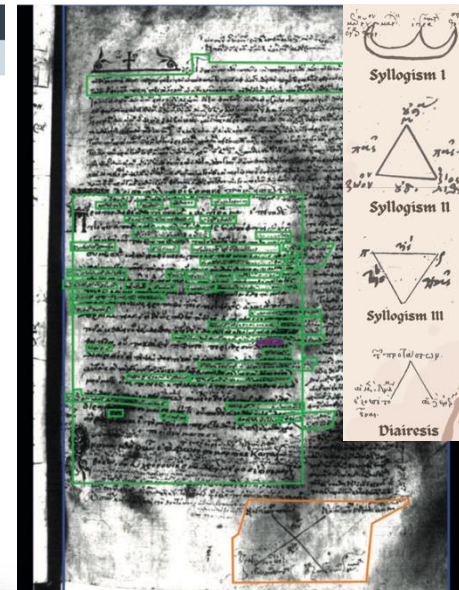
Love Father

Postscript by Beryl Pritchard (Hodge, Graves)

Please get 2 photos (passport x4cm large) and send them to Endi for your Yug visa. Better get some extra copies as you never know. Love, Mother. BG: An Aviacco plane from Barcelona-Palma crashed Friday night at Puigpugent. Only two passengers aboard.

Robert Graves (1895-1985) was a poet, novelist, mythographer, critic and historian. Probably best known as the author of *Claudius*, and *Claudius the God* and as a survivor and poet of the Great War, he is one of the great figures of 20th-Century English Poetry and Literature.

Tomas Graves
Writer and musician * (Date: 27 January 1953)
VIA LSC Authority
Robert Graves' third son by Beryl

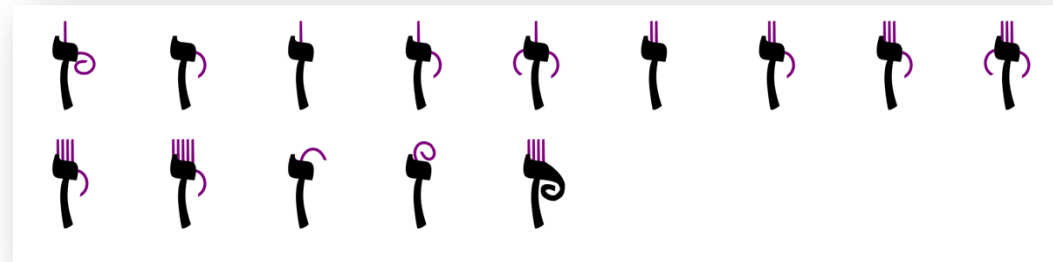


Syllogism I

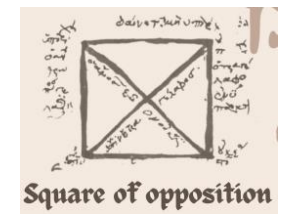
Syllogism II

Syllogism III

Diresis



Aristotele, *De interpretatione*

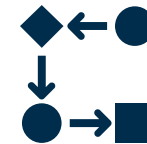


Square of opposition

Courtesy of Richard Thelen (KIT), Neil MacKinnon (KIT), Regina Ciancio (CNR-IOM), Germaine Götzelmann (KIT), Danah Tonne (KIT), Lilly Osburg (KIT)

FAIR Digital Objects

Digital Object: any identifiable piece of digital information that can be stored, managed, accessed, and exchanged by computer systems. It may represent data, software, metadata, a document, an image, a workflow, or another digital resource.

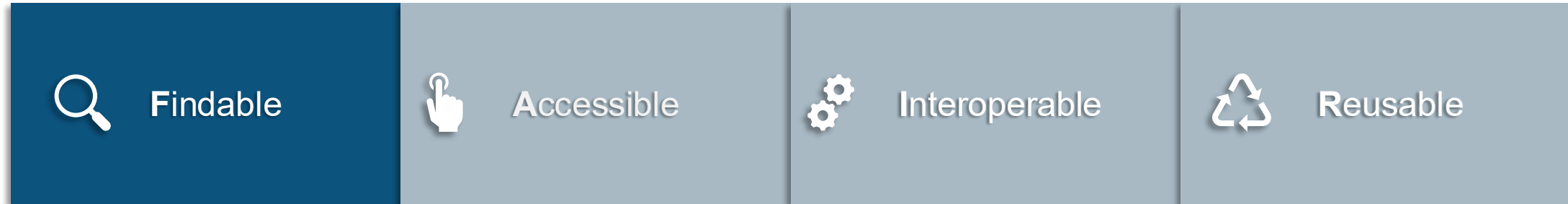


The FAIR Guiding Principles



<https://www.go-fair.org/fair-principles/>

The FAIR Guiding Principles



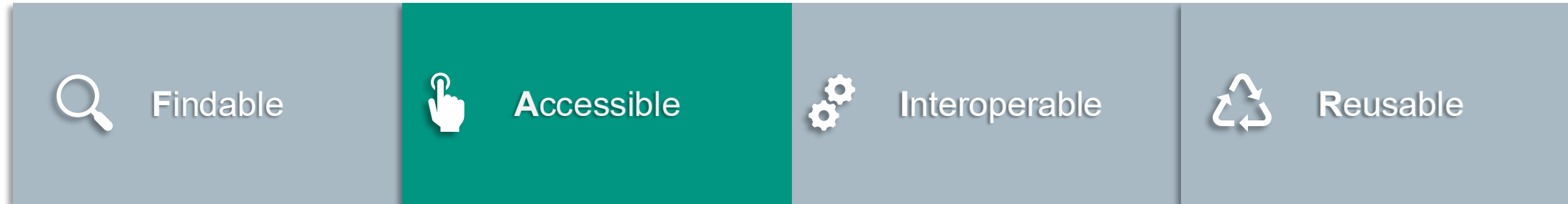
(Meta)data should be easy to find for both humans and computers



Globally unique persistent identifiers (PID)

<https://www.go-fair.org/fair-principles/>

The FAIR Guiding Principles



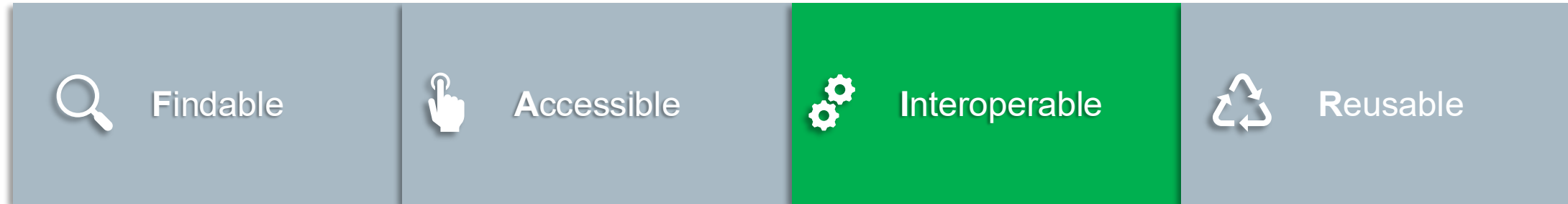
It should be known how (meta)data can be accessed



(Meta)data repositories, authorization & authentication

<https://www.go-fair.org/fair-principles/>

The FAIR Guiding Principles



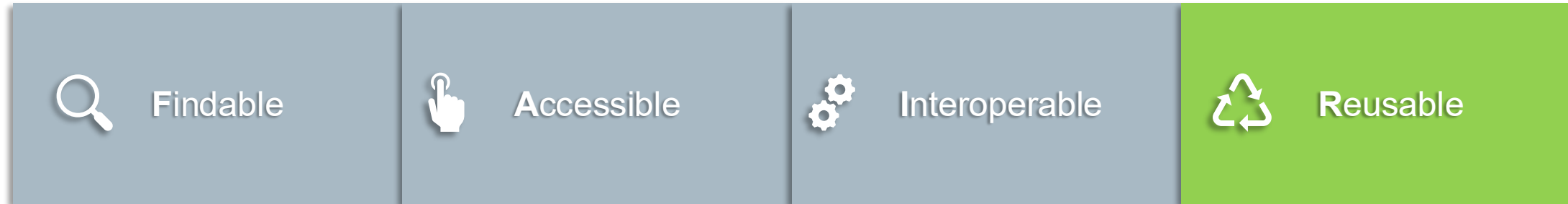
Data should be exchanged and interpreted by humans and computers



Structured metadata (schemas, vocabularies)

<https://www.go-fair.org/fair-principles/>

The FAIR Guiding Principles



It should be clear how data can be reused and/or replicated

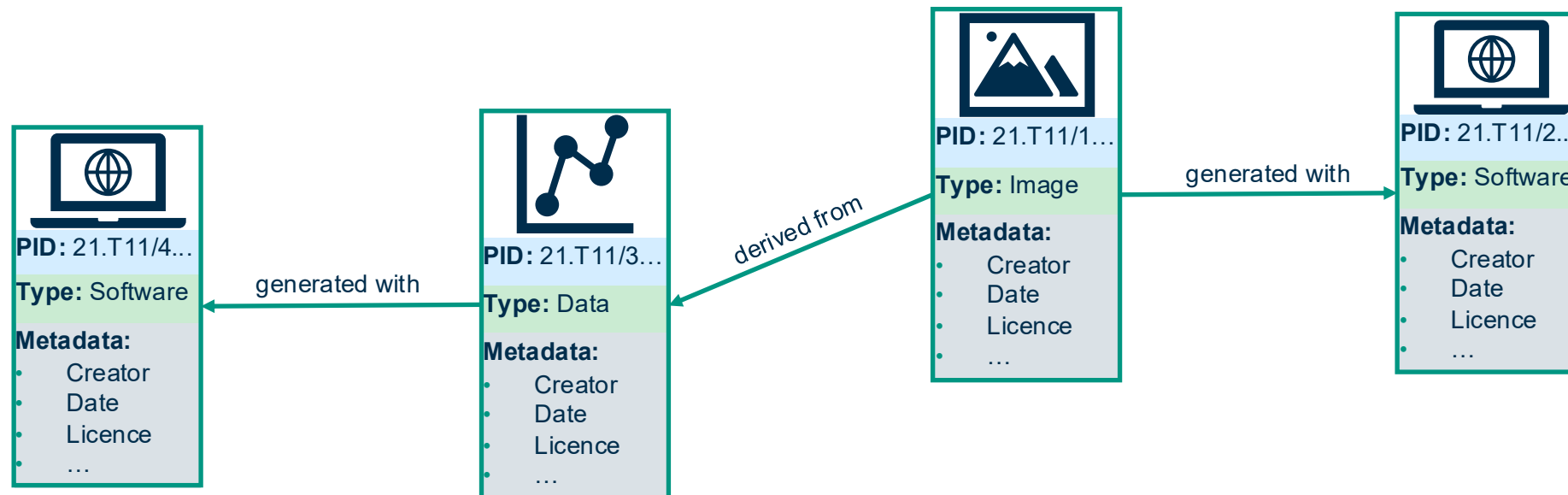


Licences, rich (provenance) metadata

<https://www.go-fair.org/fair-principles/>

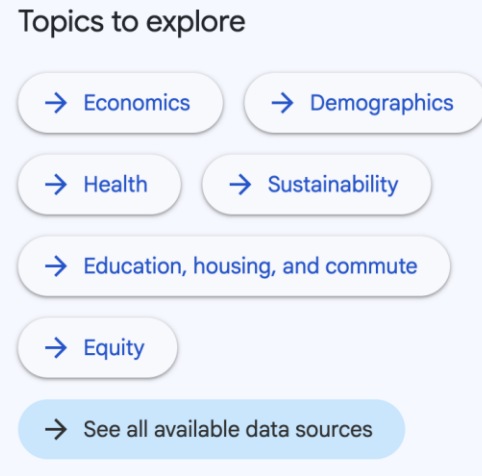
FAIR Digital Objects

FAIR Digital Object: digital object that is structured and described in a way that makes it FAIR. It contains the digital content and the essential machine-readable information.



Why are FDOs useful?

Data Commons brings together the world's public data, making it simple to explore



Data Space: framework containing a set of data sources (and their relations) regardless of their format, location, or model.

Data Commons synthesizes a single graph from these different data sources. It links **references** to the same entities (such as cities, counties, organizations, etc.) across different datasets to nodes on the graph, so that users can access data about a particular entity **aggregated from different sources without data cleaning** or joining.

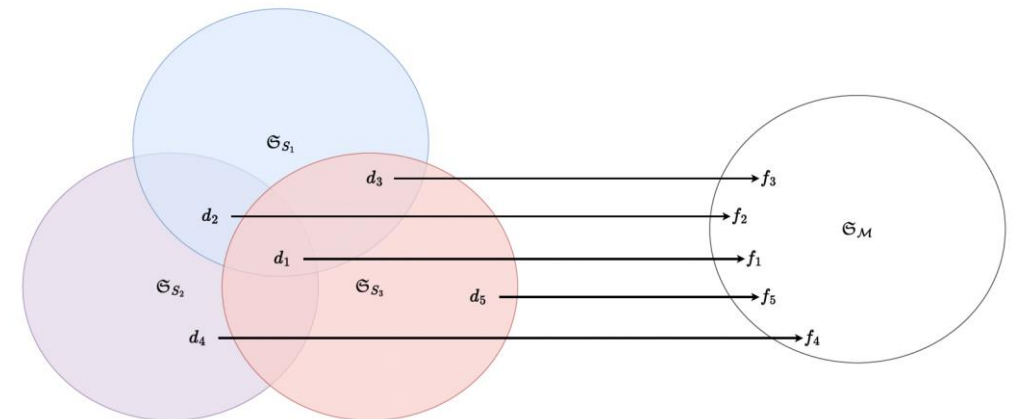


Figure 3.2.: Illustration of fully aligned FAIRified data spaces by projection to a meta-level.

Courtesy of Nicolas Blumenröhr (KIT)

What are the advantages of FDOs?

- **Standardized descriptions:** harmonised representation, alignment of different data spaces, better data reusability/reproducibility, reduced duplication of effort
- **Standardized type definitions:** easier exchange of information across systems and organizations
- **PIDs:** persistently findable, reference (pointer) independent of storage location, no effect if location changes
- **Lightweight layer:** retrieve data from multiple repositories without touching or migrating the data
- **Customizable:** can include/link additional info which is not included in the specific repository
- **Machine-actionable:** machine-readable format, automated workflows/data pipelines, AI-ready without requiring manual intervention (at least quicker and more straightforward, no need of multiple APIs)
- **Operation FDOs:** set to act on specific target FDOs (according to their type)



Example: NMR Search



Nuclear Magnetic Resonance Metadata Query

NFFA-Europe Data Management and Virtual Access Service

Rossella Aversa

Search for something...

Search

Resource Type

- Dataset 21899
- Study 5174
- Project 58
- Repository 2

Source

- chemotion-repository.net 22990
- nmriv.org 4141

NMR Method

- ¹H nuclear magnetic resonance spectroscopy 4116
- ¹³C nuclear magnetic resonance spectroscopy 3559
- infrared absorption spectroscopy 3137
- distortionless enhancement with polarization transfer 1836
- mass spectrometry 1752

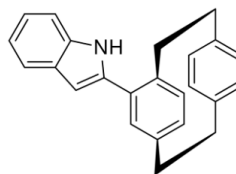
+ Show more

NMR Solvent

- dimethyl sulfoxide-d6 610
- dideuterium oxide 365
- deuterated chloroform 128

Showing 1 -20 out of 10000

↑↓ Name (ascending)



(rac)-2-([2.2]paracyclophan-4-yl)-1H-indole

CRS-39445 + 21/05/2024

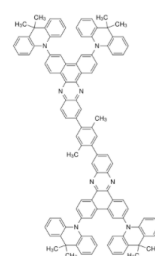
21.T11981/c8319d84-0835-4f26-aa0d-6262c203c4dc

Tilman Köhler Study chemotion-repository.net AGPL-3.0

323.4302399999986 g/mol

Show Related Items 8

Open



11,11'-(2,5-dimethyl-1,4-phenylene)bis(3,6-bis(9,9-dimethylacridin-10(9H)-yl)dibenzo[a,c]phenazine)

CRS-46008 + 25/11/2024

21.T11981/219f2a17-8b21-4110-bfce-97e06474a596

Jason Hofmann Study chemotion-repository.net AGPL-3.0

1491.860280000001 g/mol

Show Related Items 10

Open



3-(3-hydroxypropyl)-4-(3-methoxyprop-1-yn-1-yl)-6-(trifluoromethyl)-1H-isochromen-1-one

CRS-33628 + 02/06/2023

21.T11981/2782486c-968e-4e8d-8456-c69043d7e1e3

[NMR Search](#)

Example: NMR Search

Nuclear Magnetic Resonance

NFFA-Europe Data Management and Virtual

Search for something...

Resource Type

- Dataset 21899
- Study 5174
- Project 58
- Repository 2

Source

- chemotion-repository.net 22990
- nmrXiv.org 4141

NMR Method

- 1H nuclear magnetic resonance spectroscopy 4116
- 13C nuclear magnetic resonance spectroscopy 3559
- infrared absorption spectroscopy 3137
- distortionless enhancement with polarization transfer 4836
- mass spectrometry 1752
- [+ Show more](#)

NMR Solvent

- dimethyl sulfoxide-d6 610
- dideuterium oxide 365
- deuterated chloroform 128

Chemotion-Repository

Reactions | Samples | Scheme-only reactions

ID	Embargo	Author	Published on	Sample
CRS-61129	LB_2025-09-05	Lisa Burkart	2026-05-29	-
CRS-63954	-	Simone Gräßle	2026-05-26	-
CRS-63962	-	Simone Gräßle	2026-05-26	-

Sample Published on 2024-05-21

Molecule/Material name: (rac)-2-[(2,2)paracyclophan-4-yl]-1H-indole
 Contributor: Tilman Hans Köhler¹
 Author: Tilman Hans Köhler¹
¹. Institute of Organic Chemistry, Karlsruhe Institute of Technology, Germany

nmrXiv

Cenaptnmr

DOI: 10.57992/nmrXiv.p33

License: Creative Commons Zero v1.0 Universal (CC0 1.0) | Published: December 21, 2023 | Created: August 29, 2022

Pan-assay interference compounds (PAINS)

Info | Samples | Files | License

Cite this sample | Format: APA

Bisson, J., Kim, S. B., Chen, S.-N., Simmler, C., Ray, G. J., Youn, I., Pauli, G. F., Lankin, D. C., & Achanta, P. S. (2023). Eucalyptol 400 MHz in CDCl3 NMR data [Dataset]. nmrXiv. <https://doi.org/10.57992/NMRXIV.P33.S218>

Eucalyptol 400 MHz in CDCl3 NMR data

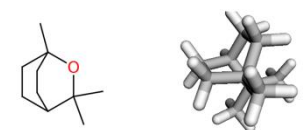
DOI: 10.57992/nmrXiv.p33.s218

Description: Eucalyptol 400 MHz in CDCl3 NMR data

Sample

Molecular Composition

InChI=1S/C10H18O/c1-9(2)8-4-6-10(3,11-9)7-5-8/h8H,4-7H2,1-3H3



NMR Search

Example: NMR Search



Nuclear Magnetic Resonance Metadata Query

NFFA-Europe Data Management and Virtual Access Service

Rossella Aversa

Search for something...

Search

Resource Type

- Dataset 21899
- Study 5174
- Project 58
- Repository 2

Source

- chemotion-repository.net 22990
- nmriv.org 4141

NMR Method

- ¹H nuclear magnetic resonance spectroscopy 4116
- ¹³C nuclear magnetic resonance spectroscopy 3559
- infrared absorption spectroscopy 3137
- distortionless enhancement with polarization transfer 1836
- mass spectrometry 1752

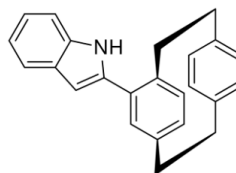
+ Show more

NMR Solvent

- dimethyl sulfoxide-d6 610
- dideuterium oxide 365
- deuterated chloroform 128

Showing 1 -20 out of 10000

↑↓ Name (ascending)



(rac)-2-([2.2]paracyclophan-4-yl)-1H-indole

CRS-39445 + 21/05/2024

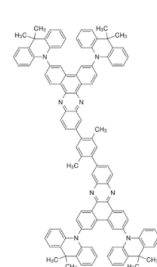
21.T11981/c8319d84-0835-4f26-aa0d-6262c203c4dc

Tilman Köhler Study chemotion-repository.net AGPL-3.0

323.4302399999986 g/mol

Show Related Items 8

Open



11,11'-(2,5-dimethyl-1,4-phenylene)bis(3,6-bis(9,9-dimethylacridin-10(9H)-yl)dibenzo[a,c]phenazine)

CRS-46008 + 25/11/2024

21.T11981/219f2a17-8b21-4110-bfce-97e06474a596

Jason Hofmann Study chemotion-repository.net AGPL-3.0

1491.860280000001 g/mol

Show Related Items 10

Open



3-(3-hydroxypropyl)-4-(3-methoxyprop-1-yn-1-yl)-6-(trifluoromethyl)-1H-isochromen-1-one

CRS-33628 + 02/06/2023

21.T11981/2782486c-968e-4e8d-8456-c69043d7e1e3

[NMR Search](#)

Example: NMR Search

Inspect Result

21.T11981/c8319d84-0835-4f26-aa0d-6262c203c4dc [Copy](#)

Key	Value
contact	^ ID Köhler, Tilman (Karlsruhe Institute of Technology)
KernelInformationProfile	^ 21.T11148/b9b76f887845e32d29f7
digitalObjectType	^ 21.T11148/ca9fd0b2414177b79ac2
hadPrimarySource	^ 21.T11981/935ad20c-e8f7-485d-8987-b4f22431ff4b
isMetadataFor	^ 21.T11981/cf1b03f0-3074-4a9e-9da8-2238734a4fda

Showing 1 to 10 of 21 entries

[Open](#)

(rac)-2-((2,2)paracyclophan-4-yl)-1H-indole
CRS-38443 + 21052024
21.T11981/c8319d84-0835-4f26-aa0d-6262c203c4dc
Tilman Köhler Study chemotion-repository.net ADPL-3.0
323.4302399999986 g/mol

infrared absorption spectroscopy (IR)
CRD-38443 + 21052024
21.T11981/c31ee95-5c18-497a-ac6b-ce3783428bb2
Tilman Köhler Dataset chemotion-repository.net
CC-BY-SA-4.0 323.4302399999986 g/mol
infrared absorption spectroscopy

heteronuclear single quantum coherence (HSQC)
CRD-38443 + 21052024
21.T11981/c169c08f-c47d-4809-6781-3ab184e08771
Tilman Köhler Dataset chemotion-repository.net
CC-BY-SA-4.0 323.4302399999986 g/mol
heteronuclear single quantum coherence

electron ionisation mass spectrometry (EI-MS)
CRD-38441 + 21052024
21.T11981/f8c204b-8202-45ed-07ee-14ed4885115
Tilman Köhler Dataset chemotion-repository.net
CC-BY-SA-4.0 323.4302399999986 g/mol
electron ionisation mass spectrometry

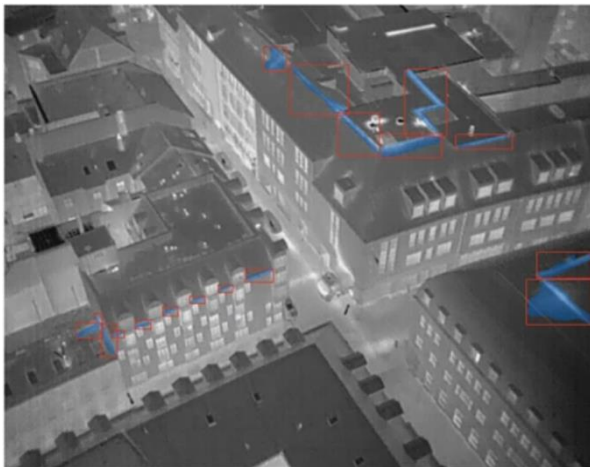
deuterated chloroform 128

CRS-33628 + 02/06/2023
21.T11981/2782486c-968e-4e8d-8456-c69043d7e1e3

[Open](#)

[NMR Search](#)

Example: Thermal Bridges on Building Rooftops



PID Information Record

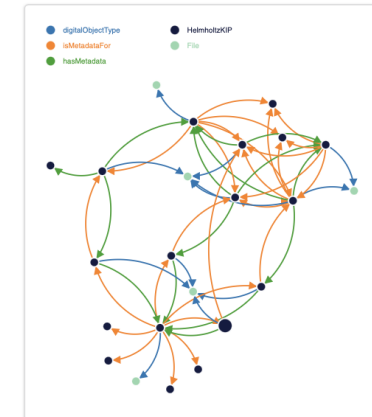
Type	Value
kernelInformationProfile	21.T11148/b9b76f887845e32d29f7
dateModified	2022-08-26T00:00:00+00:00
checksum	{ "sha512sum": "c51b5782d7f7b07607f7b8c
dateCreated	2022-05-30T00:00:00+00:00
digitalObjectLocation	https://zenodo.org/record/6517768/files/Flu
version	1.0.0
digitalObjectType	21.11152/3052e4f1-4dd6-4c40-9a16-0a6b8
licenseUrl	https://creativecommons.org/licenses/by/4.0/
hasMetadata	21.11152/ba370aa3-6422-428c-9ff7-c2ef4
isMetadataFor	21.11152/365fd8cf-8e86-41b8-9d0e-b816f
topic	http://vocabularies.unesco.org/thesaurus/cc
contact	https://orcid.org/0000-0002-9082-9095

FDO Badge

FAIR DO 21.11152/eabeb5bc-0514-47c9-bcd2-98f0253843d8

Copy as

FAIR DO Graph



PID Component

Copy Code

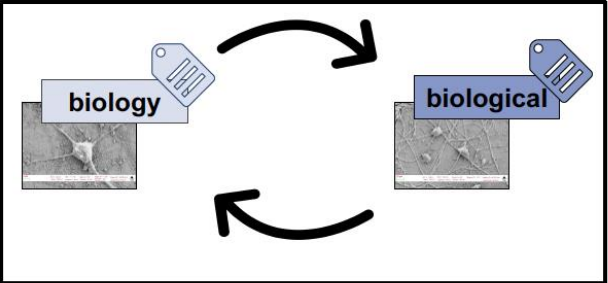
[TBBR Dataset](#)

Example: Relabeling of SEM training datasets

 Data sets with different label terms are not directly reusable for Machine Learning (ML) applications.

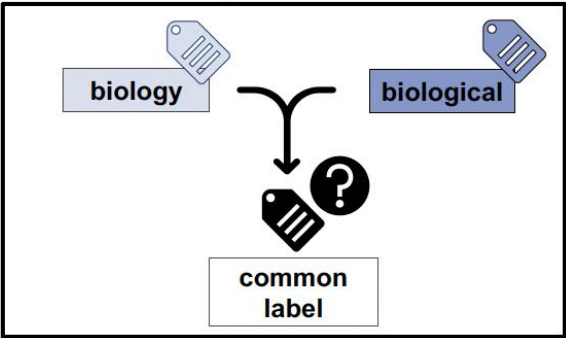
 Automated relabelling is enabled by using a vocabulary-based schema and FAIR Digital Objects (FAIR DOs).

Usually different label terms are used to describe the same category, i.e. have the same meaning.

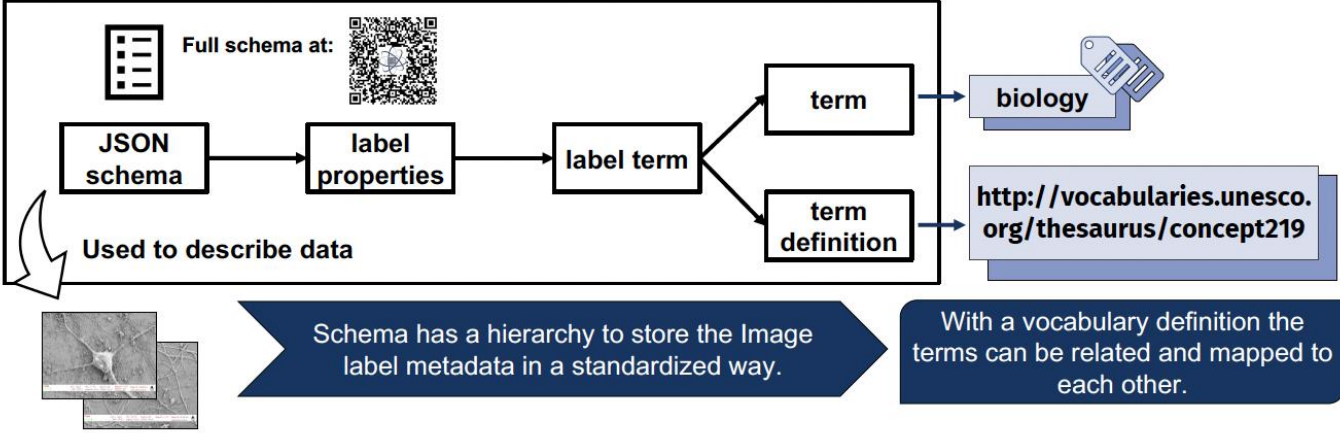


Example with *Scanning Electron Microscopy (SEM) image data. Provided by R. Aversa et. al. <http://doi.org/10.23728/b2share.19cc2afd23e34b92b36a1dfd0113a89f>

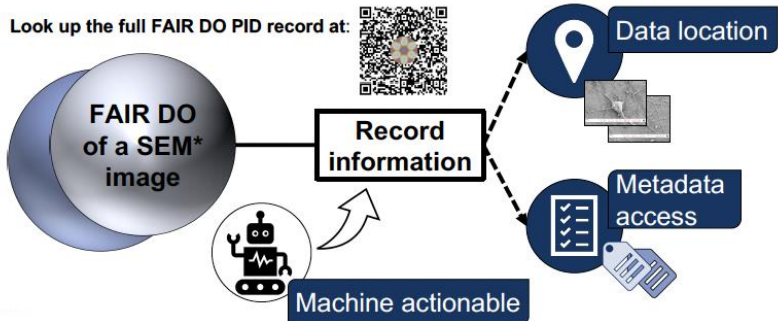
Labels can be mapped to each other, i.e. relabelled.



Metadata Schema for Label Information



FAIR DO for Machine Actionability



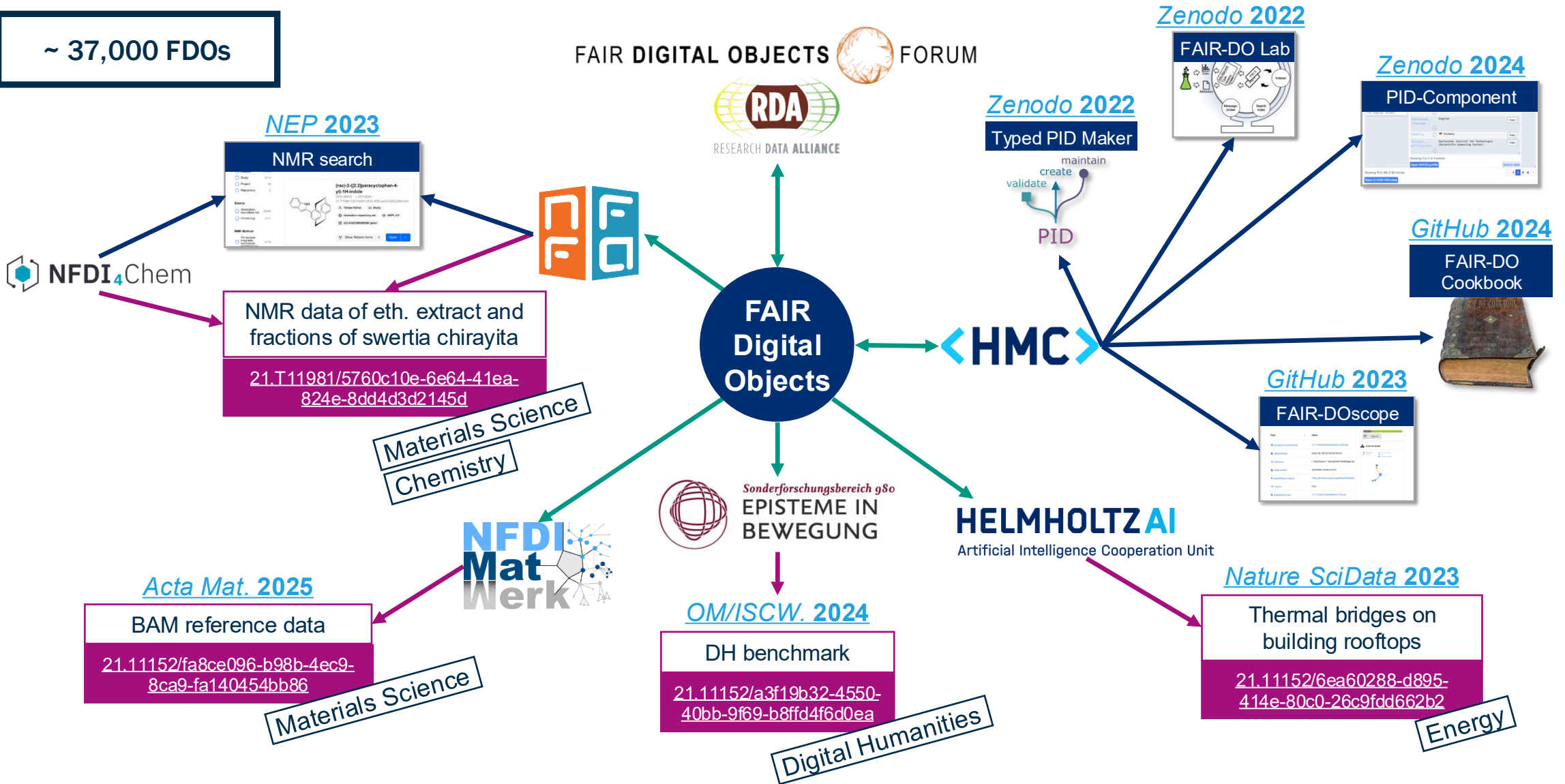
Conclusions

- Using a metadata schema and vocabularies for data description provides standardization.
- The record information of a FAIR DO contains all information required to access the data and its metadata.
- Machines can evaluate the information record and carry out the relabelling using that information.

FAIR Digital Objects

~ 37,000 FDOs

FAIR DIGITAL OBJECTS FORUM

Looking for other applications

- **Data explorer:** overview of data (different repos, representations, infos)
 - important information can be harmonized
 - extra metadata (e.g., telescope/db-specific) in a different record for different actions
- **Interactive interfaces:** based on geolocation, portion of the sky, ...
- Possible **operations** for machine-actionable datasets:
 - Spot if something is under embargo, for machine-driven decisions
 - Check for specific metadata in the record (because they are needed for training)
 - Show preview image
 - ...

We can provide:

- assistance for modelling FDOs
- tools for creating FDOs
- specific visualisations (we have motivated experts)



CMB
~~Sky~~ is the limit