Slides: https://tinyurl.com/lswt-2024-databus

LOD-GEOSS

Databus: FAIR Data in Energy Systems Analysis

Carsten Hoyer-Klick[1], Markus Blesl[2], Lüder von Bremen[1], Johannes Frey[3], Ulrich Frey[1], Anastasis Gainnousakis[4], **Sebastian Hellmann**[3], Christian Hofmann[5], Ludwig Hülk[5], Stefan Kronshage[1], Patrick Kuckertz[6], Christoph Muschner[5], Michaja Pehl[4], Marion Schroedter-Homscheidt[1]

[1]German Aerospace Center, Institute of Networked Energy Systems,
[2]Stuttgart University, Institute for Energy Economics and Rational Energy Use,
[3]Institut für Angewandte Informatik e.V.,
[4]Potsdam Institute for Climate Impact Research
[5]Reiner Lemoine Institute,
[6]Forschungszentrum Jülich, Techno-economic Systems Analysis (IEK-3)

Supported by: Federal Ministry for Economic Affairs and Energy, on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center · Reiner Lemoine Institut · InfAI Institute for Applied Informatics · Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung · Potsdam Institute for Climate Impact Research · Forschungszentrum Jülich

DBpedia Databus - Data Catalog Platform

LOD-GEOSS

- **Metadata Matters**:
  - Data Catalogs and Data Metadata are often neglected (takes some extra effort)
  - Why do you need them?
  - Raising awareness:
    - You might not wonder, why your data projects are slow and mediocre quality, because you probably never worked with good metadata and a good catalog

Data catalogs offer a high potential for innovation, i.e.

realize faster, cheaper & better data projects

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

DBpedia Databus - Data Catalog Platform

LOD-GEOSS

› DBpedia Databus
  - a lightweight, scalable, adaptable, powerful Data Catalog Platform
  - Open Source - https://github.com/dbpedia/databus
  - addresses the lack of Open Source data catalog software

› DBpedia Databus is deployed in these areas
  - **Supply Chain Management**
  - **Engineering & Industry 4.0**
  - **Knowledge Graphs & AI**
  - **Data Governance**
  - today: **Research Data Management**

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

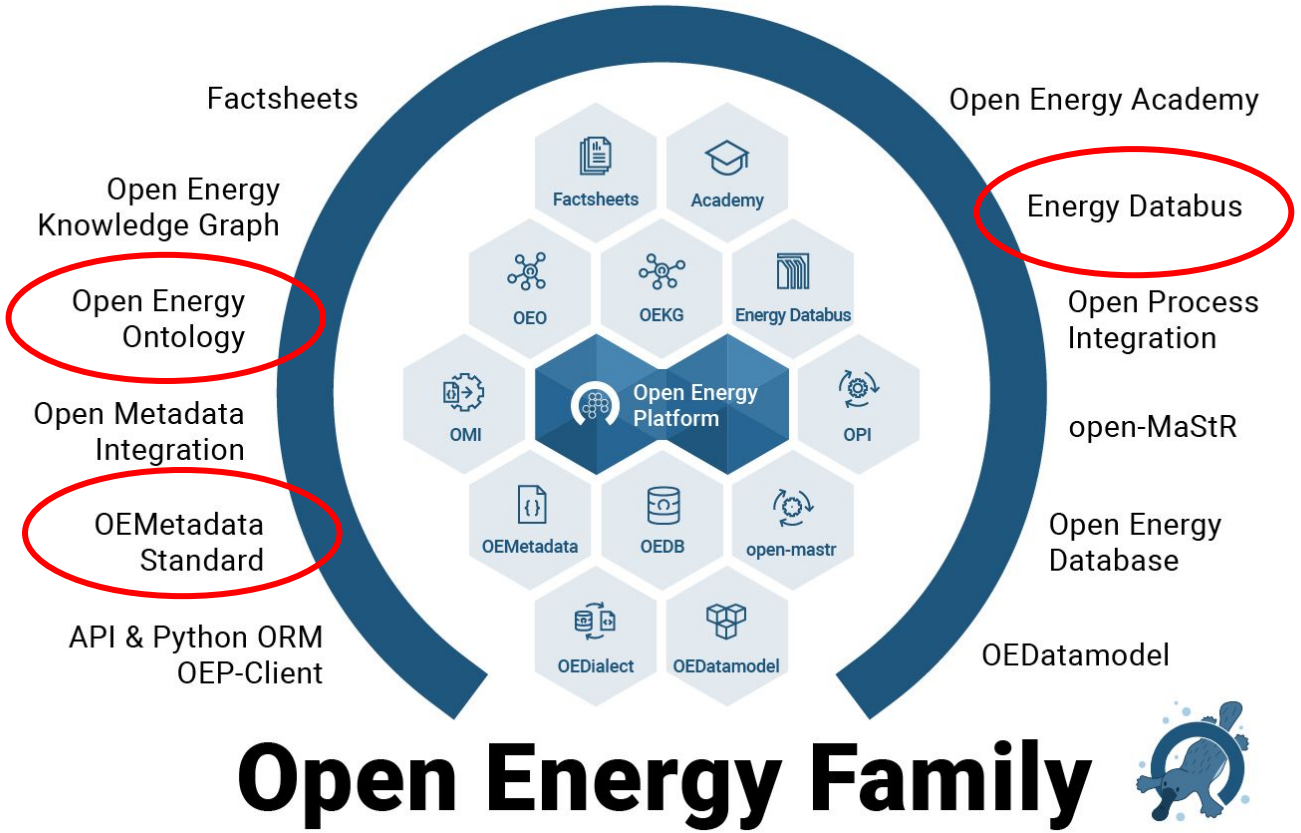
Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center · Reiner Lemoine Institut · InfAI Institute for Applied Informatics · Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung · Potsdam Institute for Climate Impact Research · Jülich Forschungszentrum

Open Energy Family

Findable Accessible Interoperable Reusable

LOD-GEOSS

- The **Open Energy Family** is an initiative for open and FAIR data in the domain of energy systems research
- Energy systems research develops models to evaluate and compare future energy scenarios like electricity demand and offer at the Energy Exchange Agency (EEX, Leipziger Strombörse)
- Development of a FAIR infrastructure within the Open Energy Family



https://openenergyplatform.org/

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

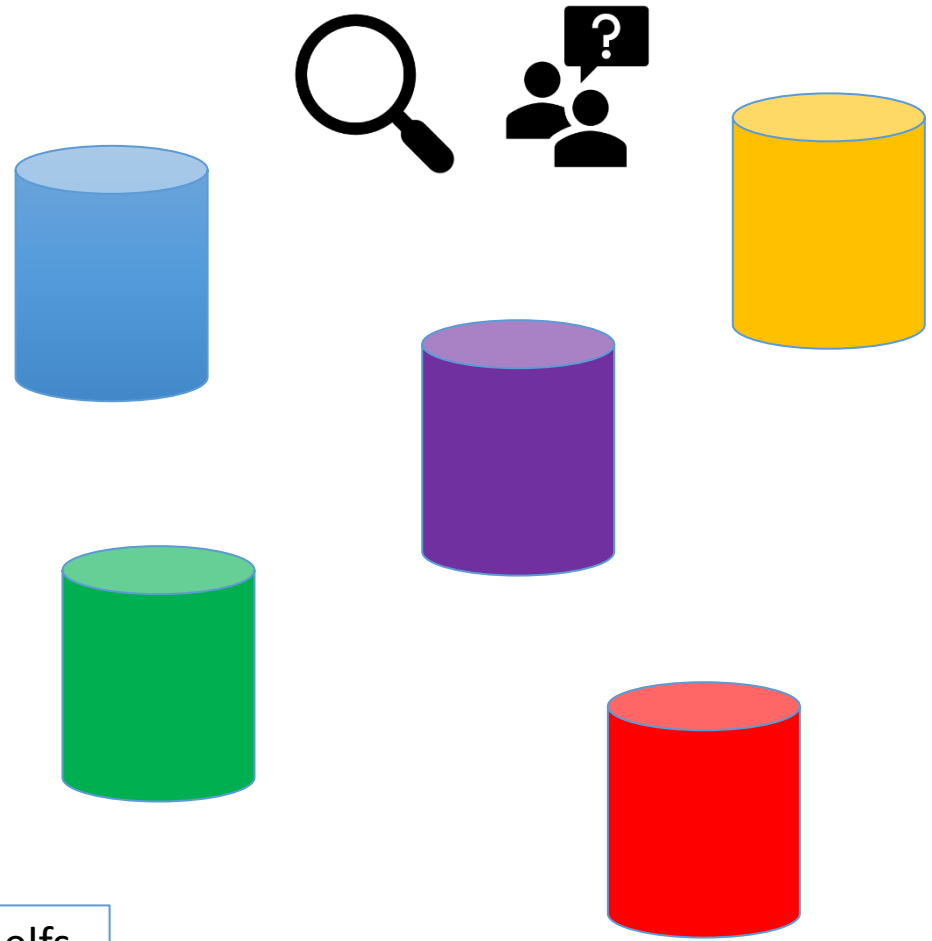
https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

DLR Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center · Reiner Lemoine Institut · InfAI Institute for Applied Informatics · Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung · Potsdam Institute for Climate Impact Research PIK · JÜLICH Forschungszentrum

Challenges I: Finding and Accessing Data

LOD-GEOSS

- Most of the research datasets are already online in energy systems research, but …
  - **Online availability (publication) ≠ Findability (Search)**



Pain points for energy systems researchers:

- Data collection is a labor intensive task
- Data cleaning, aggregation, etc.
  is repeated by many researchers with different results
- Data quality is often unknown

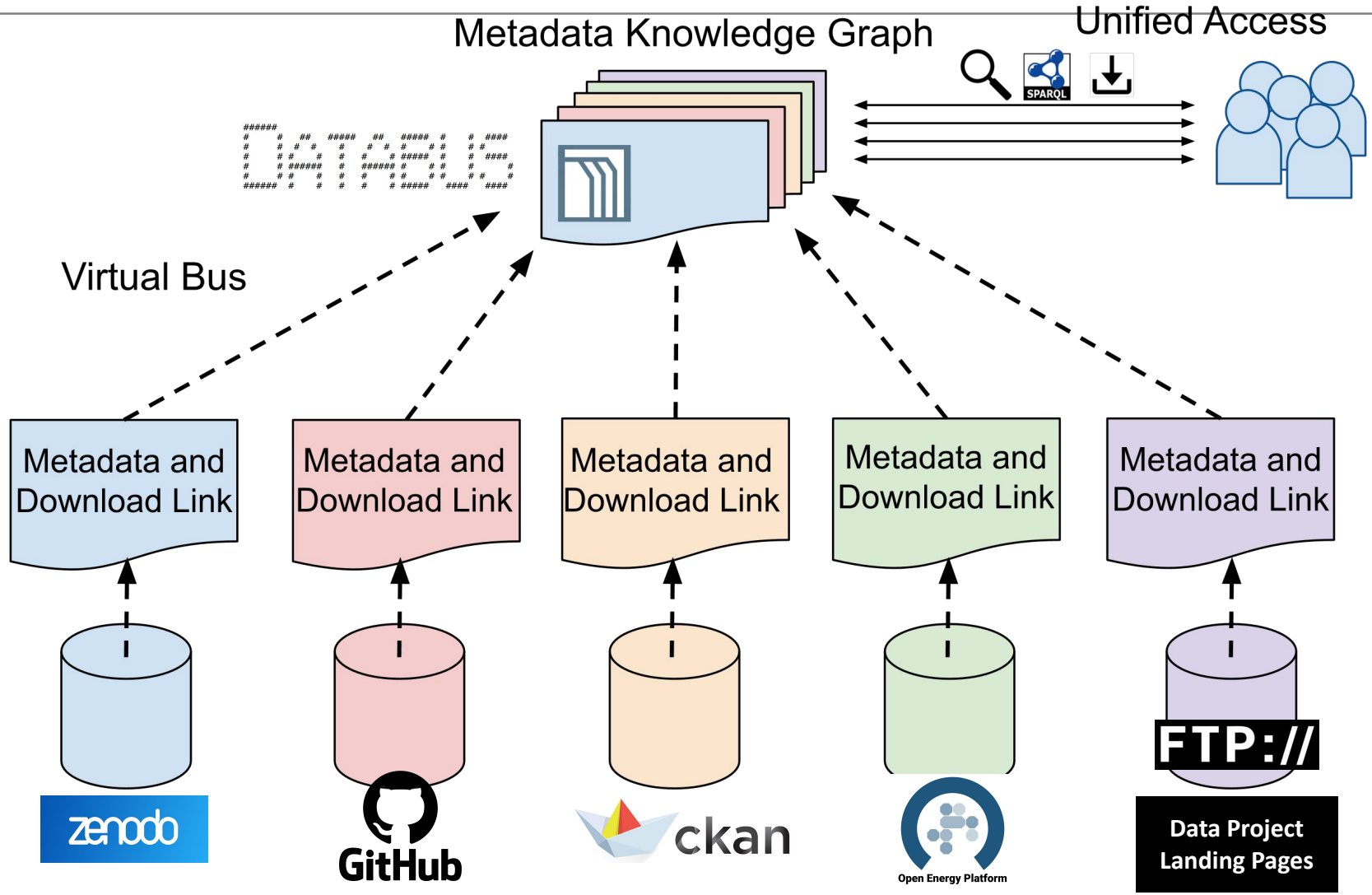
Imagine a library without a catalog and systematic numbers on the shelfs

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

Solution I: Semantic Metadata Catalog

- 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



Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

Solution II: Effective Search over Data (Findability ≠ Online Availability)

LOD-GEOSS

- **Effective search needs domain customization**
  - cf. “No Free Lunch in Search and Optimization” (NFL Theorems) Wolpert and Macready (1997)

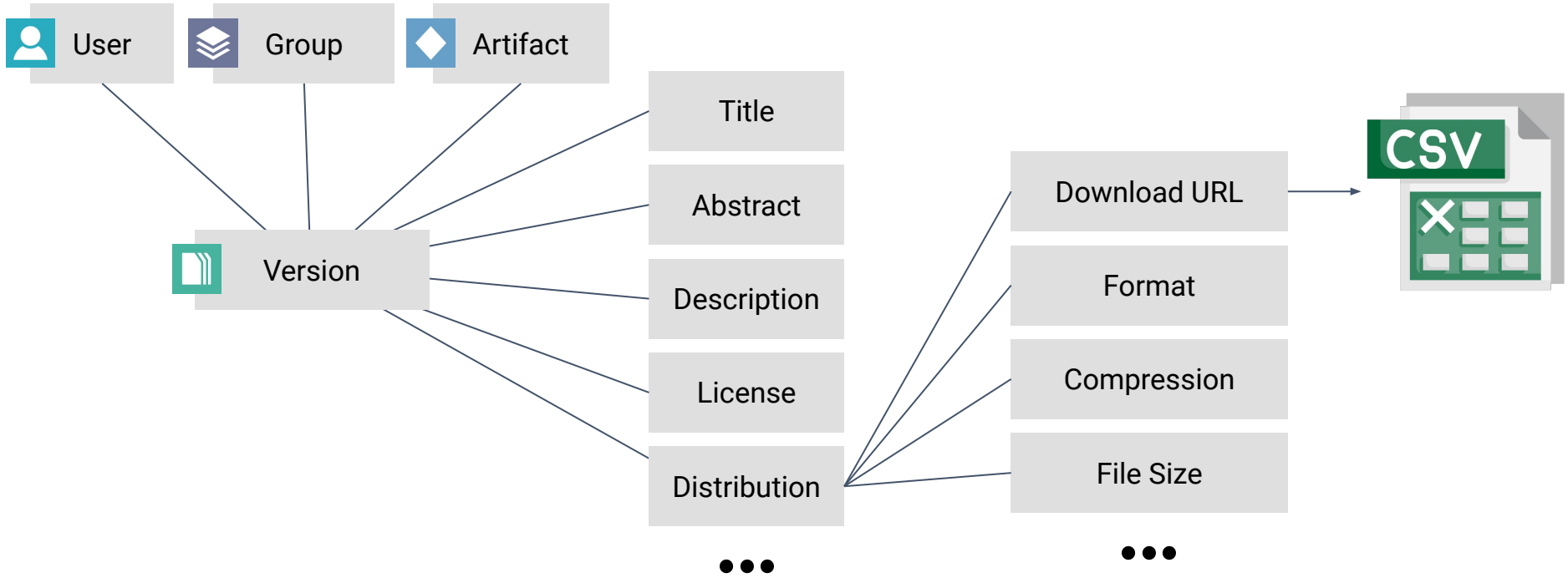
- Metadata Overlay Search System (MOSS)
  1. extends the Databus core graph with custom metadata subgraphs
  2. indexes these subgraphs for searching

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

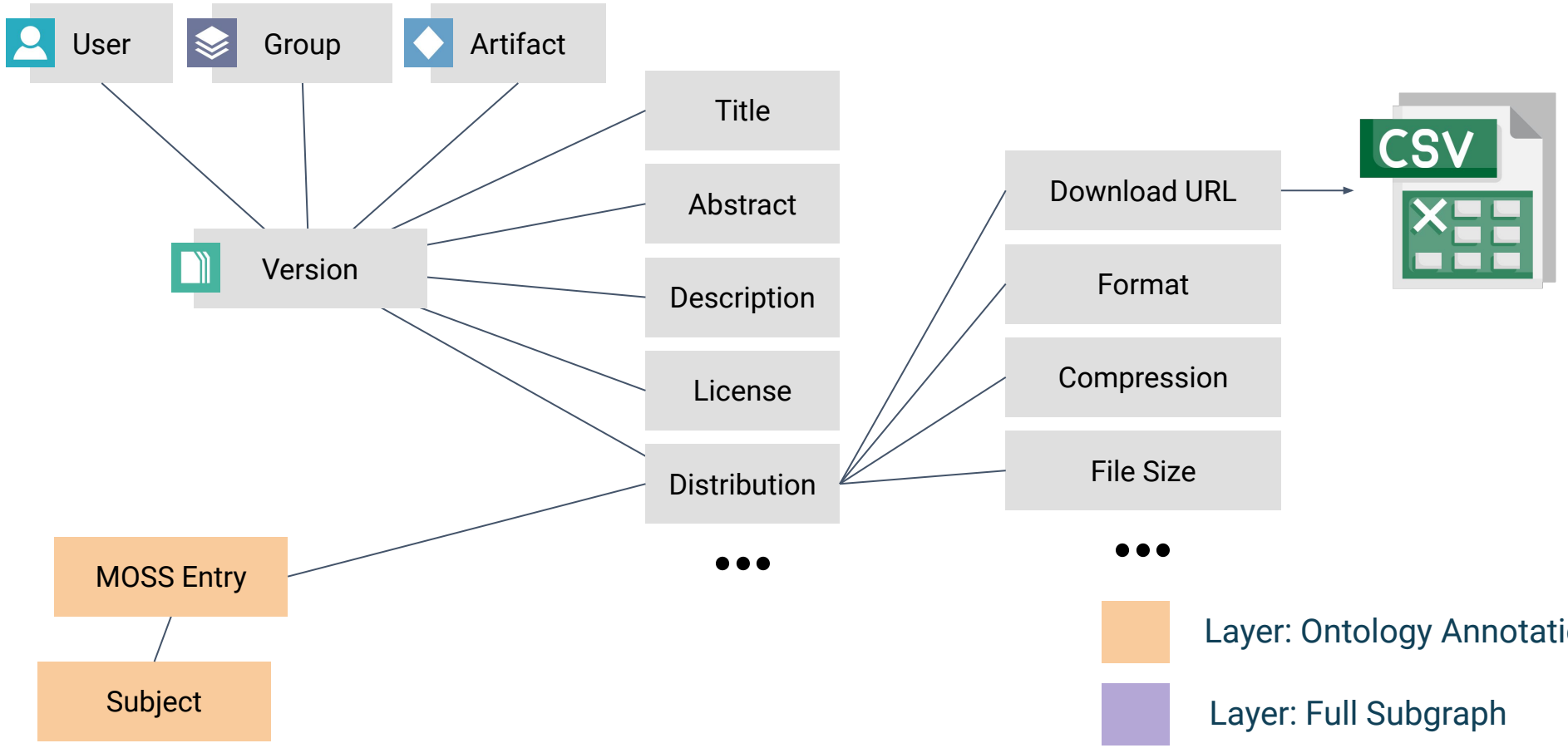
DLR Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center
REINER LEMOINE INSTITUT
InfAI Institute for Applied Informatics
Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung
PIK Potsdam Institute for Climate Impact Research
JÜLICH Forschungszentrum

Core Metadata (improves DCAT significantly)



https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

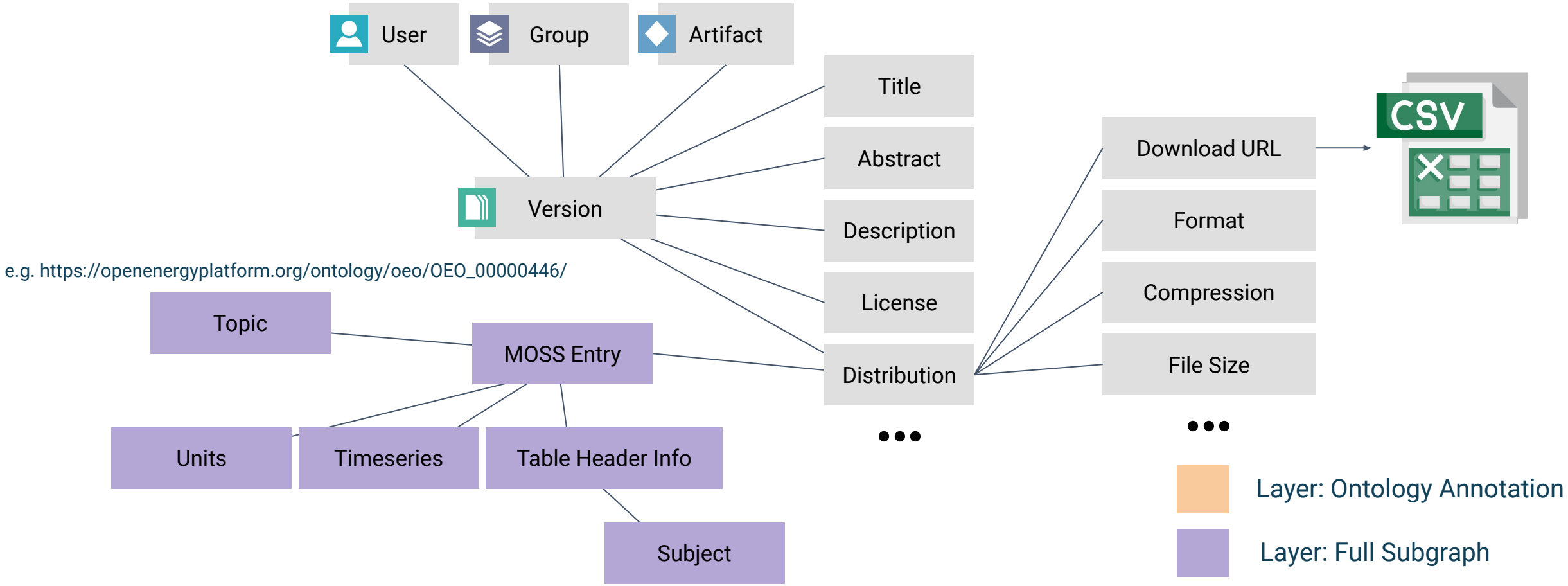
Metadata Subgraph Extensions (Example Layer 1)



OEO e.g.
https://openenergyplatform.org/ontology/oeo/OEO_00000446/

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

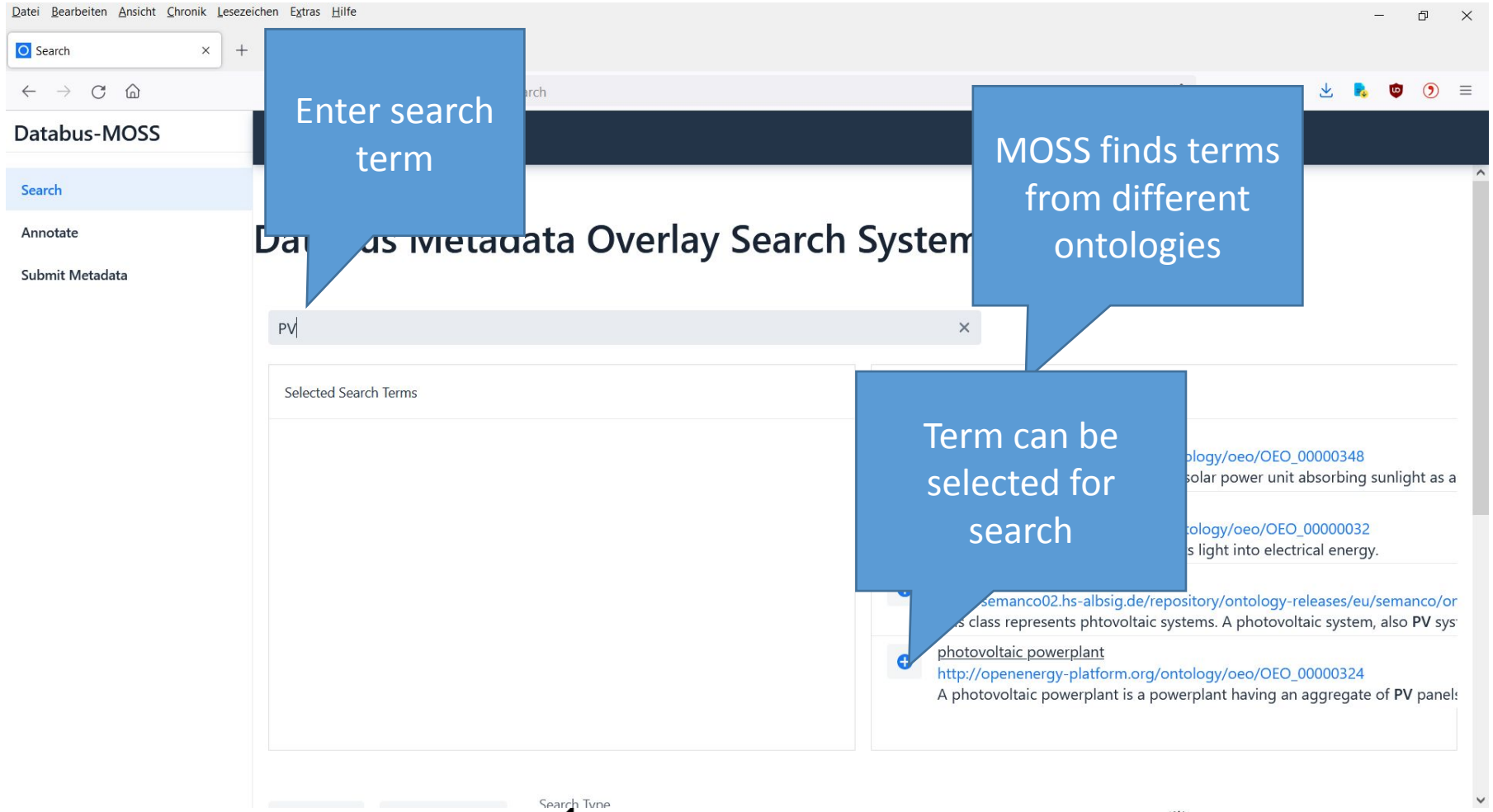
Metadata Subgraph Extensions (Example Layer 2)



https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

Searching data with the MOSS (Metadata Overlay Search System)

LOD-GEOSS

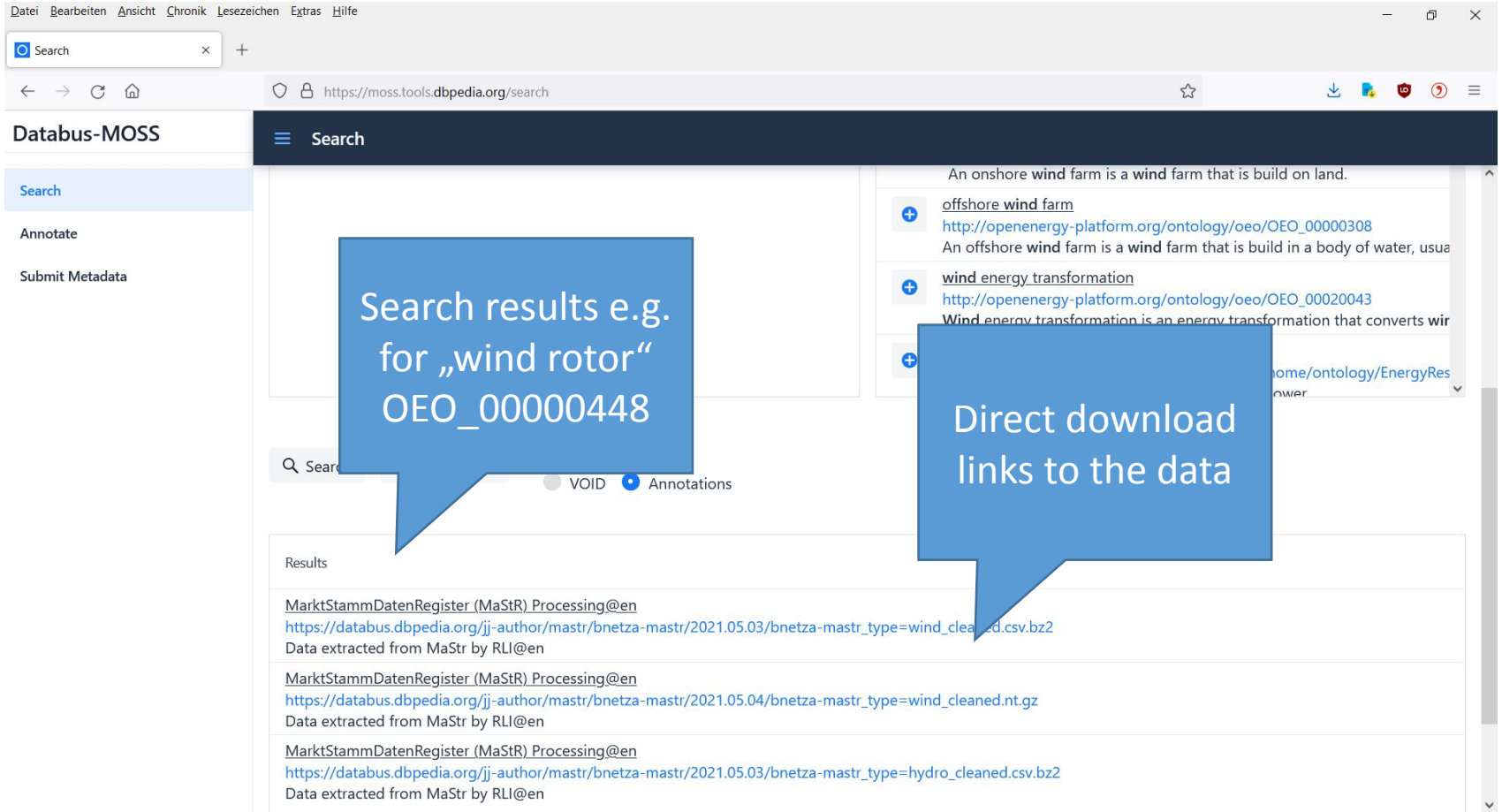


Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

DLR Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center
REINER LEMOINE INSTITUT
InfAI Institute for Applied Informatics
Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung
PIK Potsdam Institute for Climate Impact Research
JÜLICH Forschungszentrum

Searching with MOSS



Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

Features of MOSS

LOD-GEOSS

Engine:

- Built on the Databus Ontology and Prov-O
- Multiple parallel layers of metadata possible
- Graph Wiki tracks revisions (GStore)
- Data quality validated with SHACL
- Terminology service
- Graph indexer (Lucene/Elastic Search)
- Search over one or multiple Databuses

User Interface:

- Customizable forms for data entry
- Ontology autocompletion
- Customizable Search and Facets

Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

DLR Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center
REINER LEMOINE INSTITUT
InfAI Institute for Applied Informatics
Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung
PIK Potsdam Institute for Climate Impact Research
JÜLICH Forschungszentrum

Summary and Take Away

LOD-GEOSS

Databus

- Databus is comparable to Open Source Huggingface, Google Dataset Search, Maven Central for Data
- Databus has a very different purpose and mission to online data publishing portals like Zenodo, Invenio, CKAN, Leibniz Data Manager, Open Data Portal Sachsen
- It's gaining adoption
- It's Open Source
  - we offer support and managed hosting

Transforming Energy Systems Research

- FAIR and Open Data is
  1. a necessity for research
  2. a productivity factor
- Open Energy Family succeeded in producing useful standards:
  - Open Energy Ontology
  - Open Energy Metadata (JSON-LD)
  - Databus as Data Catalog
- Supports researchers in finding relevant data, trace provenance and automate workflows for models and scenarios

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

DLR Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center
REINER LEMOINE INSTITUT
InfAI Institute for Applied Informatics
Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung
PIK Potsdam Institute for Climate Impact Research
JÜLICH Forschungszentrum

Watch the video and vote for us to win:

Open Energy Databus

LOD-GEOSS

https://buergerbeteiligung.sachsen.de/portal/smwa/beteiligung/themen/1040556

or

https://tinyurl.com/vote-open-energy-databus

Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag

https://tinyurl.com/lswt-2024-databus
Research Grant 03EI1005A-G

DLR Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center
REINER LEMOINE INSTITUT
InfAI Institute for Applied Informatics
Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung
PIK Potsdam Institute for Climate Impact Research
JÜLICH Forschungszentrum

LOD-GEOSS

Contact us

Institut für Angewandte Informatik (InfAI)

KILT Competence Center & DBpedia Association

Sebastian Hellmann

Email: hellmann@infai.org

Supported by:

Federal Ministry for Economic Affairs and Energy

on the basis of a decision by the German Bundestag

License

Based on slides by German Aerospace Center:

„ FAIR Data in systems analysis“ © Hoyer-Klick, et. Al. | CC BY 4.0