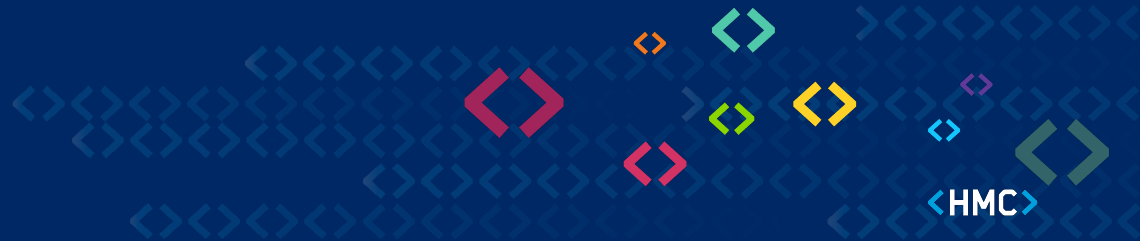


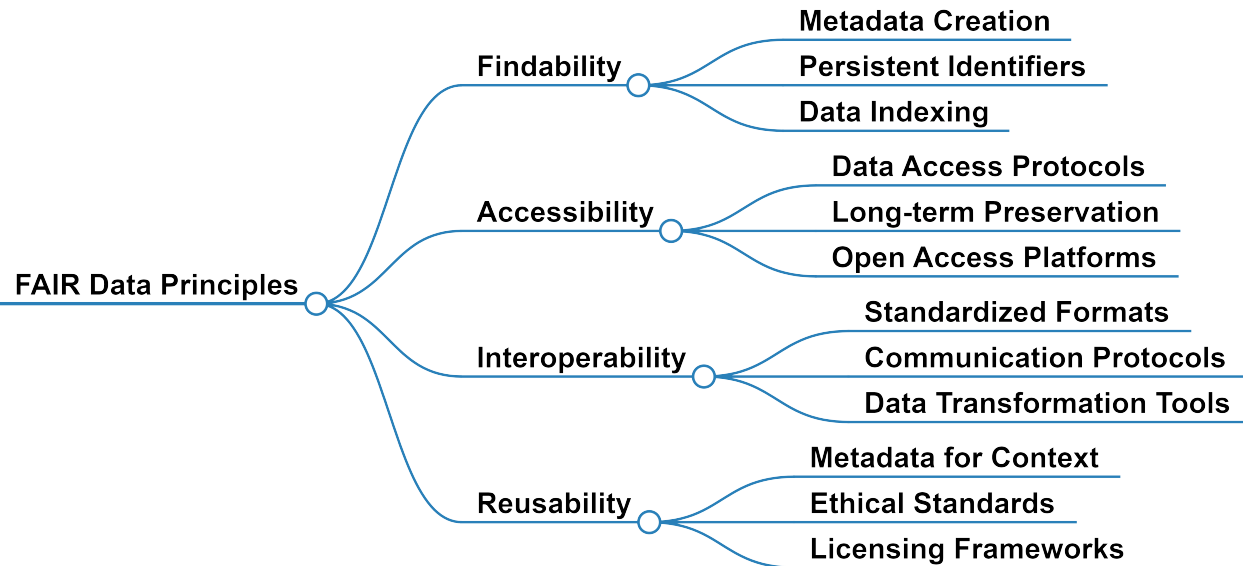
Fundamentals of Scientific Metadata for Energy

28 & 29-04-2025

Anis Koubaa



10:00	Fundamentals of Metadata <i>online</i>	<i>Mr Mohamed Anis Koubaa</i> 10:00 - 11:00
11:00	Hands On <i>online</i>	<i>Mr Mohamed Anis Koubaa</i> 11:00 - 11:30
	Break <i>online</i>	11:30 - 12:00
12:00	Controlled Vocabularies <i>online</i>	<i>Mr Mohamed Anis Koubaa</i> 12:00 - 12:45
	Platforms <i>online</i>	<i>Mr Mohamed Anis Koubaa</i> 12:45 - 13:15
13:00	Break <i>online</i>	13:15 - 13:35
	Automation, Part 1 <i>online</i>	<i>Mr Mohamed Anis Koubaa</i> 13:35 - 14:00
14:00		



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



- 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

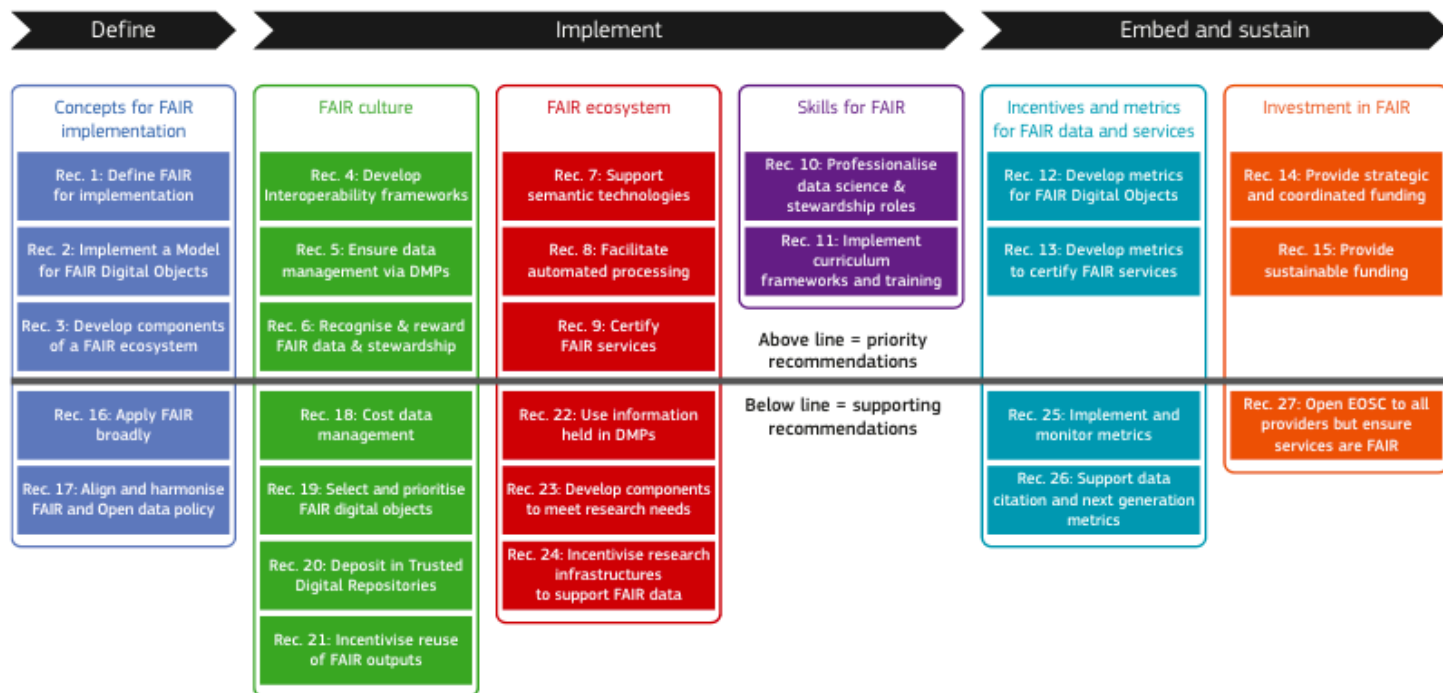
1.7.2 Step 2: Implement – culture, technology and skills for FAIR practice

- » 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

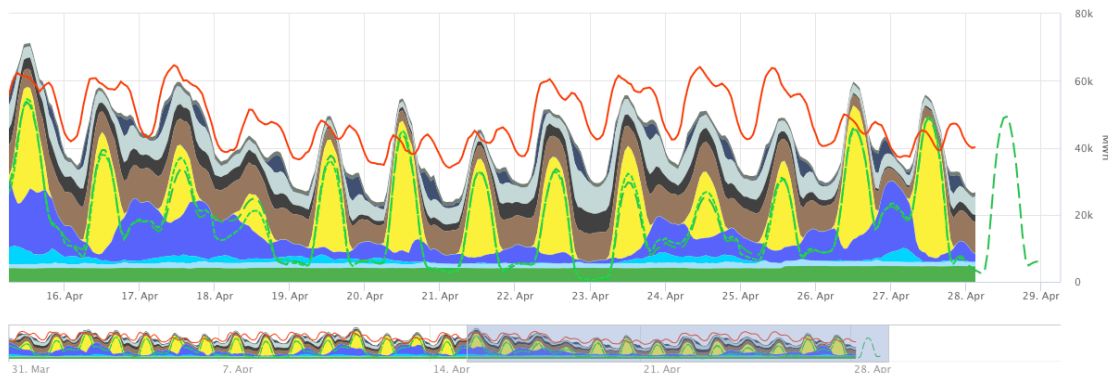
From: Final Report and Action Plan from the European Commission Expert Group on FAIR Data



Index to FAIR Action Plan recommendations

Actual generation, Forecasted generation day ahead and actual consumption

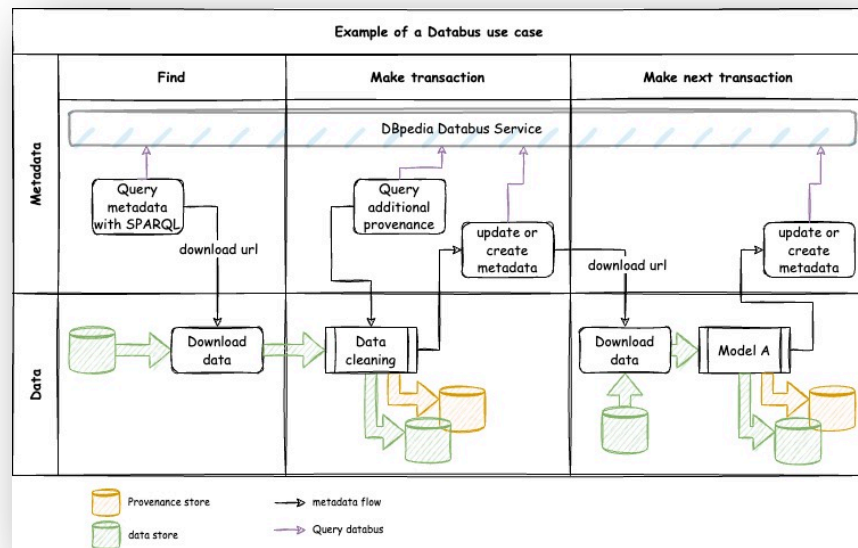
- Many data bases exist, each in its own flavour
 - Data access
 - Data format
 - Data licenses (if at all)
 - Sometimes hard to find
- Data collection is a labor intensive task
- Data cleaning, aggregation, etc. is repeated by many researchers with different results
- Data quality is often unknown



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

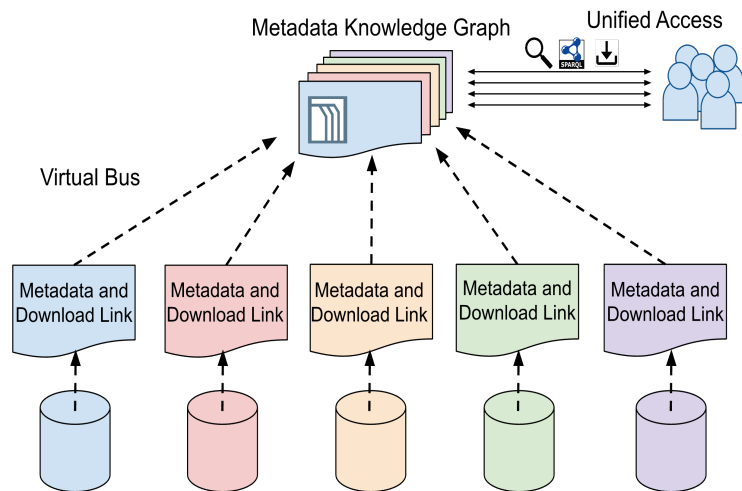
Example Application, when things work well

- 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



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?

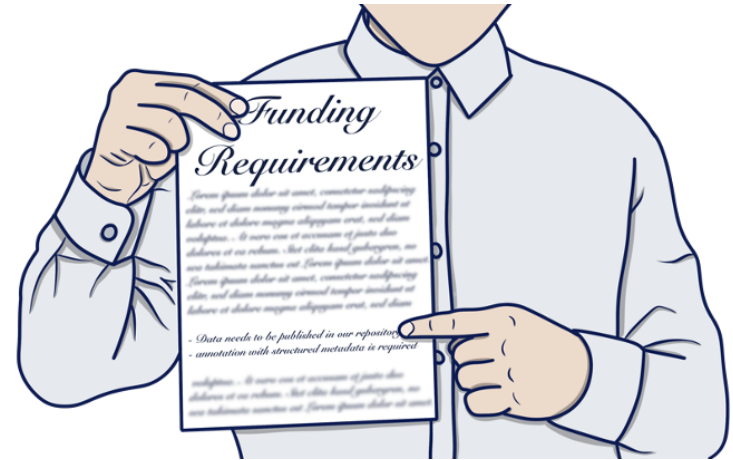
- Documentation and data quality

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?



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

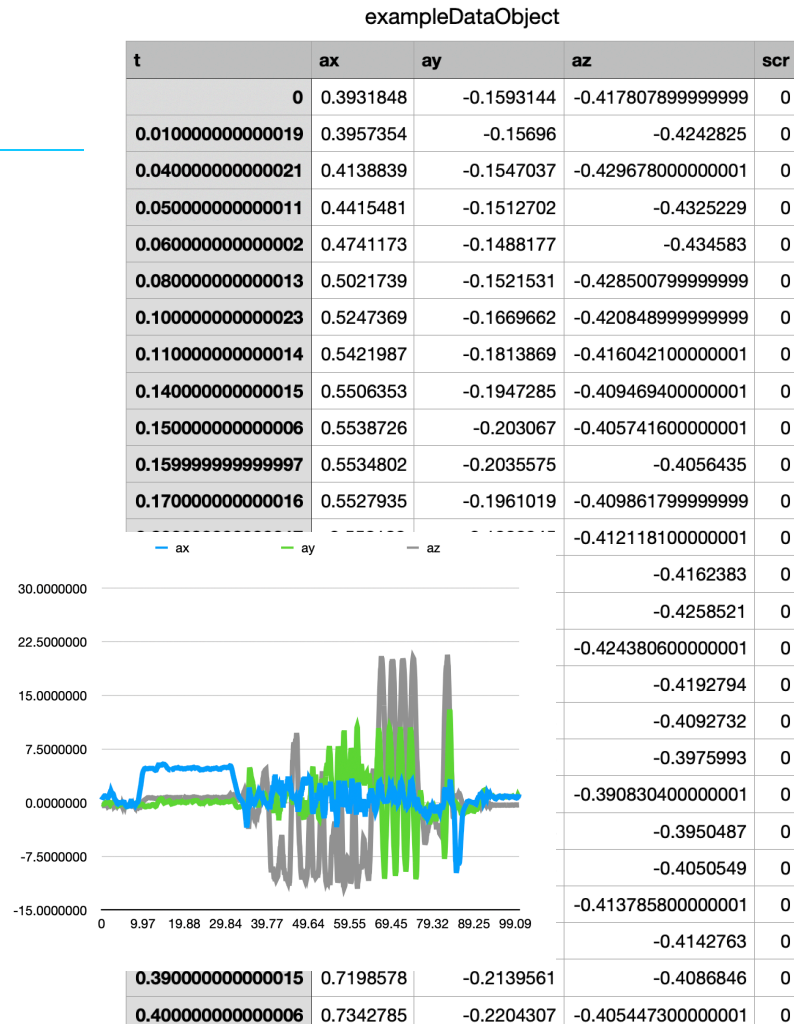
t	ax	ay	az	scr
0	0.3931848	-0.1593144	-0.417807899999999	0
0.01000000000000019	0.3957354	-0.15696	-0.4242825	0
0.04000000000000021	0.4138839	-0.1547037	-0.429678000000001	0
0.05000000000000011	0.4415481	-0.1512702	-0.4325229	0
0.06000000000000002	0.4741173	-0.1488177	-0.434583	0
0.08000000000000013	0.5021739	-0.1521531	-0.428500799999999	0
0.10000000000000023	0.5247369	-0.1669662	-0.420848999999999	0
0.11000000000000014	0.5421987	-0.1813869	-0.416042100000001	0
0.14000000000000015	0.5506353	-0.1947285	-0.409469400000001	0
0.15000000000000006	0.5538726	-0.203067	-0.405741600000001	0
0.1599999999999997	0.5534802	-0.2035575	-0.4056435	0
0.17000000000000016	0.5527935	-0.1961019	-0.409861799999999	0
0.20000000000000017	0.558189	-0.1908045	-0.412118100000001	0
0.21000000000000008	0.5764356	-0.1865862	-0.4162383	0
0.2199999999999999	0.589581	-0.18639	-0.4258521	0
0.25	0.6049827	-0.1941399	-0.424380600000001	0
0.26000000000000019	0.619992	-0.206991	-0.4192794	0
0.27000000000000001	0.6320583	-0.2191554	-0.4092732	0
0.30000000000000011	0.6392196	-0.2279844	-0.3975993	0
0.31000000000000002	0.6465771	-0.2317122	-0.390830400000001	0
0.32000000000000022	0.6583491	-0.2291616	-0.3950487	0
0.34000000000000003	0.6725736	-0.2220984	-0.4050549	0
0.36000000000000014	0.6905259	-0.216801	-0.413785800000001	0
0.37000000000000005	0.7047504	-0.2139561	-0.4142763	0
0.39000000000000015	0.7198578	-0.2139561	-0.4086846	0
0.40000000000000006	0.7342785	-0.2204307	-0.405447300000001	0

Steps towards well described data

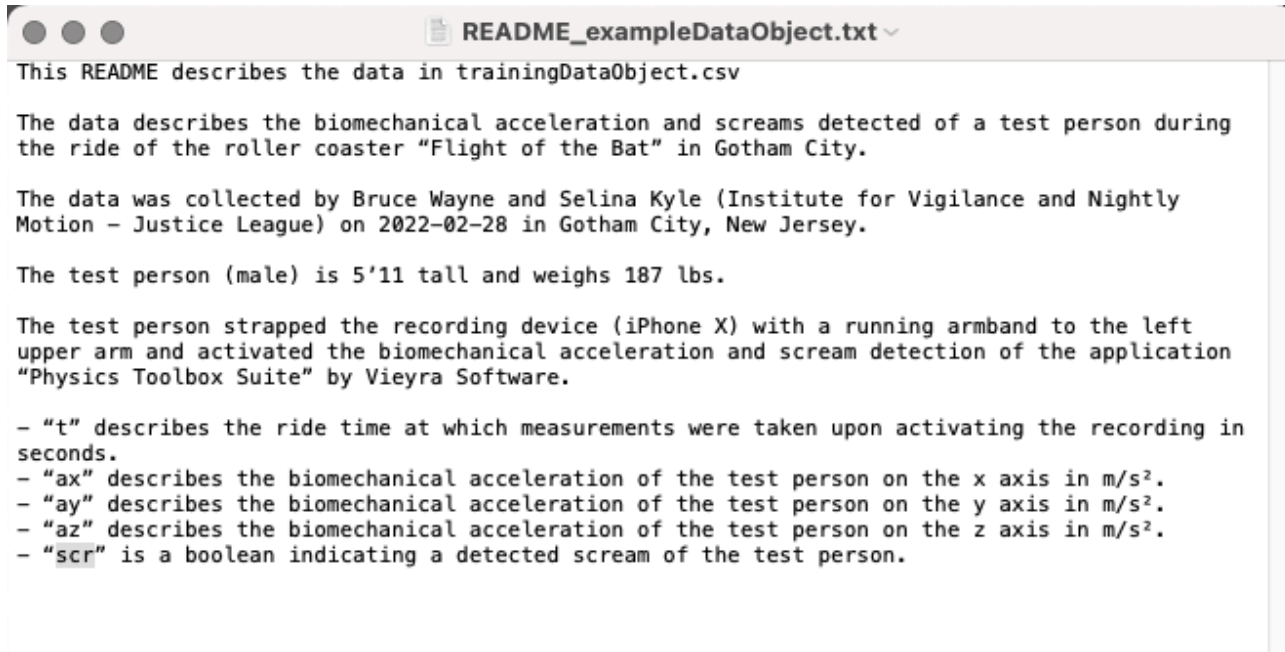
Enriching data with metadata is a key concept for the data output of scientific research to be FAIR.

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.



Steps towards well described data



```
README_exampleDataObject.txt
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.

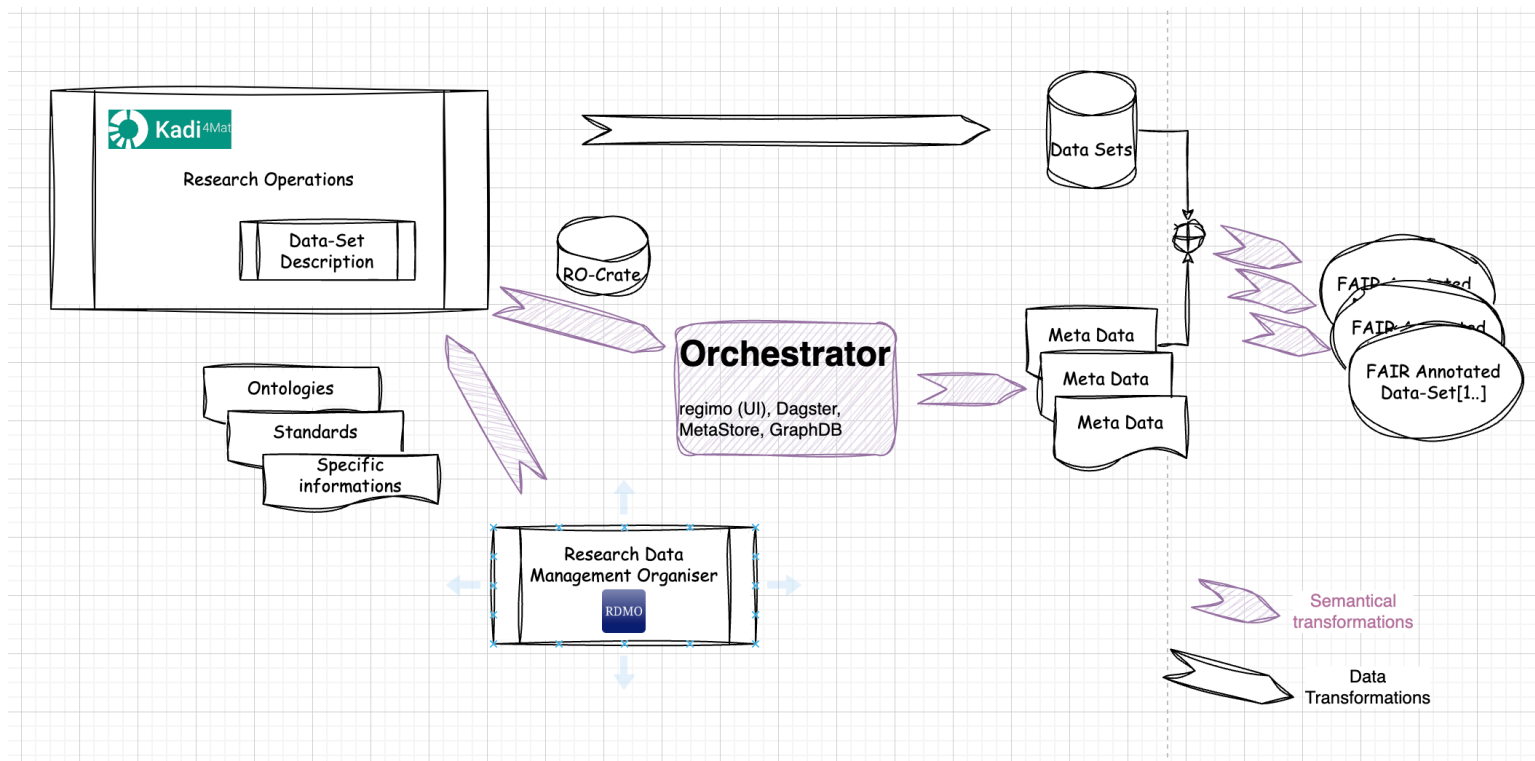
The data was collected by Bruce Wayne and Selina Kyle (Institute for Vigilance and Nightly Motion - Justice League) on 2022-02-28 in Gotham City, New Jersey.

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/s².
- "ay" describes the biomechanical acceleration of the test person on the y axis in m/s².
- "az" describes the biomechanical acceleration of the test person on the z axis in m/s².
- "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



+

-

LOWER SAXONY

NETHERLANDS

The Hague

Rotterdam

Antwerp

Brussels

BELO

DE-CE

E-CE

Luxembourg

SAARLAND

Nuremberg

BAVARIA

BALEN-WÜRTTEMBERG

Munich

VIENNA

Bratislava

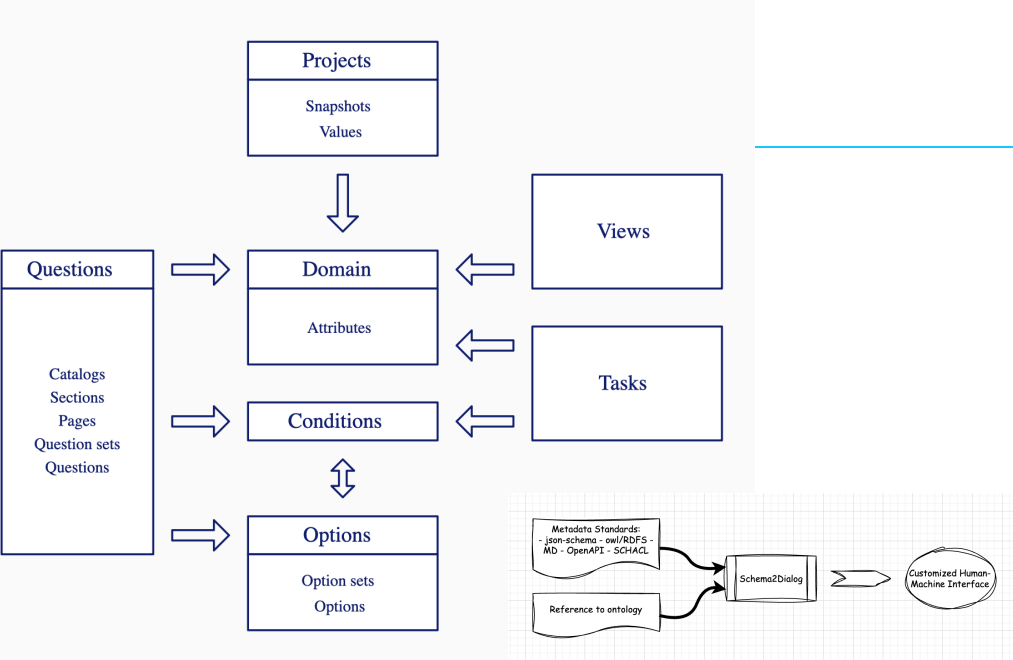
Karlsruher Institut für Technologie (KIT)

URL

<https://rdmo.forschungsdaten.info/>

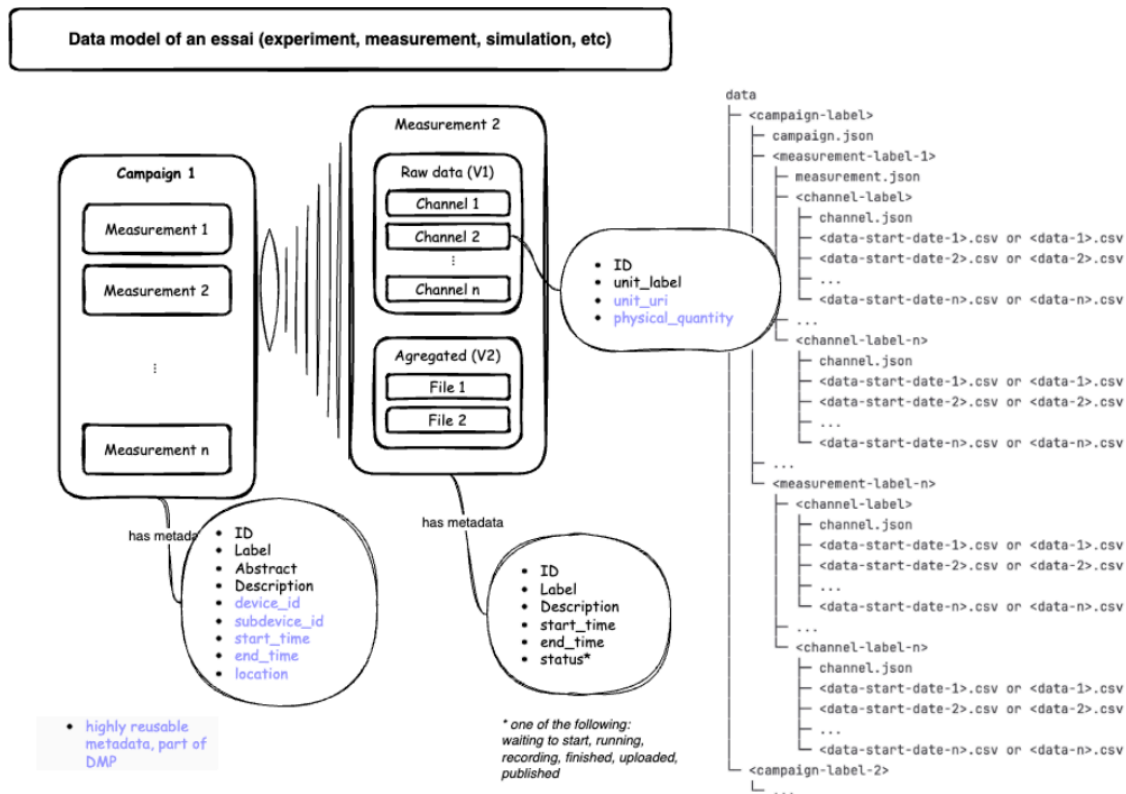
Kontakt

Kerstin Vanessa Wedlich-Zachodin



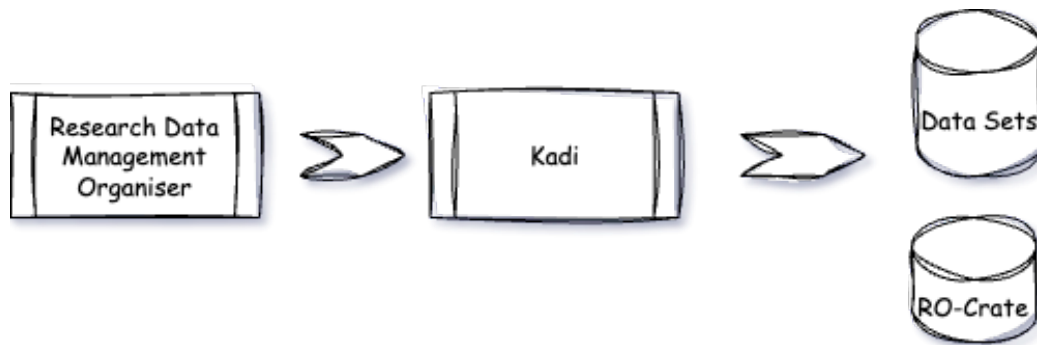
Overview of the RDMO data model

Data Model of an essai



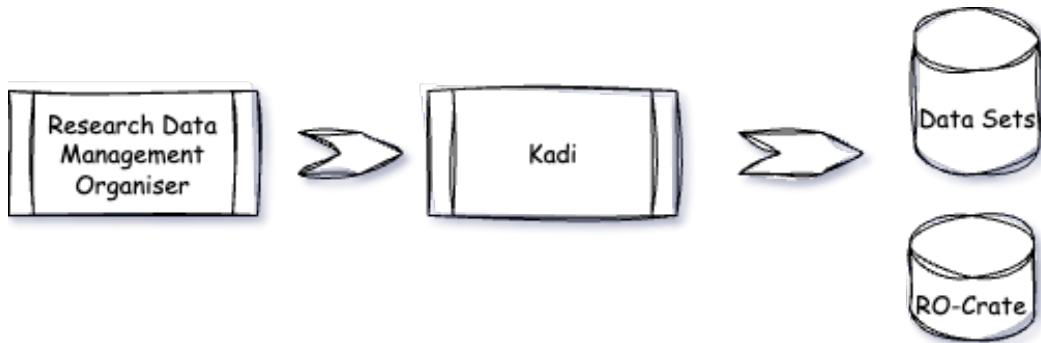
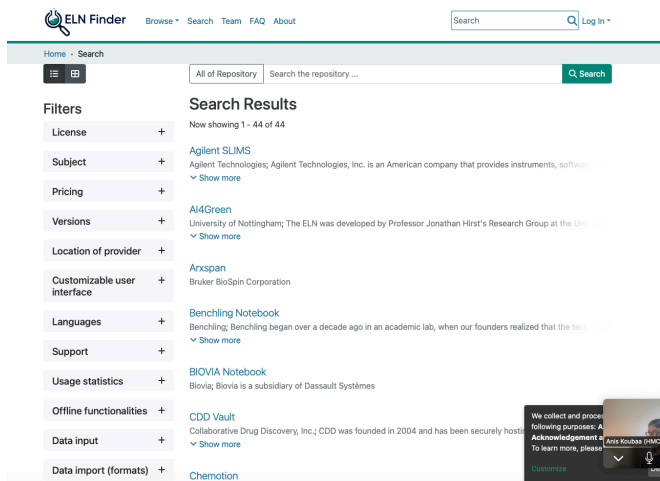
Recommendation 2 : Timing

- 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.



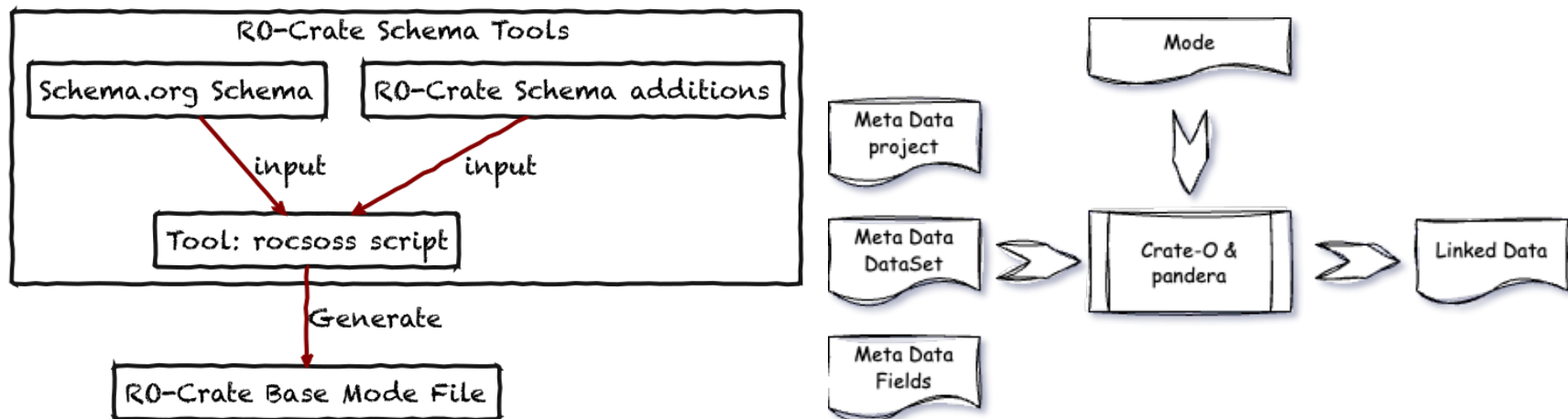
Recommendation 2 : Timing

- 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.



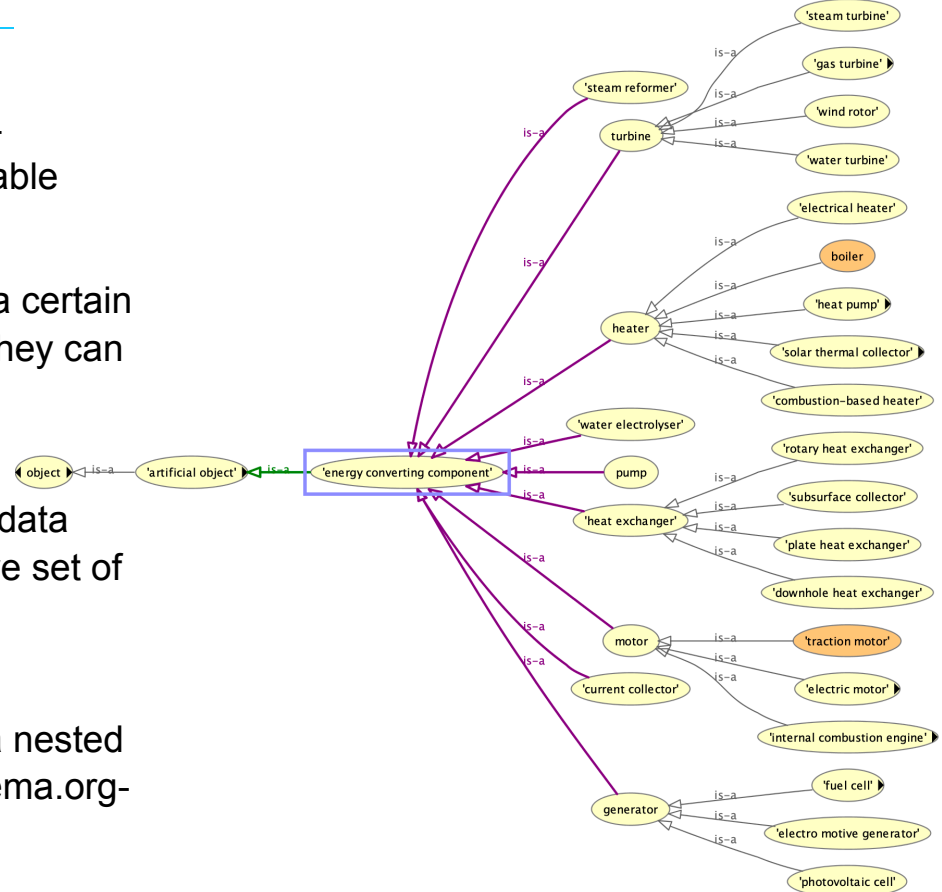
Recommendation 3 : Metadata Records Editing

- 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 (**SBE0**) 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

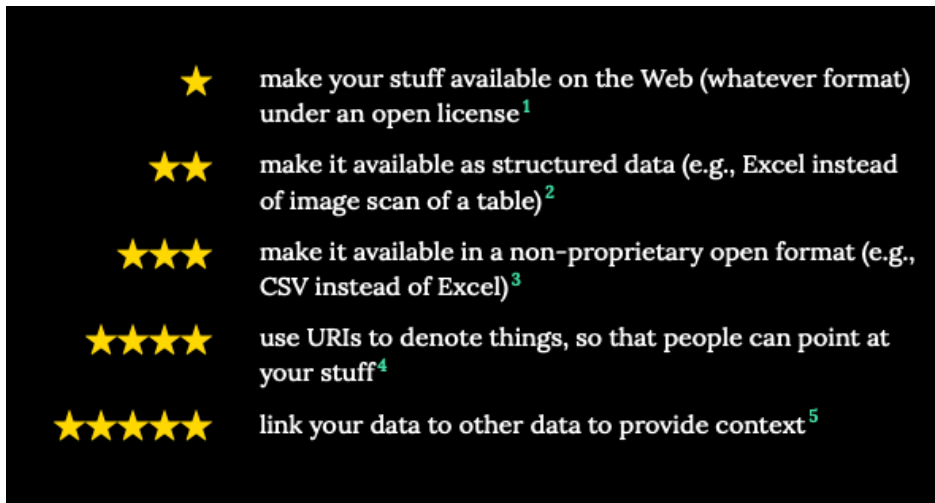
- A metadata standard for „energy related data“
- Based on existing technologies and standards as “Frictionless Data” and “DataCite”
- Implemented as JSON-LD to be human and machine readable
- Latest release (v2.0) is “ontology ready”

- **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

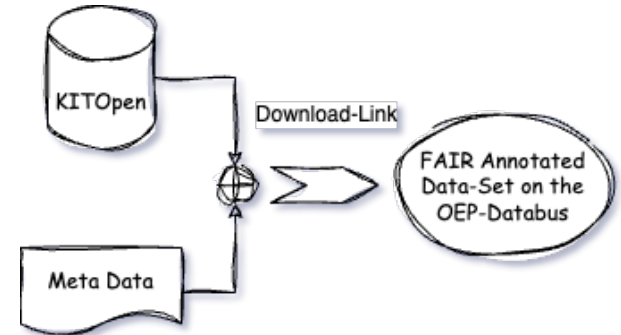
- A metadata standard for „energy related data“
- Based on existing technologies and standards as “Frictionless Data” and “DataCite”
- Implemented as JSON-LD to be human and machine readable
- Latest release (v2.0) is “ontology ready”
- Target: 5-star Linked Open Data



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

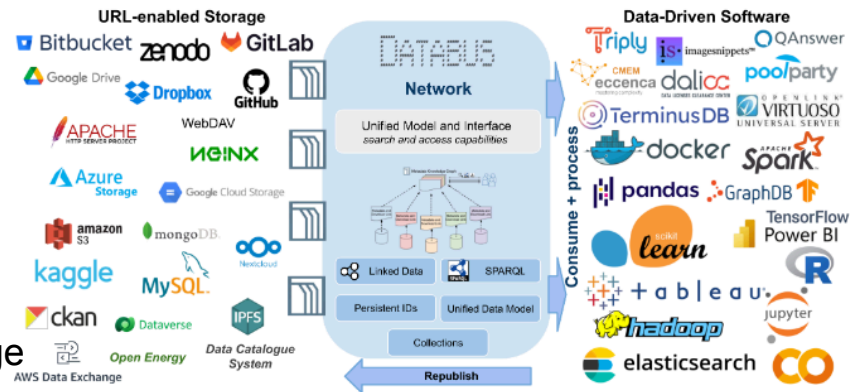
Recommendation 5 : Publish Metadata

- Make your metadata discoverable and accessible by publishing in a searchable resource (e.g. trustworthy repository) even if the data themselves 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 reconstructable 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.


Database



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.


Scenario Bundles



Ontology

Ontology refers 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.

Ontology

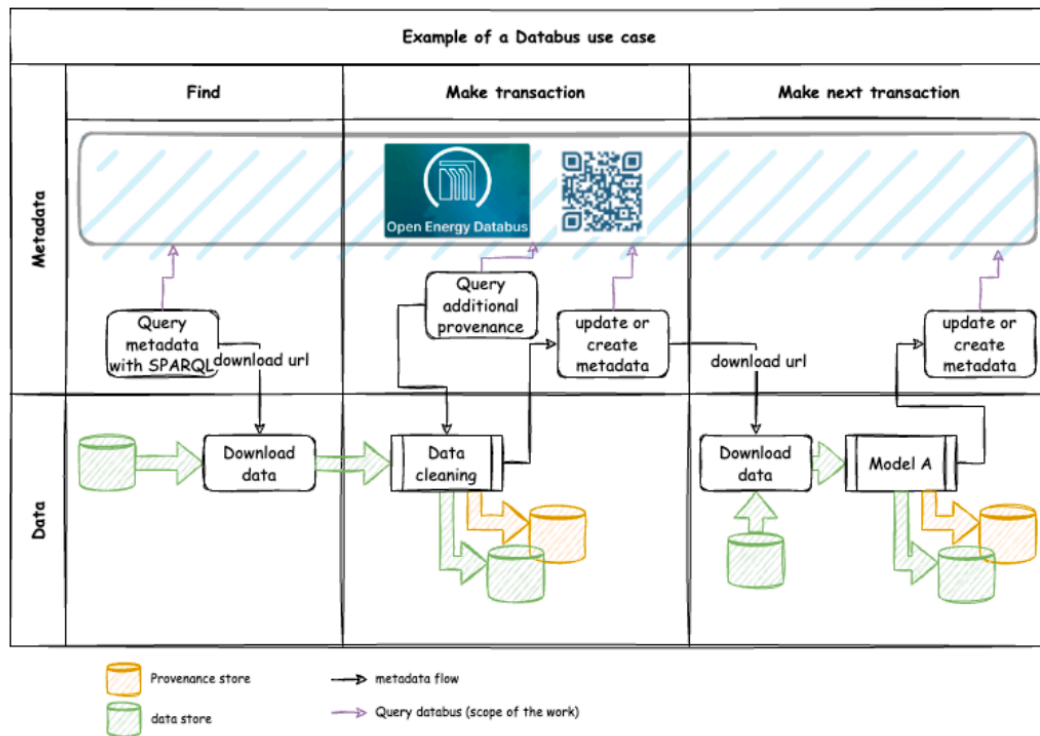


Academy

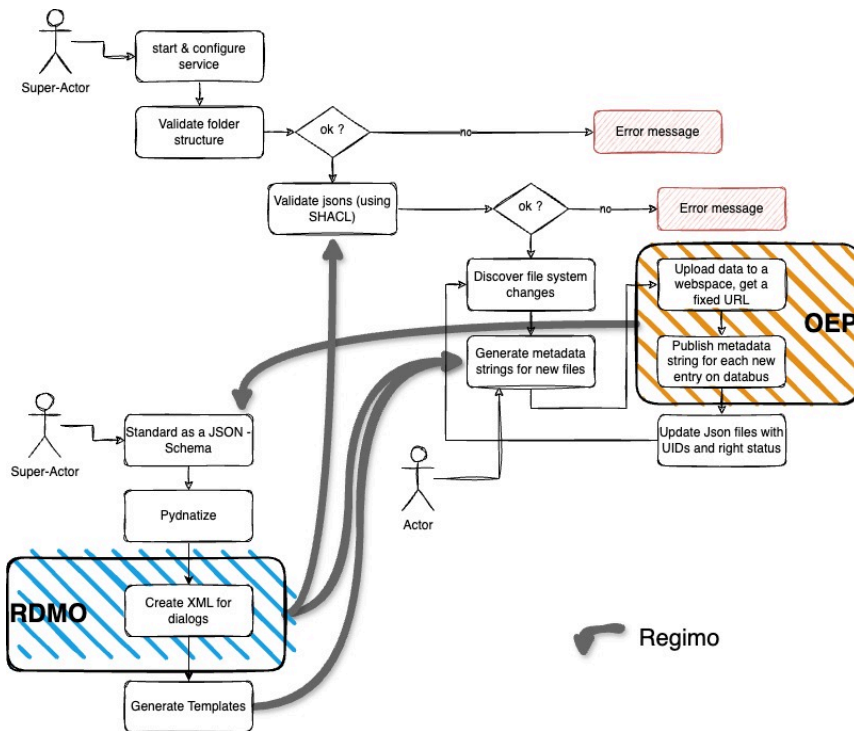
The Open Energy Academy (OEA) provides courses as well as dedicated tutorials covering important topics around the Open Energy Family (OEF) tools and the Open Energy Platform (OEP). You will also find short answers to urgent questions.

Academy

The Open Energy Platform - Databus use case



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

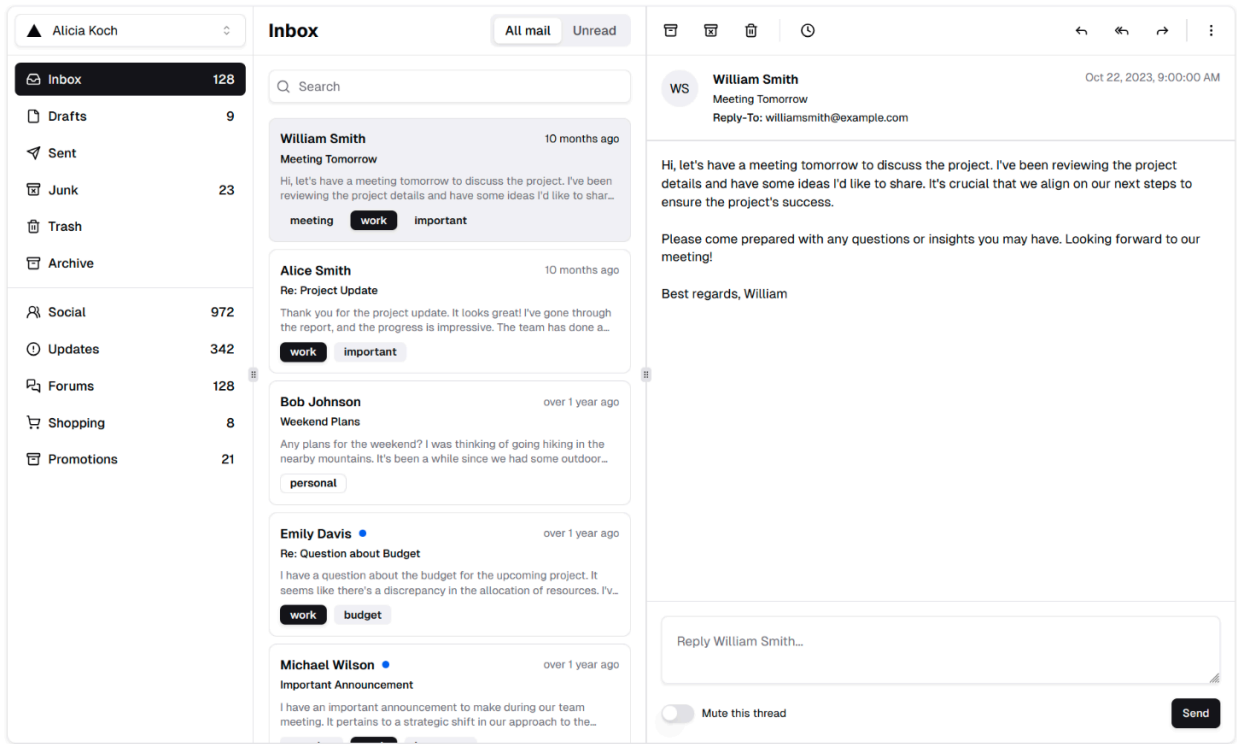


Examples of Activities within IAI

- 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

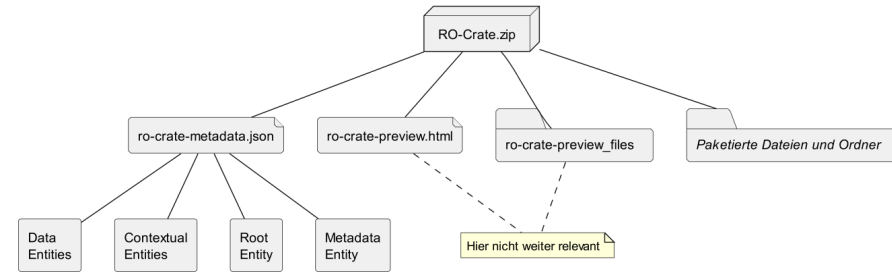
► ***General usable Concept for IAI***

Edit & Send

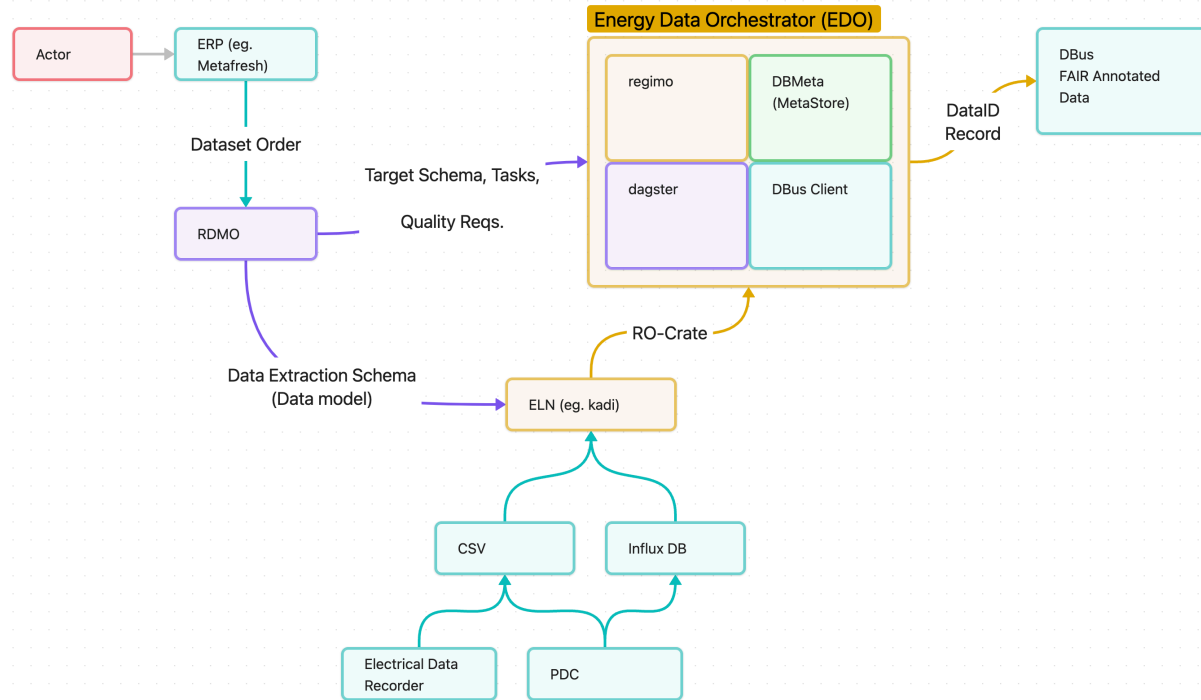


RO-Crate

- 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)