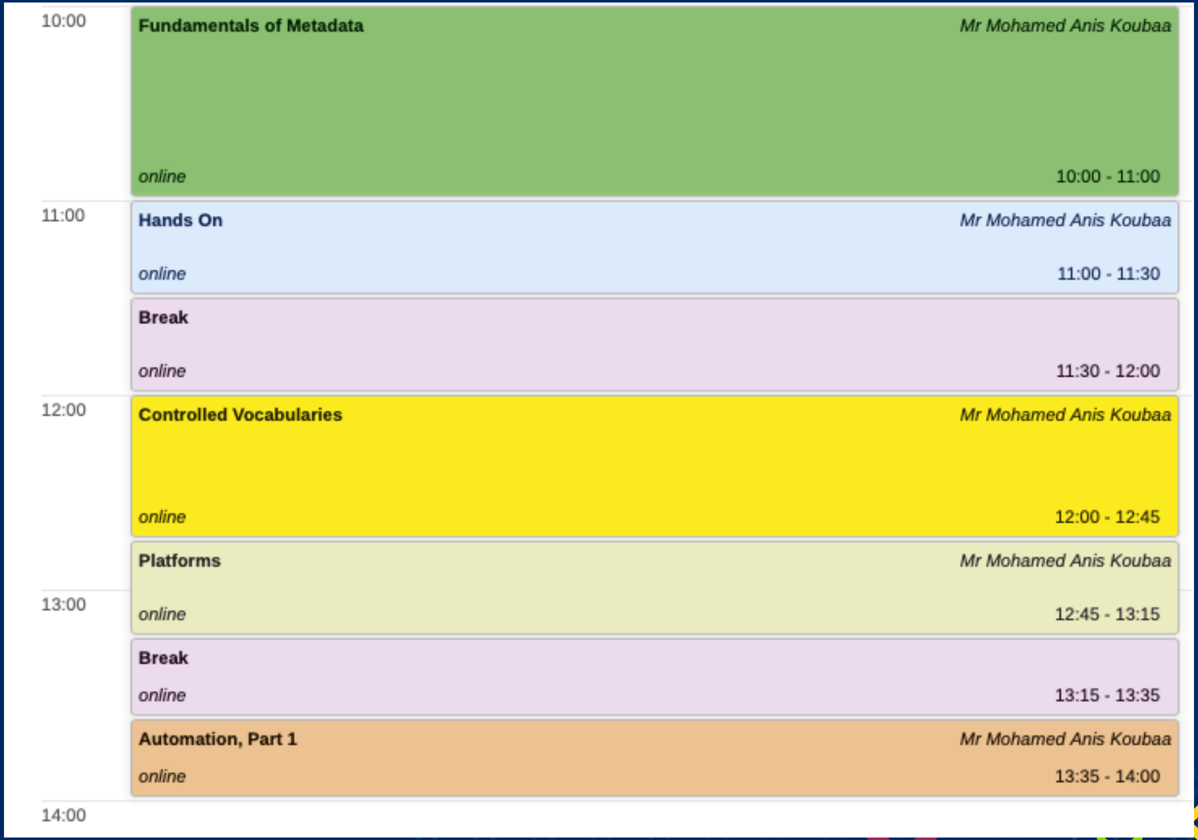
Fundamentals of Scientific Metadata for Energy

28 & 29-04-2025
Anis Koubaa



10:00

Fundamentals of Metadata *Mr Mohamed Anis Koubaa*

online 10:00 - 11:00

11:00

Hands On *Mr Mohamed Anis Koubaa*

online 11:00 - 11:30

Break

online 11:30 - 12:00

12:00

Controlled Vocabularies *Mr Mohamed Anis Koubaa*

online 12:00 - 12:45

Platforms *Mr Mohamed Anis Koubaa*

13:00

online 12:45 - 13:15

Break

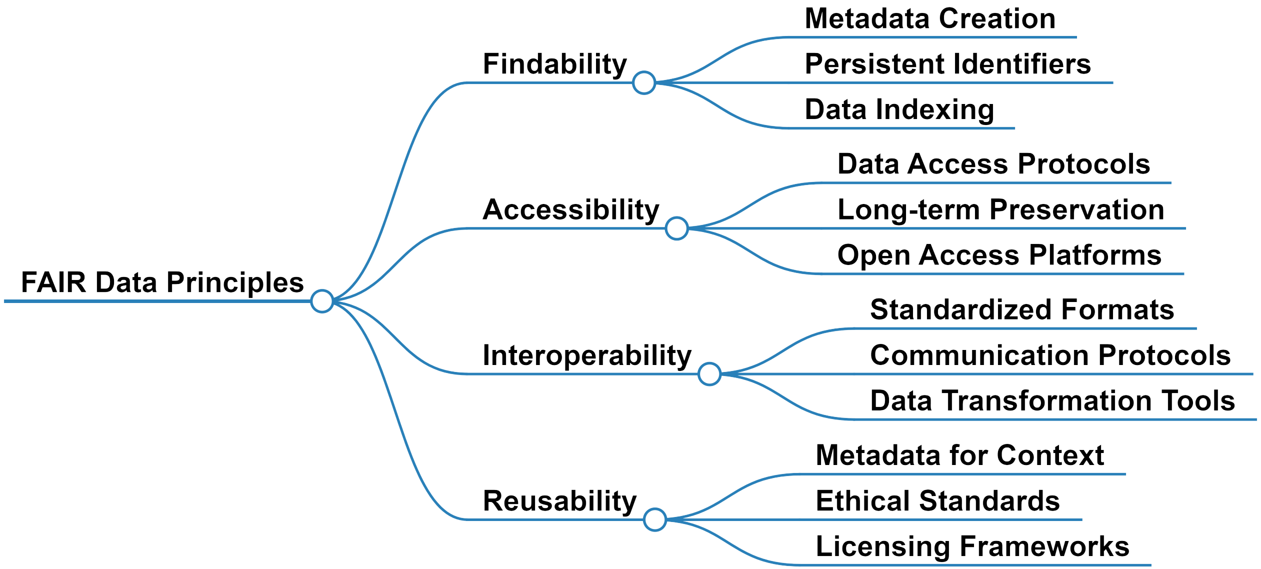
online 13:15 - 13:35

Automation, Part 1 *Mr Mohamed Anis Koubaa*

online 13:35 - 14:00

14:00

Energy Hub @ HMC
Metadata and FAIRness



- FAIR Data Principles
  - Findability
    - Metadata Creation
    - Persistent Identifiers
    - Data Indexing
  - Accessibility
    - Data Access Protocols
    - Long-term Preservation
    - Open Access Platforms
  - Interoperability
    - Standardized Formats
    - Communication Protocols
    - Data Transformation Tools
  - Reusability
    - Metadata for Context
    - Ethical Standards
    - Licensing Frameworks

Metadata describe the data and are critical to helping users discover relevant datasets. “We love rich metadata,” says LeMay. “We want to know who made the data, where it was made, what it contains, who to credit, how to reference the dataset.”

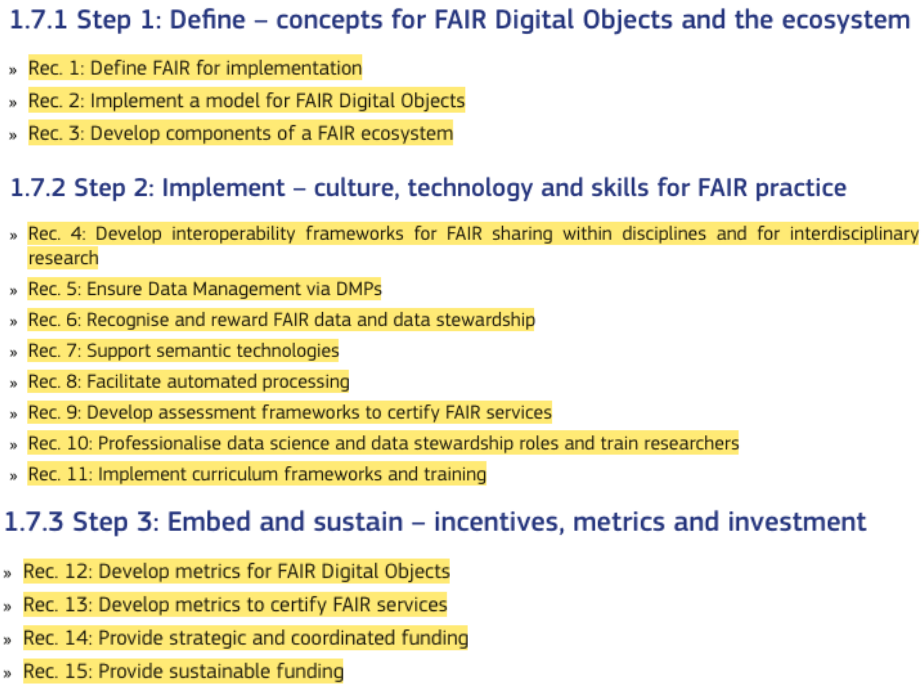
Metadata can also include keywords, field of science classification codes, the DOIs of related papers, the researchers’ ORCID identifiers, and the codes for the grants that supported the research.

NI 360 · 11 FEBRUARY 2019

"A love letter to your future self": What scientists need to know about FAIR data

Energy Hub @ HMC

- Support scientific teams in reaching higher levels of FAIRness:
  - Develop solutions for facilitating data description, by making processes automatic or semi-automatic
  - Dissemination of existing community vocabularies and standards
  - Support developing controlled vocabularies and standards, based mainly on existing extra-community standards and integrating them



1.7.1 Step 1: Define – concepts for FAIR Digital Objects and the ecosystem

» Rec. 1: Define FAIR for implementation

» Rec. 2: Implement a model for FAIR Digital Objects

» Rec. 3: Develop components of a FAIR ecosystem

1.7.2 Step 2: Implement – culture, technology and skills for FAIR practice

» Rec. 4: Develop interoperability frameworks for FAIR sharing within disciplines and for interdisciplinary research

» Rec. 5: Ensure Data Management via DMPs

» Rec. 6: Recognise and reward FAIR data and data stewardship

» Rec. 7: Support semantic technologies

» Rec. 8: Facilitate automated processing

» Rec. 9: Develop assessment frameworks to certify FAIR services

» Rec. 10: Professionalise data science and data stewardship roles and train researchers

» Rec. 11: Implement curriculum frameworks and training

1.7.3 Step 3: Embed and sustain – incentives, metrics and investment

» Rec. 12: Develop metrics for FAIR Digital Objects

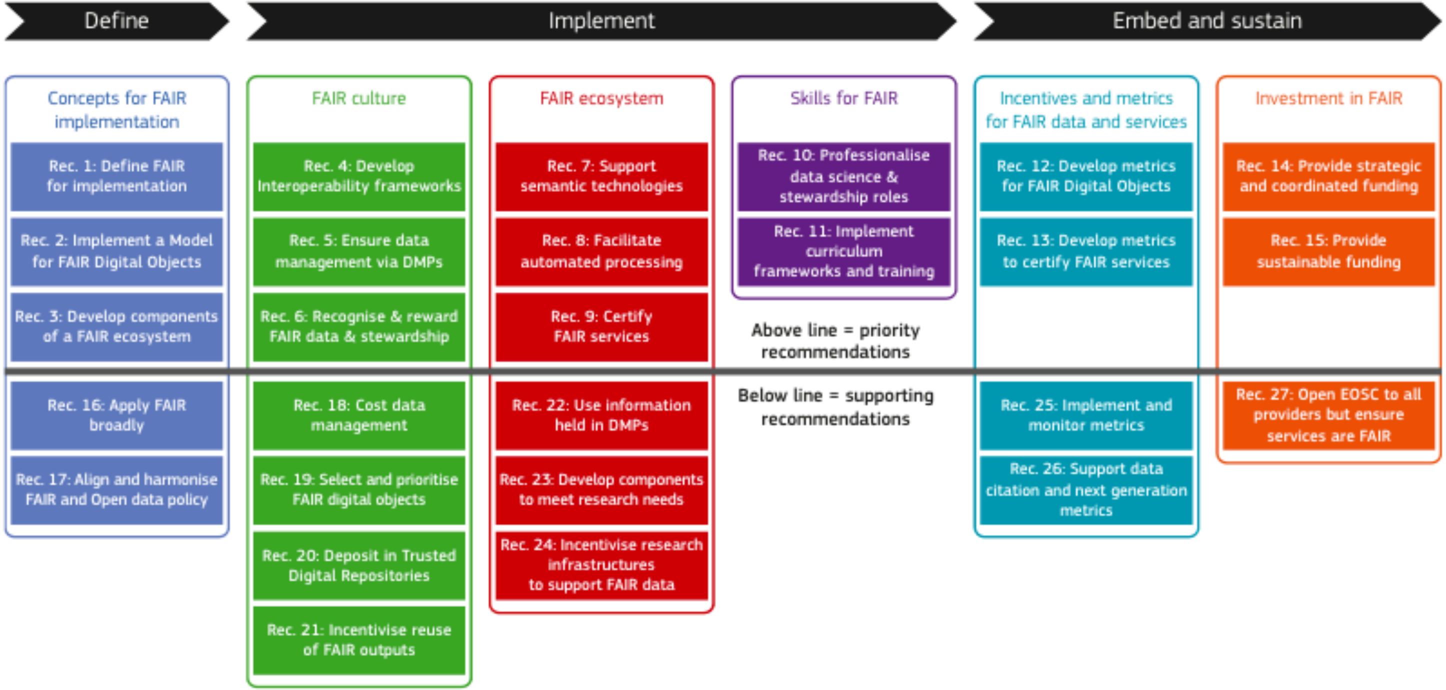
» Rec. 13: Develop metrics to certify FAIR services

» Rec. 14: Provide strategic and coordinated funding

» Rec. 15: Provide sustainable funding

From: Final Report and Action Plan from the European Commission Expert Group on FAIR Data

Energy Hub @ HMC

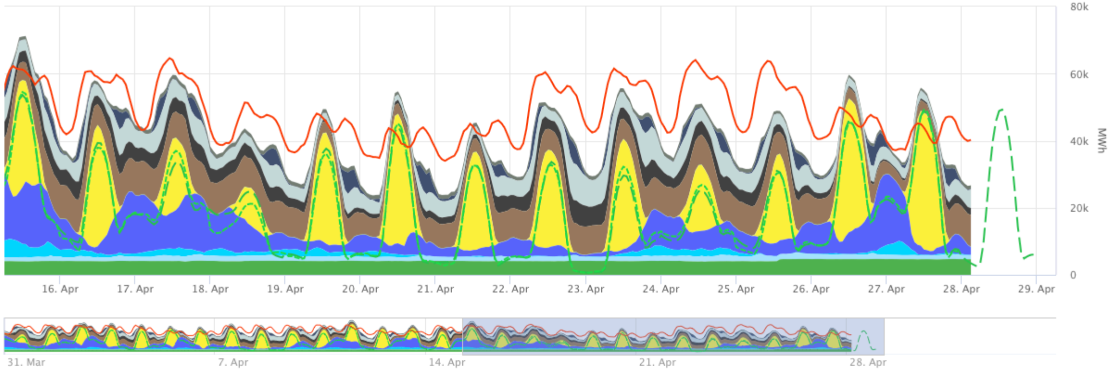


Define	Implement			Embed and sustain	
Concepts for FAIR implementation	FAIR culture	FAIR ecosystem	Skills for FAIR	Incentives and metrics for FAIR data and services	Investment in FAIR
Rec. 1: Define FAIR for implementation	Rec. 4: Develop Interoperability frameworks	Rec. 7: Support semantic technologies	Rec. 10: Professionalise data science & stewardship roles	Rec. 12: Develop metrics for FAIR Digital Objects	Rec. 14: Provide strategic and coordinated funding
Rec. 2: Implement a Model for FAIR Digital Objects	Rec. 5: Ensure data management via DMPs	Rec. 8: Facilitate automated processing	Rec. 11: Implement curriculum frameworks and training	Rec. 13: Develop metrics to certify FAIR services	Rec. 15: Provide sustainable funding
Rec. 3: Develop components of a FAIR ecosystem	Rec. 6: Recognise & reward FAIR data & stewardship	Rec. 9: Certify FAIR services	Above line = priority recommendations		
Rec. 16: Apply FAIR broadly	Rec. 18: Cost data management	Rec. 22: Use information held in DMPs	Below line = supporting recommendations	Rec. 25: Implement and monitor metrics	Rec. 27: Open EOSC to all providers but ensure services are FAIR
Rec. 17: Align and harmonise FAIR and Open data policy	Rec. 19: Select and prioritise FAIR digital objects	Rec. 23: Develop components to meet research needs		Rec. 26: Support data citation and next generation metrics	
	Rec. 20: Deposit in Trusted Digital Repositories	Rec. 24: Incentivise research infrastructures to support FAIR data			
	Rec. 21: Incentivise reuse of FAIR outputs				

Index to FAIR Action Plan recommendations

Actual generation, Forecasted generation day ahead and actual consumption

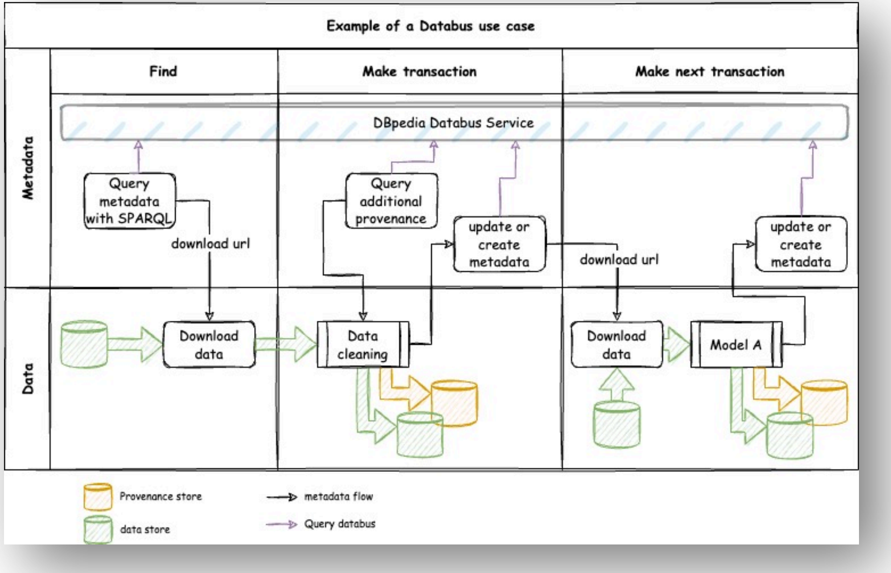
- Many data bases exist, each in its own flavour
  - Data access
  - Data format
  - Data licenses (if at all)
  - Sometimes hard to find
- Data collection is a labor intensive task
- Data cleaning, aggregation, etc. is repeated by many researchers with different results
- Data quality is often unknown



From: https://www.smard.de/

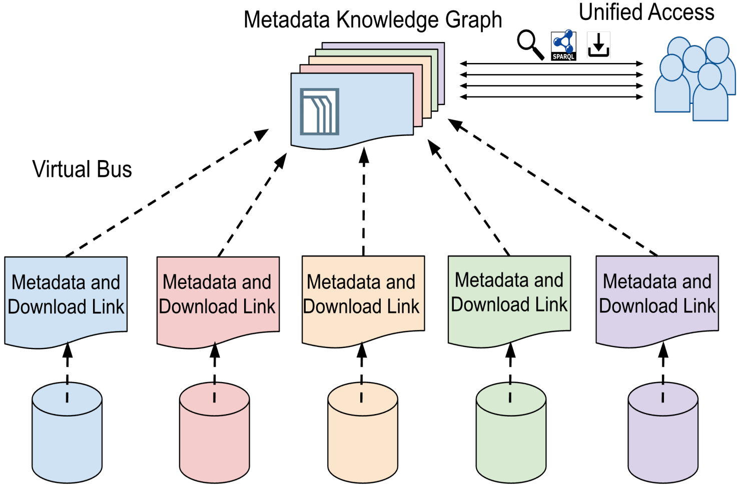
Example Application, when things work well

- Automatic Metadata Registration on a target Platform
- A Databus (for example the DBPedia Databus) is a distributed database architecture, where provider can publish and search within standardised metadata
- A bridge between a DMP-Tool and a target platform is realised
- Augmented Metadata reuse through management of hierarchical structures



Example Application, when things work well

- A metadata catalog harvests the (rich) metadata from the available data sources
- The catalog can be used to discover data
- The metadata contains a URI to the actual data or to an API


Metadata Knowledge Graph
Unified Access
SPARQL
Virtual Bus
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link

Funding Requirements

Example Requirements, From the DFG Checklist: Handling of research data

- Data description

  How does your project generate new data?

  Is existing data reused?

  Which data types (in terms of data formats like image data, text data or measurement data) arise in your project and in what way are they further processed?

  To what extent do these arise or what is the anticipated data volume?

- Documentation and data quality

  What approaches are being taken to describe the data in a comprehensible manner (such as the use of available metadata, documentation standards or ontologies)?

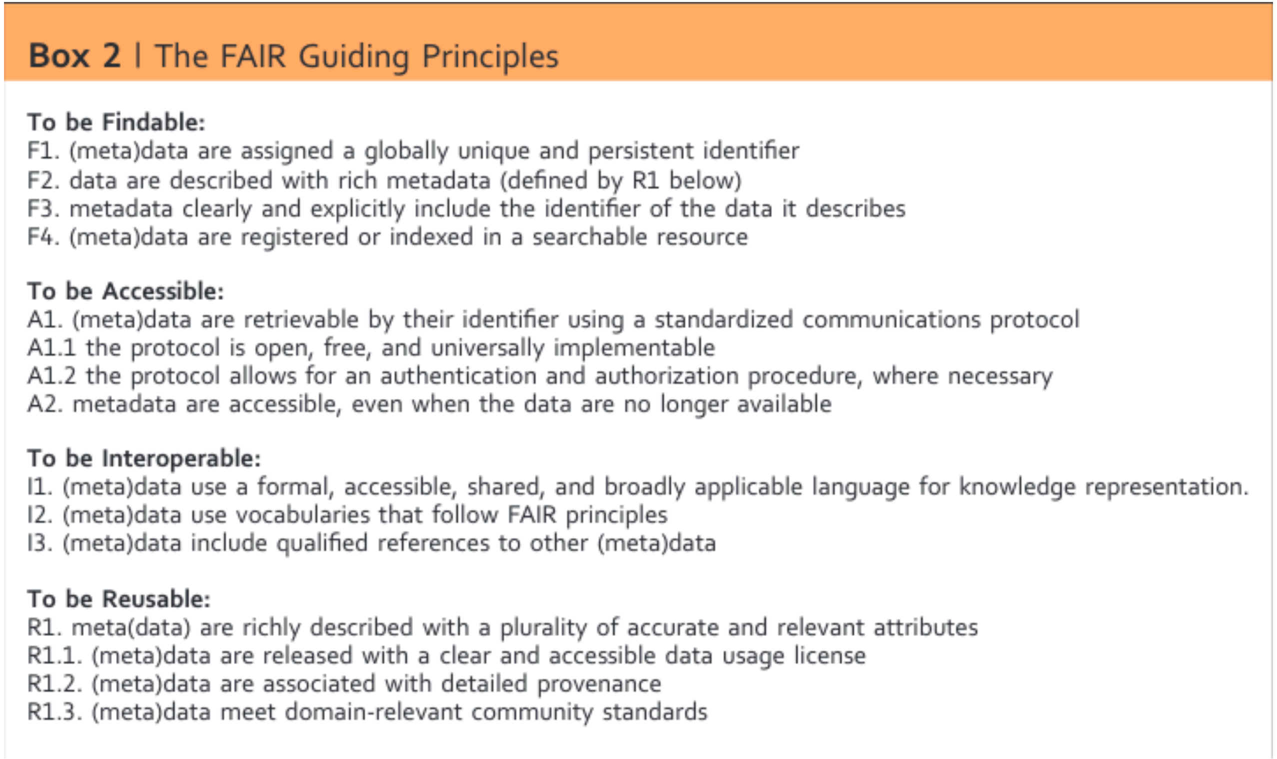
  What measures are being adopted to ensure high data quality?

  Are quality controls in place and if so, how do they operate?

  Which digital methods and tools (e.g. software) are required to use the data?

Funding
Requirements
- Data needs to be published in our repository
- annotation with structured metadata is required

FAIR Guiding Principles



Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

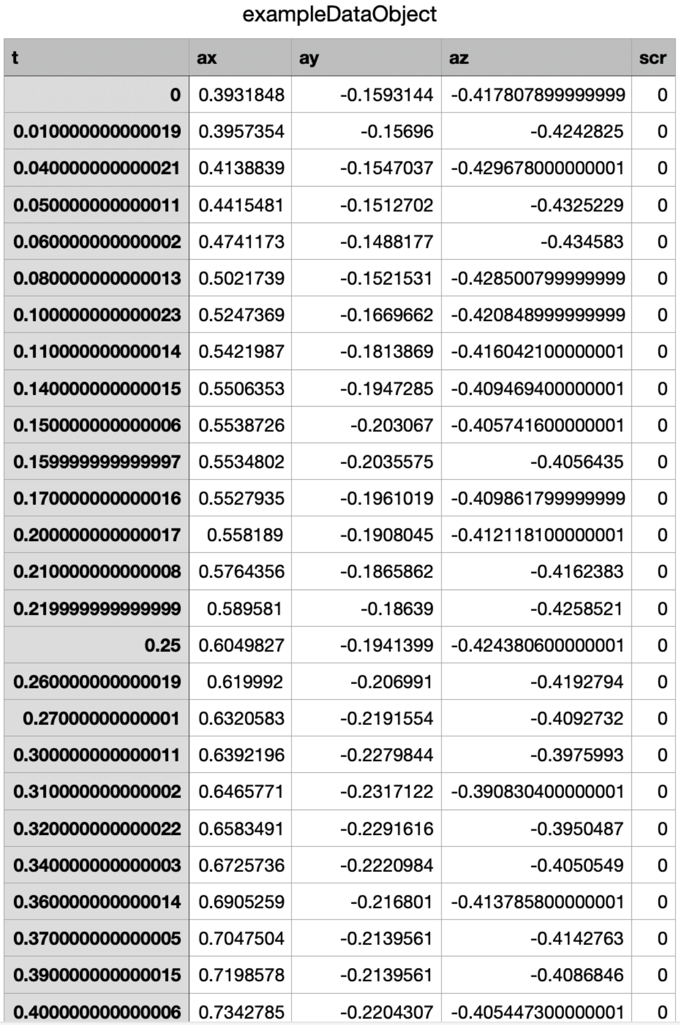
From: The FAIR Guiding Principles for scientific data management and stewardship Mark D. Wilkinson et al. (Published: 15 March 2016)

Steps towards well described data

Enriching data with metadata is a key concept for the data output of scientific research to be FAIR.

Data processing software and custom code often do not support the annotation with metadata out-of-the-box.

This confronts data creators and maintainers with challenges to annotate their data.



exampleDataObject

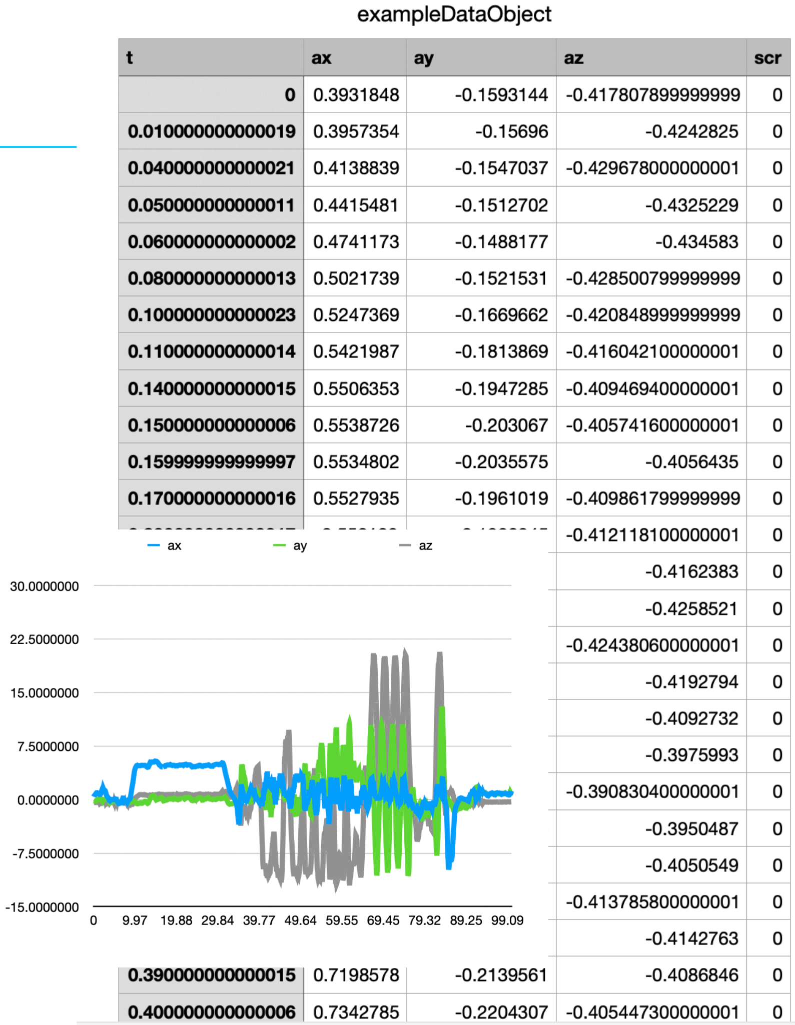
t	ax	ay	az	scr
0	0.3931848	-0.1593144	-0.417807899999999	0
0.010000000000019	0.3957354	-0.15696	-0.4242825	0
0.040000000000021	0.4138839	-0.1547037	-0.429678000000001	0
0.050000000000011	0.4415481	-0.1512702	-0.4325229	0
0.060000000000002	0.4741173	-0.1488177	-0.434583	0
0.080000000000013	0.5021739	-0.1521531	-0.428500799999999	0
0.100000000000023	0.5247369	-0.1669662	-0.420848999999999	0
0.110000000000014	0.5421987	-0.1813869	-0.416042100000001	0
0.140000000000015	0.5506353	-0.1947285	-0.409469400000001	0
0.150000000000006	0.5538726	-0.203067	-0.405741600000001	0
0.159999999999997	0.5534802	-0.2035575	-0.4056435	0
0.170000000000016	0.5527935	-0.1961019	-0.409861799999999	0
0.200000000000017	0.558189	-0.1908045	-0.412118100000001	0
0.210000000000008	0.5764356	-0.1865862	-0.4162383	0
0.219999999999999	0.589581	-0.18639	-0.4258521	0
0.25	0.6049827	-0.1941399	-0.424380600000001	0
0.260000000000019	0.619992	-0.206991	-0.4192794	0
0.27000000000001	0.6320583	-0.2191554	-0.4092732	0
0.300000000000011	0.6392196	-0.2279844	-0.3975993	0
0.310000000000002	0.6465771	-0.2317122	-0.390830400000001	0
0.320000000000022	0.6583491	-0.2291616	-0.3950487	0
0.340000000000003	0.6725736	-0.2220984	-0.4050549	0
0.360000000000014	0.6905259	-0.216801	-0.413785800000001	0
0.370000000000005	0.7047504	-0.2139561	-0.4142763	0
0.390000000000015	0.7198578	-0.2139561	-0.4086846	0
0.400000000000006	0.7342785	-0.2204307	-0.405447300000001	0

Steps towards well described data

Enriching data with metadata is a key concept for the data output of scientific research to be FAIR.

Data processing software and custom code often do not support the annotation with metadata out-of-the-box.

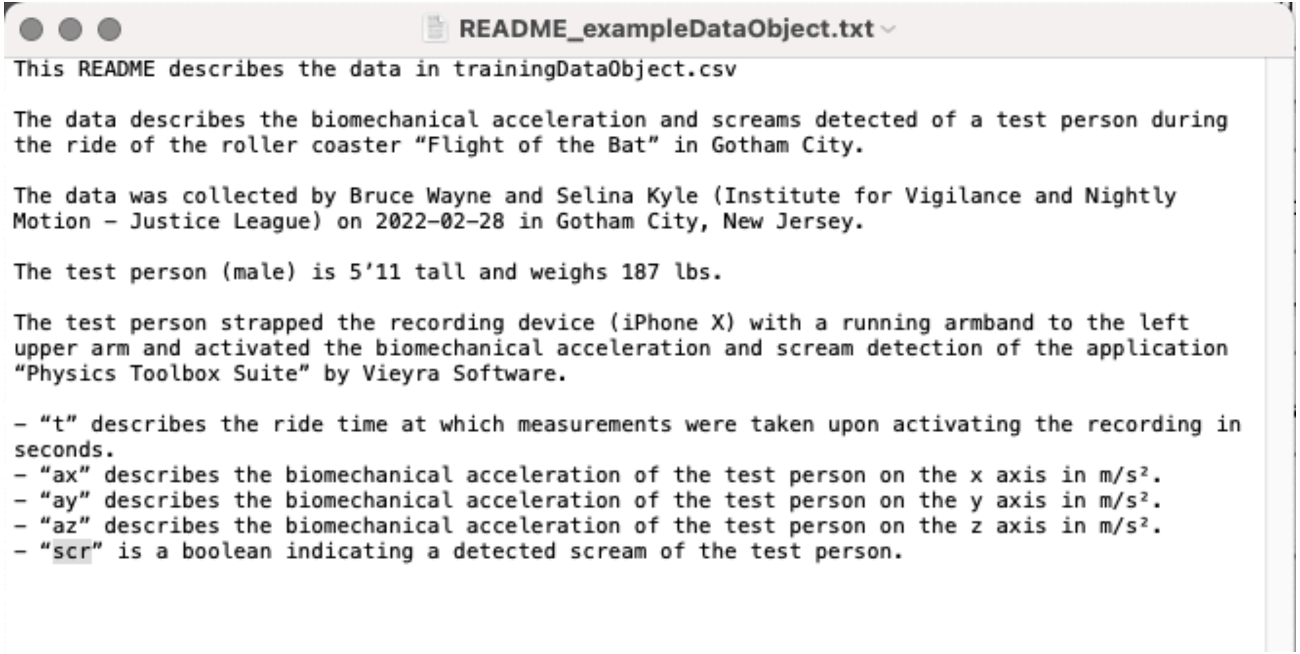
This confronts data creators and maintainers with challenges to annotate their data.



exampleDataObject

t	ax	ay	az	scr
0	0.3931848	-0.1593144	-0.417807899999999	0
0.010000000000019	0.3957354	-0.15696	-0.4242825	0
0.040000000000021	0.4138839	-0.1547037	-0.429678000000001	0
0.050000000000011	0.4415481	-0.1512702	-0.4325229	0
0.060000000000002	0.4741173	-0.1488177	-0.434583	0
0.080000000000013	0.5021739	-0.1521531	-0.428500799999999	0
0.100000000000023	0.5247369	-0.1669662	-0.420848999999999	0
0.110000000000014	0.5421987	-0.1813869	-0.416042100000001	0
0.140000000000015	0.5506353	-0.1947285	-0.409469400000001	0
0.150000000000006	0.5538726	-0.203067	-0.405741600000001	0
0.159999999999997	0.5534802	-0.2035575	-0.4056435	0
0.170000000000016	0.5527935	-0.1961019	-0.409861799999999	0
			-0.412118100000001	0
			-0.4162383	0
			-0.4258521	0
			-0.424380600000001	0
			-0.4192794	0
			-0.4092732	0
			-0.3975993	0
			-0.390830400000001	0
			-0.3950487	0
			-0.4050549	0
			-0.413785800000001	0
			-0.4142763	0
0.390000000000015	0.7198578	-0.2139561	-0.4086846	0
0.400000000000006	0.7342785	-0.2204307	-0.405447300000001	0

Steps towards well described data



This README describes the data in trainingDataObject.csv

The data describes the biomechanical acceleration and screams detected of a test person during the ride of the roller coaster “Flight of the Bat” in Gotham City.

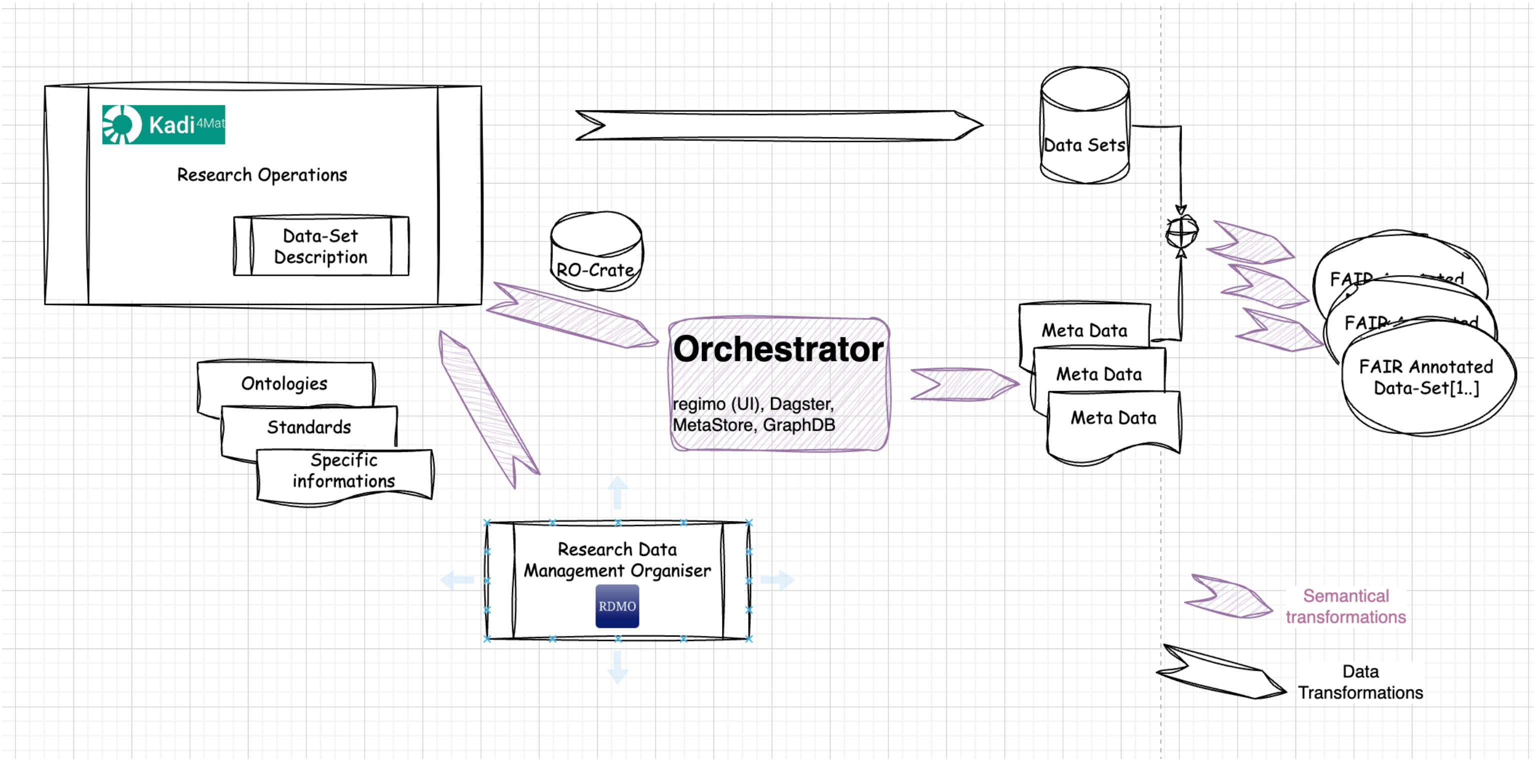
The data was collected by Bruce Wayne and Selina Kyle (Institute for Vigilance and Nightly Motion – Justice League) on 2022-02-28 in Gotham City, New Jersey.

The test person (male) is 5’11 tall and weighs 187 lbs.

The test person strapped the recording device (iPhone X) with a running armband to the left upper arm and activated the biomechanical acceleration and scream detection of the application “Physics Toolbox Suite” by Vieyra Software.

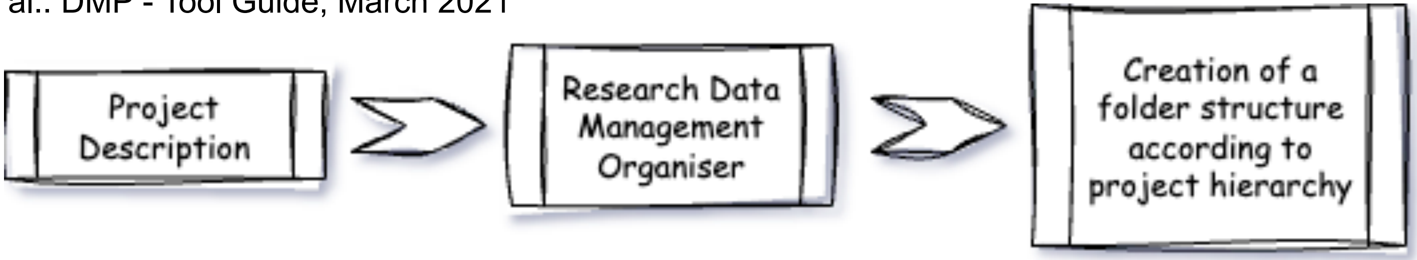
- “t” describes the ride time at which measurements were taken upon activating the recording in seconds.
- “ax” describes the biomechanical acceleration of the test person on the x axis in m/s^2.
- “ay” describes the biomechanical acceleration of the test person on the y axis in m/s^2.
- “az” describes the biomechanical acceleration of the test person on the z axis in m/s^2.
- “scr” is a boolean indicating a detected scream of the test person.

Overview of a vision

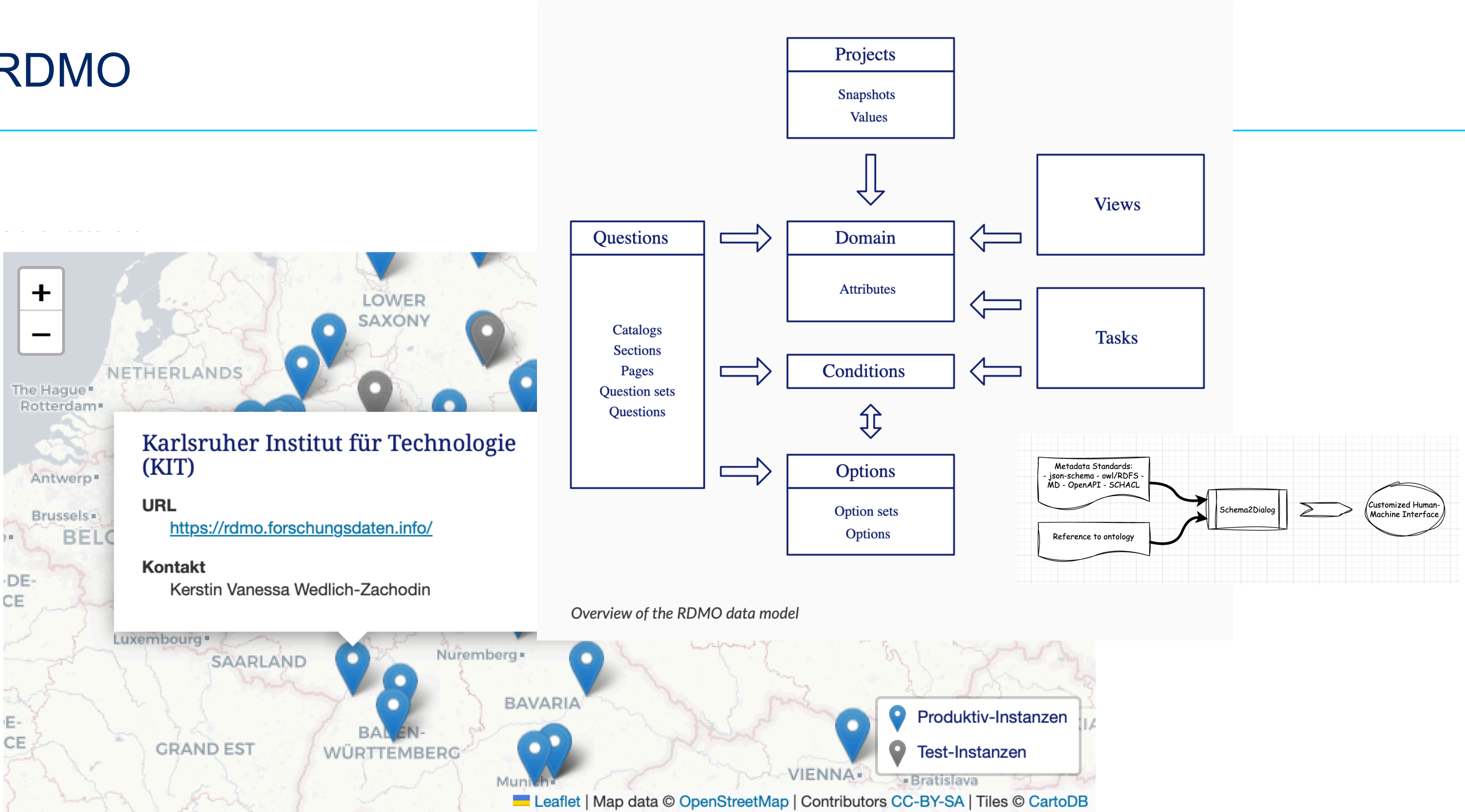


Recommendation 1 : Planing

- Data management plans (DMPs) are increasingly required by funding agencies – either as part of the research proposal or as an early project deliverable.
- Even without a formal obligation, you may do yourself a favour by creating a data management plan for your own.
- DMPs cover aspects of Data collection and description, their documentation and requirements on the metadata to be used.
- Additionally aspects concerning the storage of data and their long-term preservation are handled during the creation of DMPs.
- For a benchmarking of existing tools to generate DMP please consult Helbig et al.. DMP - Tool Guide, March 2021


Project Description
Research Data Management Organiser
Creation of a folder structure according to project hierarchy

RDMO



Karlsruher Institut für Technologie (KIT)

URL

https://rdmo.forschungsdaten.info/

Kontakt

Kerstin Vanessa Wedlich-Zachodin

Produktiv-Instanzen

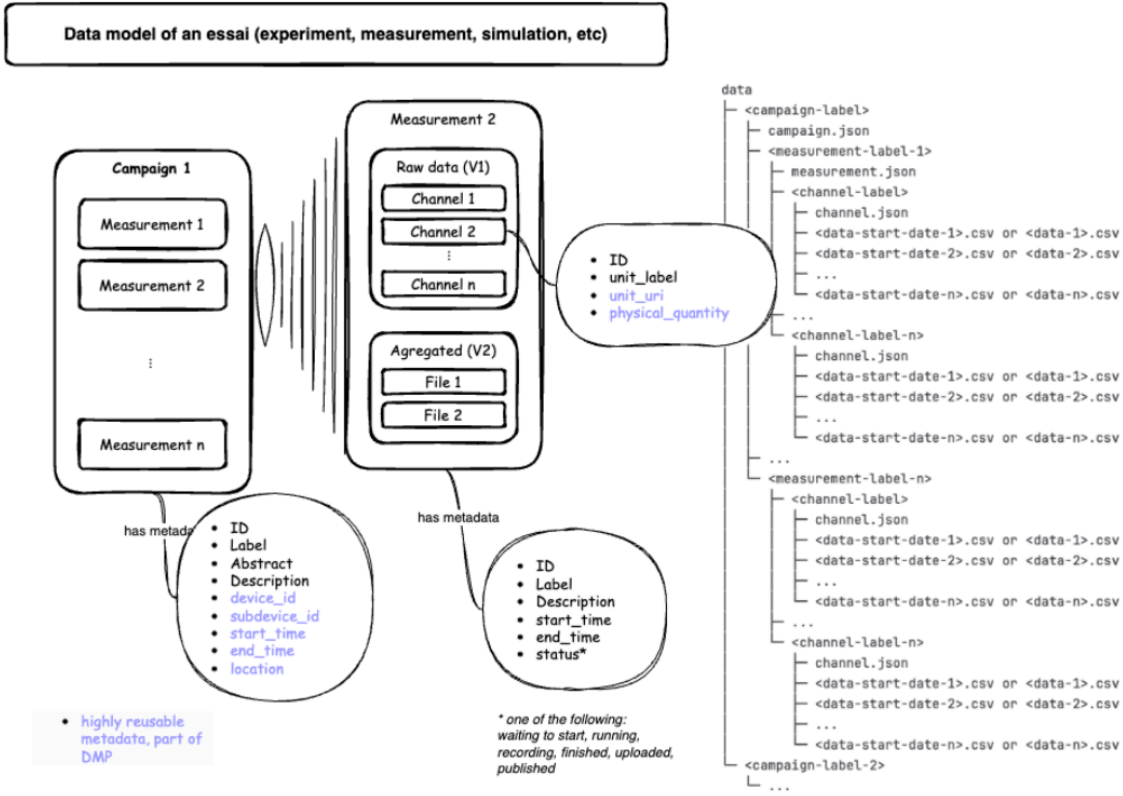
Test-Instanzen

Leaflet | Map data © OpenStreetMap | Contributors CC-BY-SA | Tiles © CartoDB

Overview of the RDMO data model

Filter 57/57

Data Model of an essai



Data model of an essai (experiment, measurement, simulation, etc)

- ID
- Label
- Abstract
- Description
- device_id
- subdevice_id
- start_time
- end_time
- location

- highly reusable metadata, part of DMP

- ID
- unit_label
- unit_uri
- physical_quantity

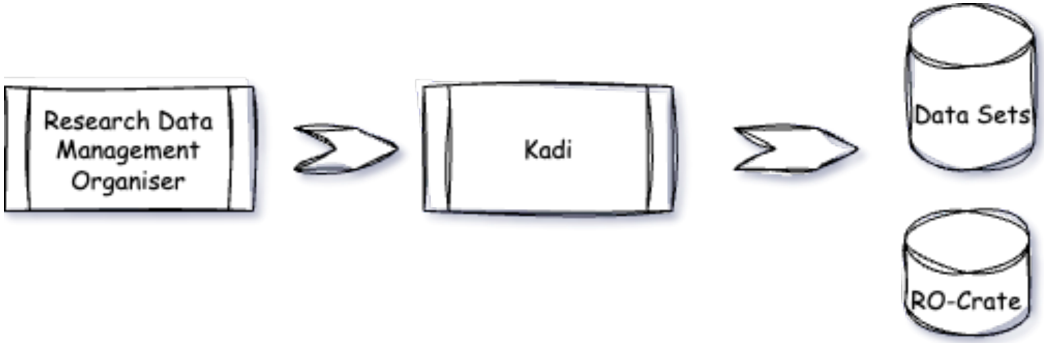
- ID
- Label
- Description
- start_time
- end_time
- status*

* one of the following: waiting to start, running, recording, finished, uploaded, published

```
data
├ <campaign-label>
│ ├ campaign.json
│ ├ <measurement-label-1>
│ │ ├ measurement.json
│ │ ├ <channel-label>
│ │ │ ├ channel.json
│ │ │ ├ <data-start-date-1>.csv or <data-1>.csv
│ │ │ ├ <data-start-date-2>.csv or <data-2>.csv
│ │ │ ├ ...
│ │ │ └ <data-start-date-n>.csv or <data-n>.csv
│ │ ├ ...
│ │ └ <channel-label-n>
│ │   ├ channel.json
│ │   ├ <data-start-date-1>.csv or <data-1>.csv
│ │   ├ <data-start-date-2>.csv or <data-2>.csv
│ │   ├ ...
│ │   └ <data-start-date-n>.csv or <data-n>.csv
│ ├ ...
│ └ <measurement-label-n>
│   ├ <channel-label>
│   │ ├ channel.json
│   │ ├ <data-start-date-1>.csv or <data-1>.csv
│   │ ├ <data-start-date-2>.csv or <data-2>.csv
│   │ ├ ...
│   │ └ <data-start-date-n>.csv or <data-n>.csv
│   ├ ...
│   └ <channel-label-n>
│     ├ channel.json
│     ├ <data-start-date-1>.csv or <data-1>.csv
│     ├ <data-start-date-2>.csv or <data-2>.csv
│     ├ ...
│     └ <data-start-date-n>.csv or <data-n>.csv
└ <campaign-label-2>
  └ ...
```

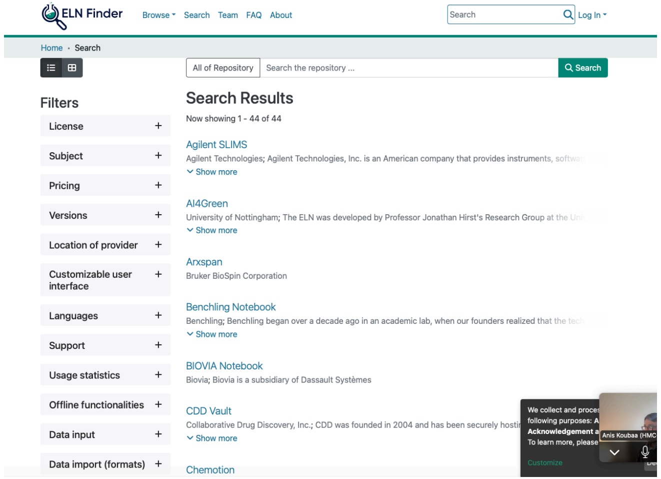
Recommendation 2 : Timing

- Start documenting metadata as early as possible!
- The effort to generate metadata retrospectively and attach them to a set of data is often very high.
- Therefore, the recording of metadata should always happen alongside or close to the generation of the research data itself.
- Using an Electronic Lab Notebook is helpful in that regard.
- The ELN Finder helps you to search and select a suitable ELN for your purposes.



Recommendation 2 : Timing

- Start documenting metadata as early as possible!
- The effort to generate metadata retrospectively and attach them to a set of data is often very high.
- Therefore, the recording of metadata should always happen alongside or close to the generation of the research data itself.
- Using an Electronic Lab Notebook is helpful in that regard.
- The ELN Finder helps you to search and select a suitable ELN for your purposes.



Search Results

Now showing 1 - 44 of 44

Agilent SLIMS
Agilent Technologies; Agilent Technologies, Inc. is an American company that provides instruments, softwa
Show more

AI4Green
University of Nottingham; The ELN was developed by Professor Jonathan Hirst's Research Group at the Uni
Show more

Arxspan
Bruker BioSpin Corporation

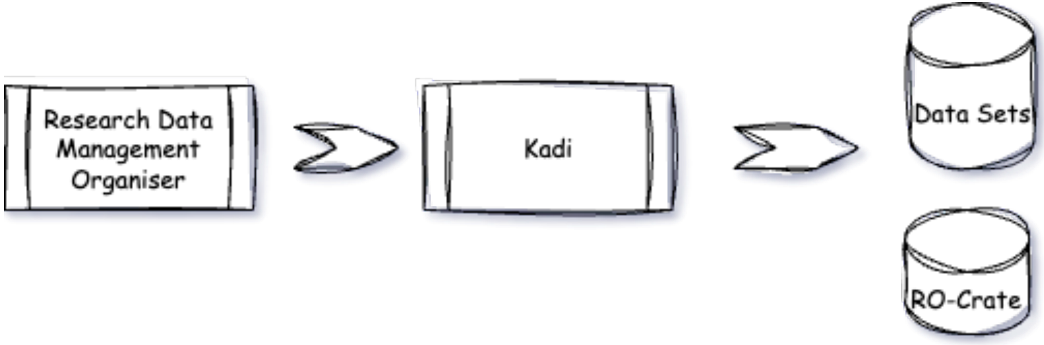
Benchling Notebook
Benchling; Benchling began over a decade ago in an academic lab, when our founders realized that the tec
Show more

BIOVIA Notebook
Biovia; Biovia is a subsidiary of Dassault Systèmes

CDD Vault
Collaborative Drug Discovery, Inc.; CDD was founded in 2004 and has been securely hosti
Show more

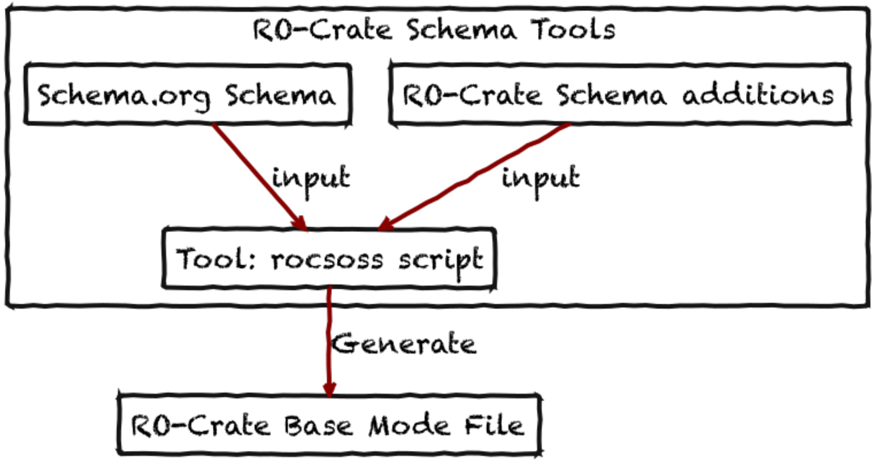
Chemotion

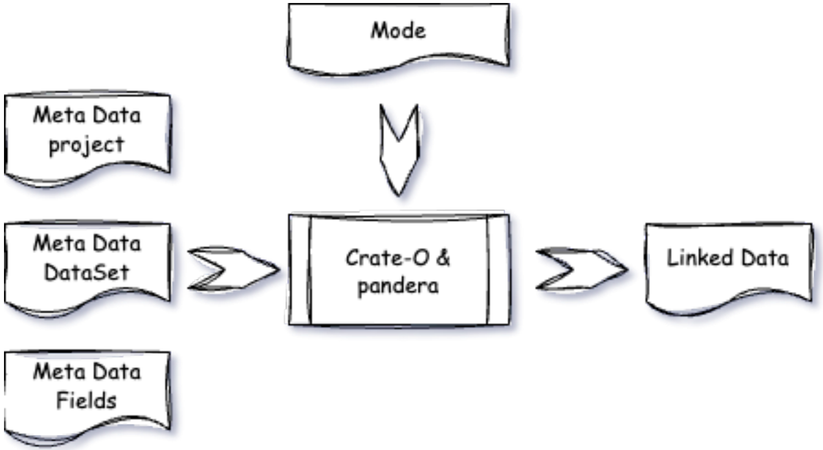
We collect and proces
following purposes: A
Acknowledgement a
To learn more, please



Recommendation 3 : Metadata Records Editing

- RO-Crates generated by ELNs should be edited to enhance their quality.
- RO-Crates are Linked Data serialised in JSON-LD, which perfectly harmonise with DataID-EcoSystem of the Databus
- Editing of Linked Data can be realised with specific editors





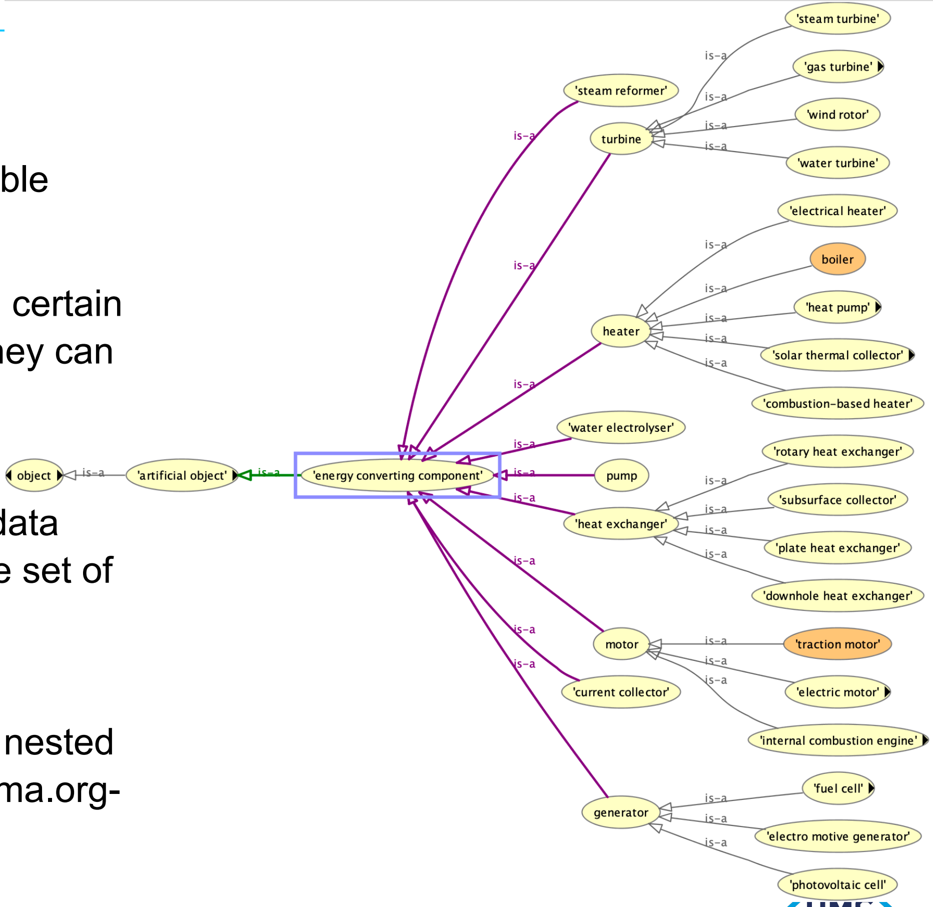
Recommendation 4 : Standards and vocabularies

- Do not reinvent the wheel; use existing (domain-specific) and machine readable standards to enable interoperability.

- Ontologies are formal descriptions of entities in a certain domain and their relationships to one another. They can compile the knowledge in a domain in a very standardised and efficient way.

- Open Energy Metadata (OEMetadata) is a metadata standard for the energy domain. It is an extensive set of metadata based on the tabular data package specifications and the FAIR principles.

- The metadata contains multiple fields (keys) in a nested JSON structure. In the context of ro-crates, schema.org-vocabulary is used.



Existing Ontology in Energy Domain

- SARGON ontology was developed to cover several ontologies in the smart grid and building automation.
- SAREF4BLDG is designed to bridge the interoperability gap between different stakeholders and the applications used to manage building information throughout the phases of the building life cycle
- SAREF4ENER is designed to focus on demand response scenarios, where consumers offer energy flexibility to the smart grid.
- EM-KPI ontology is designed to facilitate the exchange of master data and key performance indicators (KPIs) for energy management at the district and building levels, focusing solely on electricity usage
- The Smart Building Evacuation Ontology (SBEO) is a reusable framework for indoor spaces that integrates three key data models: user, building, and context. Its structure includes a user model to represent occupant characteristics and relationships, a building model to describe the layout and infrastructure, and a context model to capture dynamic changes in both the building and its occupants.
- The Flow Systems Ontology (FSO) is designed to describe the energy and mass flow relationships within systems and their components, along with the composition of the system.

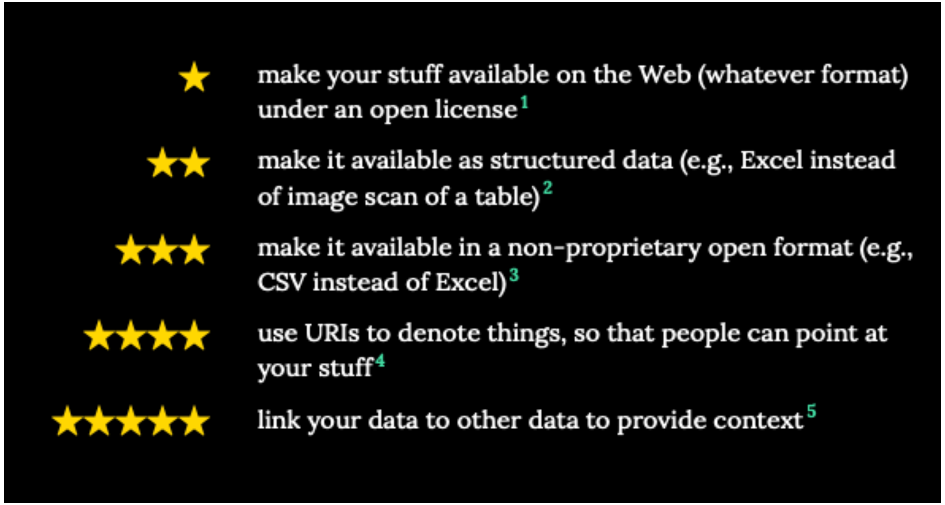
OEMetadata

- A metadata standard for „energy related data“
- Based on existing technologies and standards as “Frictionless Data” and “DataCite”
- Implemented as JSON-LD to be human and machine readable
- Latest release (v2.0) is “ontology ready”

- **Categories**
  - **General** (name, title, description)
  - **Context** (homepage, funding, contact)
  - **Spatial** (location, extent, resolution)
  - **Temporal** (referenceDate, timeseries)
  - **Source** (origin, licenses)
  - **Provenience** (contributors)
  - **Resource** (schema, fields, type, description)
  - **Review** (context and badge)

OEMetadata: 5-star Linked Open Data

- A metadata standard for „energy related data“
- Based on existing technologies and standards as “Frictionless Data” and “DataCite”
- Implemented as JSON-LD to be human and machine readable
- Latest release (v2.0) is “ontology ready”
- Target: 5-star Linked Open Data



★ make your stuff available on the Web (whatever format) under an open license[1]

★★ make it available as structured data (e.g., Excel instead of image scan of a table)[2]

★★★ make it available in a non-proprietary open format (e.g., CSV instead of Excel)[3]

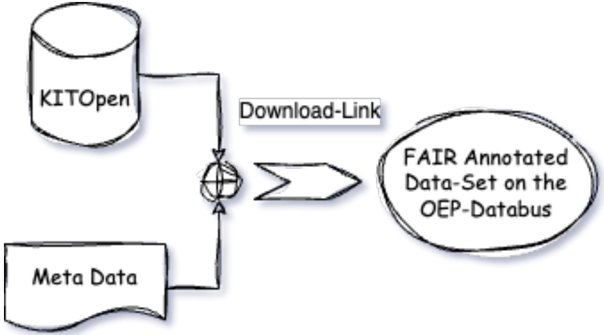
★★★★ use URIs to denote things, so that people can point at your stuff[4]

★★★★★ link your data to other data to provide context[5]

From: https://5stardata.info/en/

Recommendation 5 : Publish Metadata

- Make your metadata discoverable and accessible by publishing in a searchable resource (e.g. trustworthy repository) even if the data themself are not publicly available. Assign a persistent identifier (PID) such as a DOI to your metadata record to make it findable and citable.
- When data are uploaded to a repository one can retrieve a download-link to the actual data.
- The download-link and the gathered metadata will be used to generate an OEMetaString
- Several tutorials to helping understanding handling OEMetadata are provided by the OEP
- OEP provides a wizard for manually publishing Metadata on the Databus
- An automated Orchestrator uses the API to the Databus to register Metadata and thus make the data findable

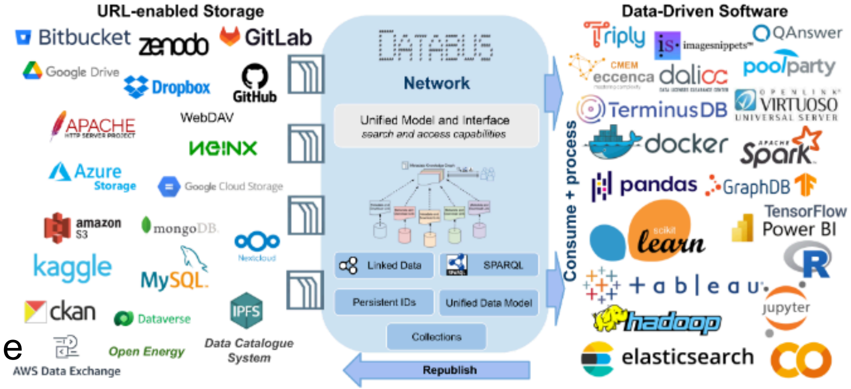


The Open Energy Platform

- OEP is a platform that facilitates agile data integration, collaboration, and automation through a structured metadata Knowledge Graph

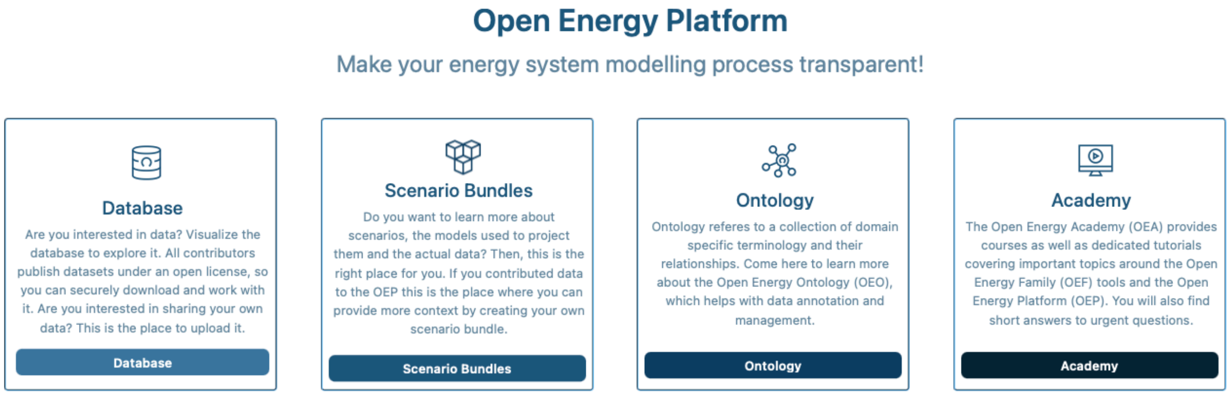
- Realising data-pipelines is a typical use case of a Databus. Data-pipelines have components (actor, action, software, test equipment, etc.) which are assigned to unique identifier

- Data-Experiments are fully reconsructable in a large ecosystem of consumer and processors



OEP - Hands On

- https://openenergyplatform.org/



Open Energy Platform

Make your energy system modelling process transparent!

Database

Are you interested in data? Visualize the database to explore it. All contributors publish datasets under an open license, so you can securely download and work with it. Are you interested in sharing your own data? This is the place to upload it.

Scenario Bundles

Do you want to learn more about scenarios, the models used to project them and the actual data? Then, this is the right place for you. If you contributed data to the OEP this is the place where you can provide more context by creating your own scenario bundle.

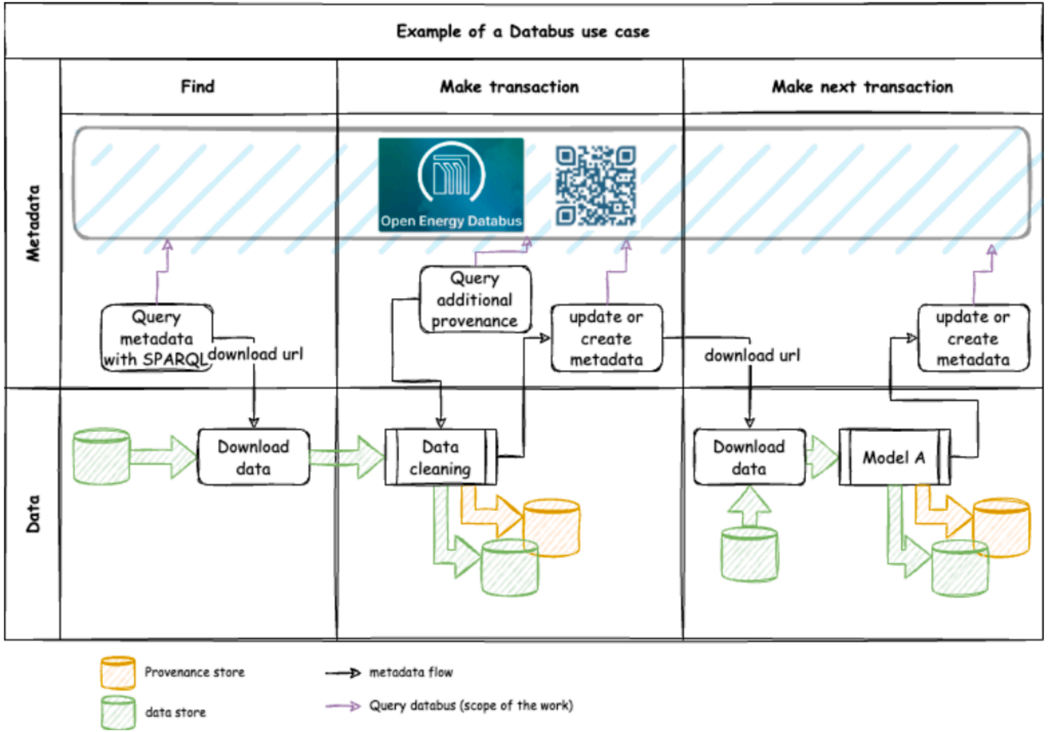
Ontology

Ontology referes to a collection of domain specific terminology and their relationships. Come here to learn more about the Open Energy Ontology (OEO), which helps with data annotation and management.

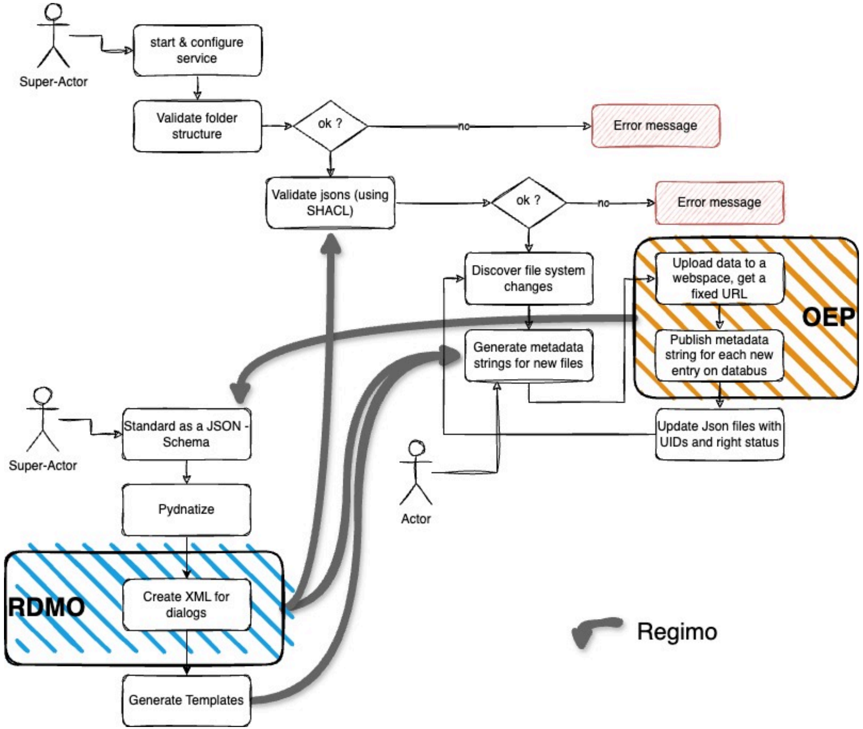
Academy

The Open Energy Academy (OEA) provides courses as well as dedicated tutorials covering important topics around the Open Energy Family (OEF) tools and the Open Energy Platform (OEP). You will also find short answers to urgent questions.

The Open Energy Platform - Databus use case



Regimo as bridge between maDMP and Target Platform (in both directions)

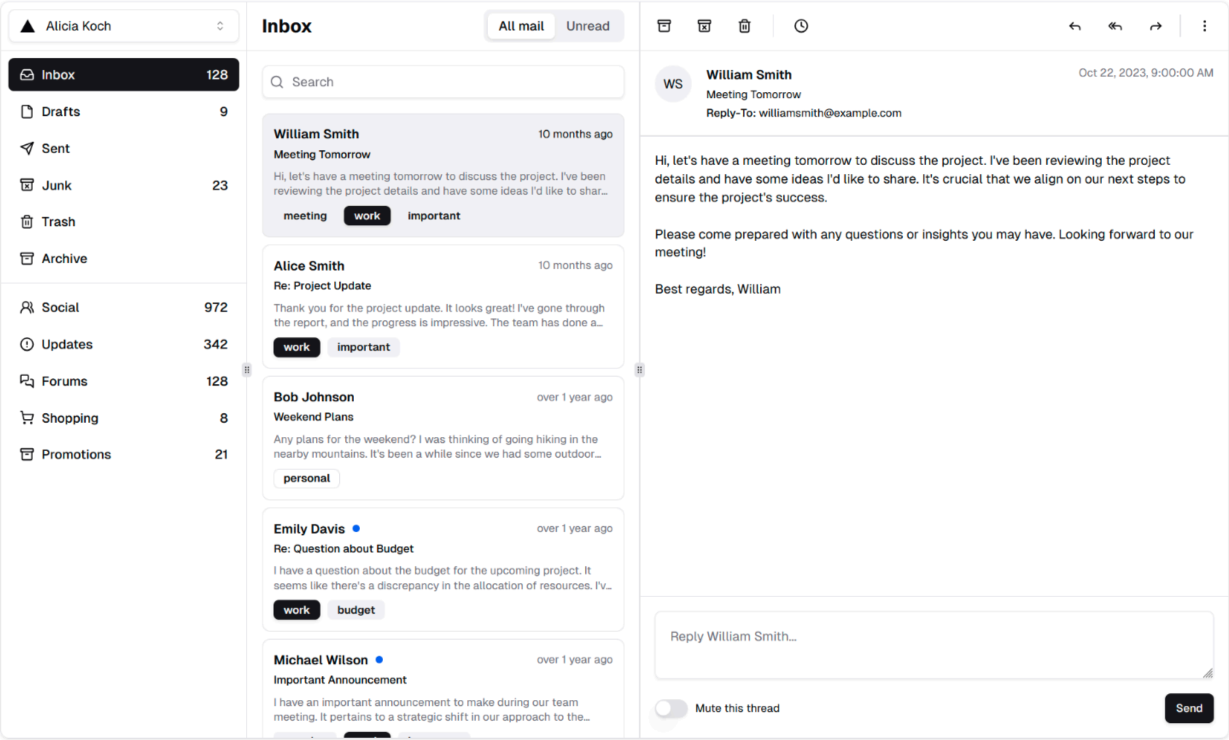

Super-Actor
start & configure service
Validate folder structure
ok ?
no
Error message
Validate jsons (using SHACL)
ok ?
no
Error message
Discover file system changes
Generate metadata strings for new files
Upload data to a webspace, get a fixed URL
Publish metadata string for each new entry on databus
OEP
Update Json files with UIDs and right status
Actor
Super-Actor
Standard as a JSON - Schema
Pydnatize
RDMO
Create XML for dialogs
Generate Templates
Regimo

Examples of Activities within IAI

- Integration of regimo in a productive DataFactory
  - IAI, Publishing of EDR-Measurements (Network frequency data)
  - IAI, Publishing of Living Labs data
  - IAI, Publishing of Gredler Areal (a small smart-grid)
- Integrate unHide - Harvestability in regimo
- Recommendation of the integration of FDO’s on Open Energy Platform

➤ ***General usable Concept for IAI***

Edit & Send



William Smith 10 months ago
Meeting Tomorrow
Hi, let's have a meeting tomorrow to discuss the project. I've been reviewing the project details and have some ideas I'd like to shar...
meeting work important

Alice Smith 10 months ago
Re: Project Update
Thank you for the project update. It looks great! I've gone through the report, and the progress is impressive. The team has done a...
work important

Bob Johnson over 1 year ago
Weekend Plans
Any plans for the weekend? I was thinking of going hiking in the nearby mountains. It's been a while since we had some outdoor...
personal

Emily Davis ● over 1 year ago
Re: Question about Budget
I have a question about the budget for the upcoming project. It seems like there's a discrepancy in the allocation of resources. I'v...
work budget

Michael Wilson ● over 1 year ago
Important Announcement
I have an important announcement to make during our team meeting. It pertains to a strategic shift in our approach to the...

WS **William Smith** Oct 22, 2023, 9:00:00 AM
Meeting Tomorrow
Reply-To: williamsmith@example.com

Hi, let's have a meeting tomorrow to discuss the project. I've been reviewing the project details and have some ideas I'd like to share. It's crucial that we align on our next steps to ensure the project's success.

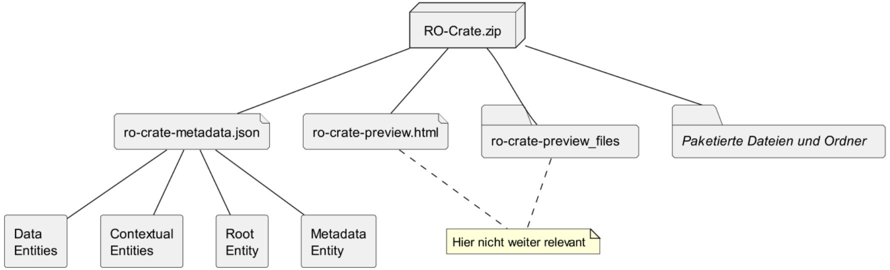
Please come prepared with any questions or insights you may have. Looking forward to our meeting!

Best regards, William

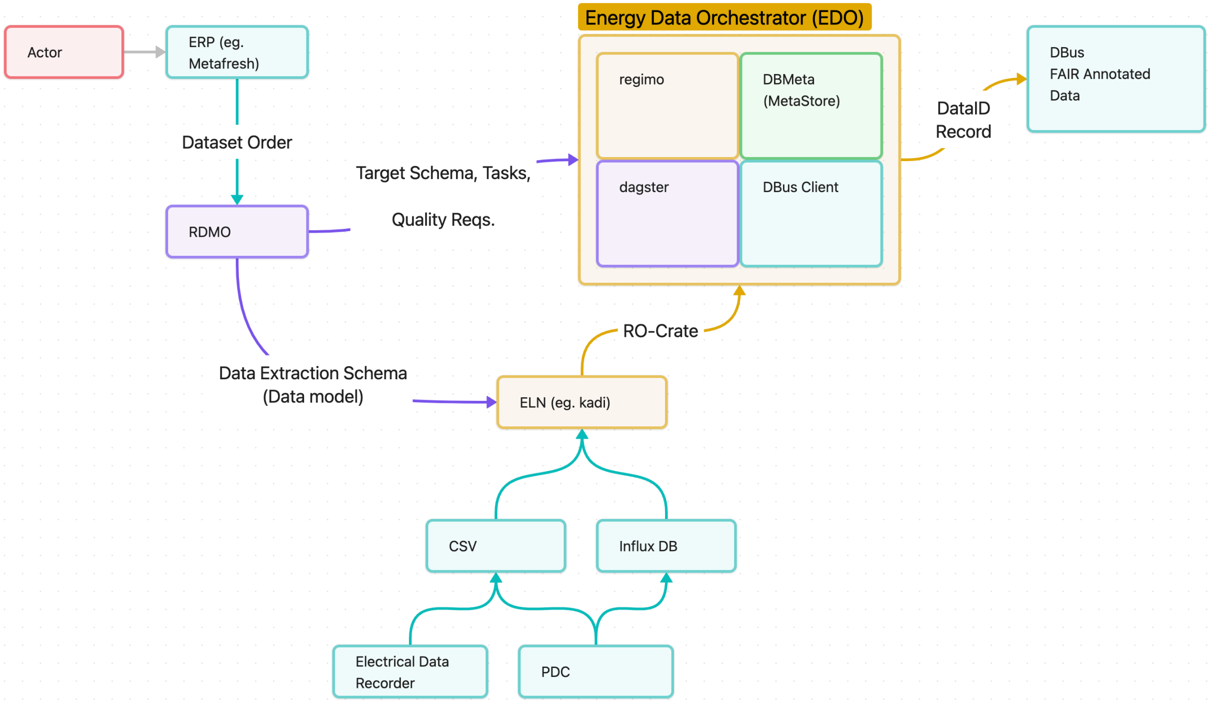
RO-Crate

- A Design, an Implementierung and an Evaluation of a Web-Editor für Research Object Crates are documented in DOI: 10.5445/IR/1000178790

- Research Object Crates are a method for enabling easy exchange of research data and supporting the reproducibility of scientific works.
- RO-Crates are created by packaging research data and their metadata together, aggregating them in a single archive
- Most of the functional Requirements were realised
- The editor can be integrated as a Web Component in own automation solution



SW-Components Overview



End of Day 1

- Discussion
- Questions for Day 2
- Contributions of present colleagues (some words to the ro - crate editor)

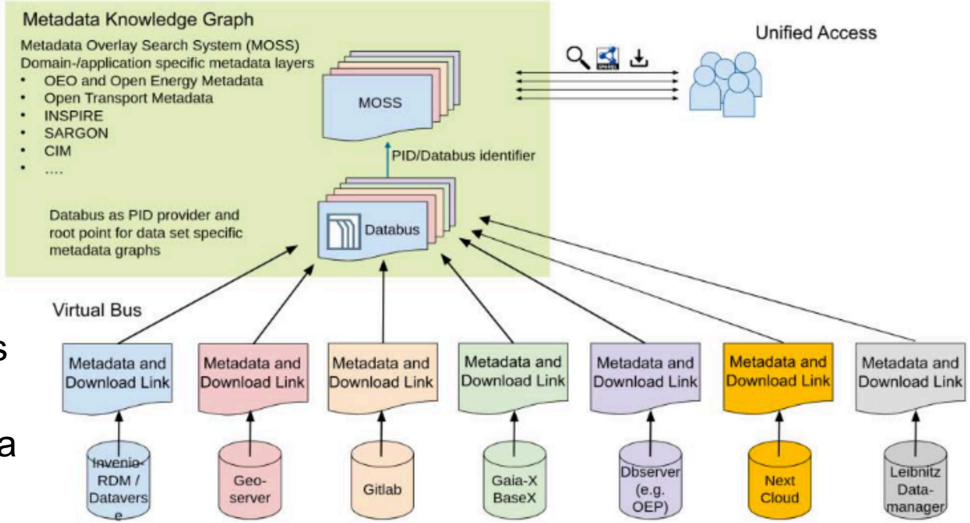
Fundamentals of Scientific Metadata for Energy

28 & 29-04-2025
Anis Koubaa

<HMC>

OEP, the Databus and the Metadata Overlay Search System

- Tables are released automatically to Databus
- additionally related metadata has been released to MOSS
- Databus and MOSS provide an advanced, flexible Cataloging System (CS)
- The CS is designed to meet FAIR data management standards across domains.
- By creating a unified metadata registry, Databus and MOSS enable researchers to add, refine, and search rich metadata records, ensuring data accessibility and interoperability across diverse repositories and research fields.


Metadata Knowledge Graph
Metadata Overlay Search System (MOSS)
Domain-/application specific metadata layers
• OEO and Open Energy Metadata
• Open Transport Metadata
• INSPIRE
• SARGON
• CIM
• ....
MOSS
PID/Databus identifier
Unified Access
Databus as PID provider and root point for data set specific metadata graphs
Databus
Virtual Bus
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Invenio-RDM / Dataverse
Geo-server
Gitlab
Gaia-X BaseX
Dbserver (e.g. OEP)
Next Cloud
Leibnitz Data-manager

OEP, the Databus and the Metadata Overlay Search System

- Data sets are registered on the Databus to create a PID and therefore a root of a metadata graph

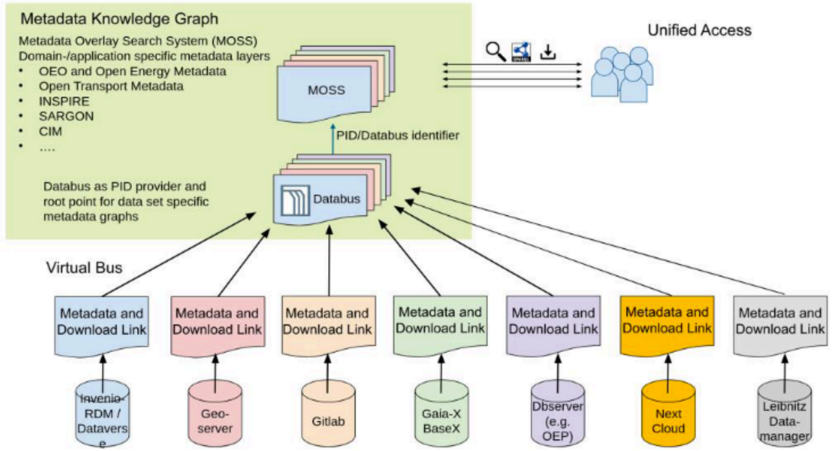
`https://example.org/janfo/energy/turbines/2022-02-02`

Base URL	User	Group	Artifact	Version

The identifiers of your metadata entries on the Databus are a composite the following parts:

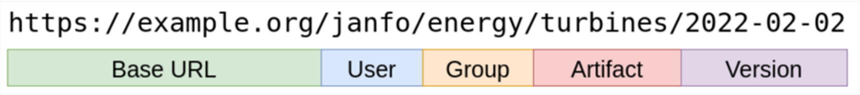
1. The **Databus base URI** *(e.g. https://databus.openenergyplatform.org)*
2. Your **account name** for that Databus *(e.g. janfo)* The username of the dataset owner on the particular Databus instance.
3. The **group name** *(e.g. energy)* The id/name of the published group.
4. The **artifact name** *(e.g. turbiunes)* The id/name of the published artifact.
5. The **version** *(e.g. 2022-02-02)* The version of the published data.
6. The **distribution name** *(e.g. cats.ttl.bz)* The id of the published data distribution.

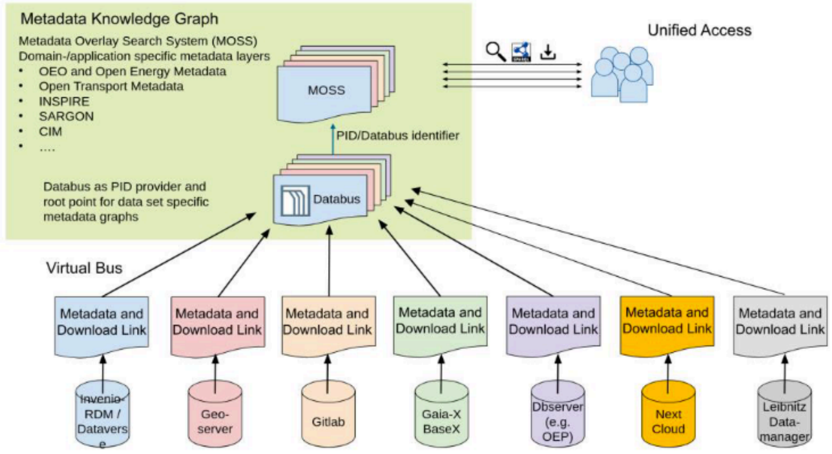
How MOOS looks like ?


Metadata Knowledge Graph
Metadata Overlay Search System (MOSS)
Domain-/application specific metadata layers
• OEO and Open Energy Metadata
• Open Transport Metadata
• INSPIRE
• SARGON
• CIM
• ....
MOSS
PID/Databus identifier
Databus as PID provider and root point for data set specific metadata graphs
Databus
Unified Access
Virtual Bus
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Invenio-RDM / Dataverse
Geo-server
Gitlab
Gaia-X BaseX
Dbserver (e.g. OEP)
Next Cloud
Leibnitz Data-manager

OEP, the Databus and the Metadata Overlay Search System

- The Databus (as well as MOSS) is committed to the same open standards as the tooling of many research data management approaches:
  - Knowledge Graphs,
  - SPARQL,
  - Linked Data,
  - DCAT & DCATAP,
  - SHACL, JSON-LD, RDF.
- Thus, it is highly interoperable and complementary to the already developed services that rely on the same standard as namely LDM, ORKG, TIB Terminology and PID services.

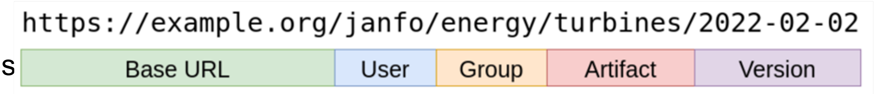




Metadata Knowledge Graph
Metadata Overlay Search System (MOSS)
Domain-/application specific metadata layers
• OEO and Open Energy Metadata
• Open Transport Metadata
• INSPIRE
• SARGON
• CIM
• ....
MOSS
PID/Databus identifier
Databus as PID provider and root point for data set specific metadata graphs
Databus
Unified Access
Virtual Bus
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Invenio-RDM / Dataverse
Geo-server
Gitlab
Gaia-X BaseX
Observer (e.g. OEP)
Next Cloud
Leibniz Data-manager

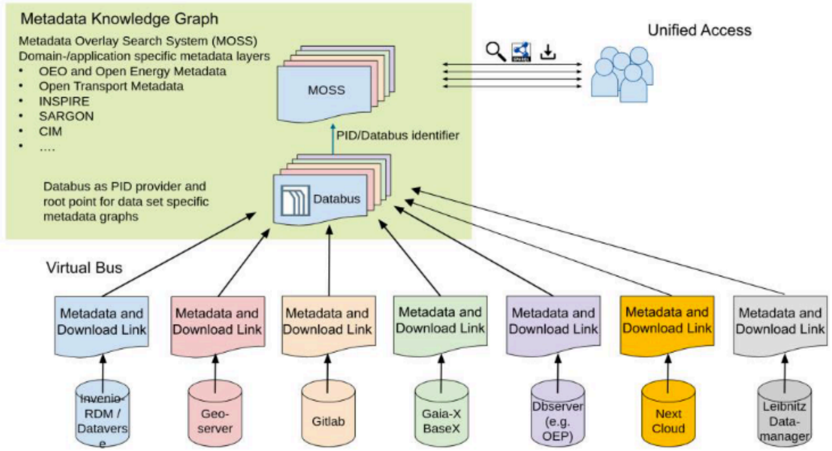
OEP, the Databus and the Metadata Overlay Search System

- MOSS allows users to annotate data sets from many repositories central place that can later be part of a larger platform.
- Initial metadata layers will focus on existing standards such as OEO and Open Energy Metdata 2.0, new layers can be added over time including an ORKG layer or others.
- Subgraphs are versioned in Git allowing traceability and versioning of metadata edits following a Wiki workflow.
- Adherence to RDF and SPARQL standard makes the collected metadata in MOSS interchangeable with LDM, federated search and other Knowledge Graph approaches.



```
https://example.org/janfo/energy/turbines/2022-02-02
```

Base URL	User	Group	Artifact	Version



Metadata Knowledge Graph
Metadata Overlay Search System (MOSS)
Domain-/application specific metadata layers
• OEO and Open Energy Metadata
• Open Transport Metadata
• INSPIRE
• SARGON
• CIM
• ....
MOSS
PID/Databus identifier
Databus as PID provider and root point for data set specific metadata graphs
Databus
Unified Access
Virtual Bus
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Metadata and Download Link
Invenio-RDM / Dataverse
Geo-server
Gitlab
Gaia-X BaseX
Observer (e.g. OEP)
Next Cloud
Leibnitz Data-manager

Hands on OEP

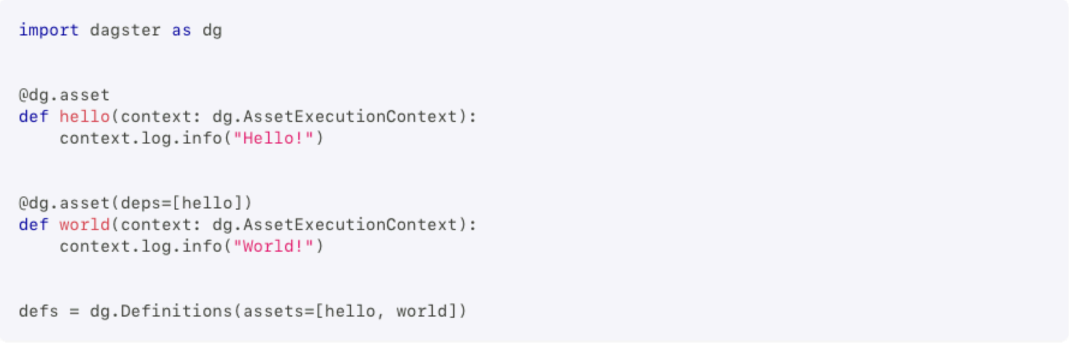
- Let’s publish some thing,
- Have a look on the content in the OEP Academy
- And download it programmatically
  - The oep-client README
  - Get your Token for the OEP
  - Think on having a specific Python Environment

Automation of Data Description

- Example data table:
  - https://openenergyplatform.org/dataedit/view/model_draft/tutorial_example_table
  - https://raw.githubusercontent.com/OpenEnergyPlatform/academy/production/docs/data/tutorial_example_table.data.json
- Documentation of the OEP Client
  - https://github.com/OpenEnergyPlatform/oep-client

Dagster

- Dagster is a data orchestrator built for data engineers, with
  - integrated lineage to provide a more complete picture of data flow,
  - observability, because broken pipelines are unavoidable, to catch problems as soon as they happen with the improved alerting suite
  - a declarative programming model,
  - and best-in-class testability.

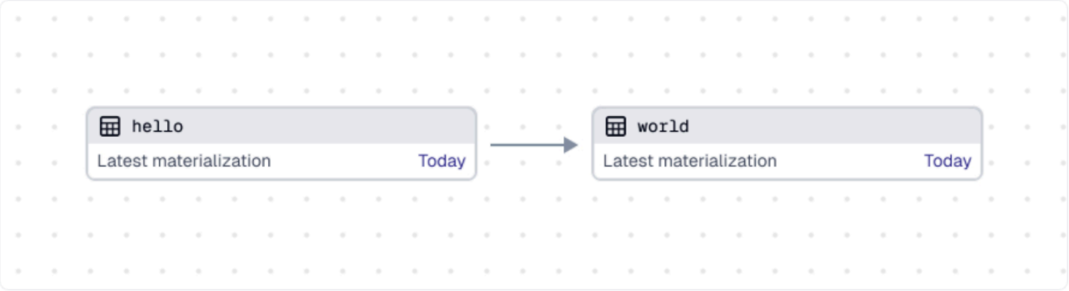


```python
import dagster as dg


@dg.asset
def hello(context: dg.AssetExecutionContext):
    context.log.info("Hello!")


@dg.asset(deps=[hello])
def world(context: dg.AssetExecutionContext):
    context.log.info("World!")


defs = dg.Definitions(assets=[hello, world])
```



End of Day 2

- Thanks for participating
- And think about it : Metadata is a love note to the future !