LOD-GEOSS

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

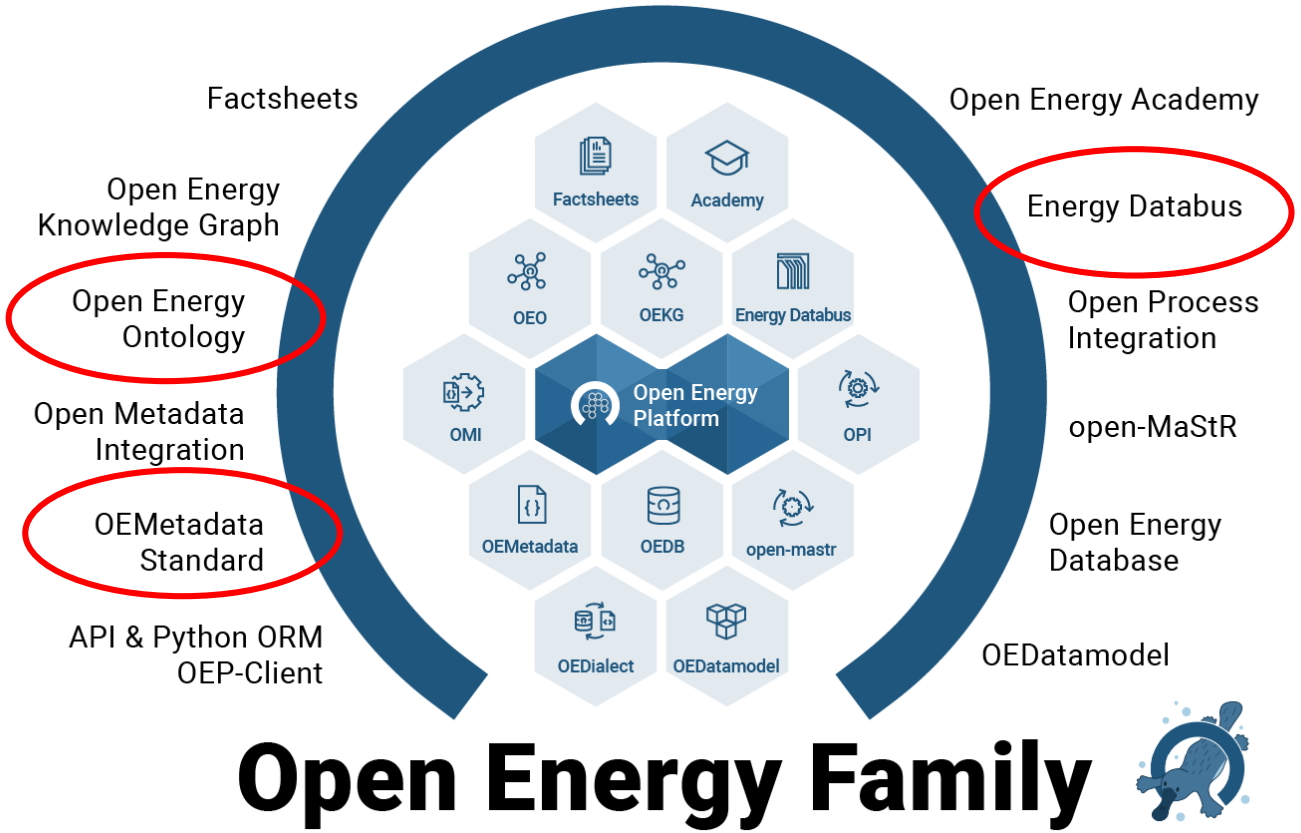
Research Grant 03EI1005A-G

Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center DLR | 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

Open Energy Family

LOD-GEOSS

- The **Open Energy Family** is an initiative for open and FAIR data in the domain of energy systems research
- Development of a FAIR infrastructure within the Open Energy Family



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

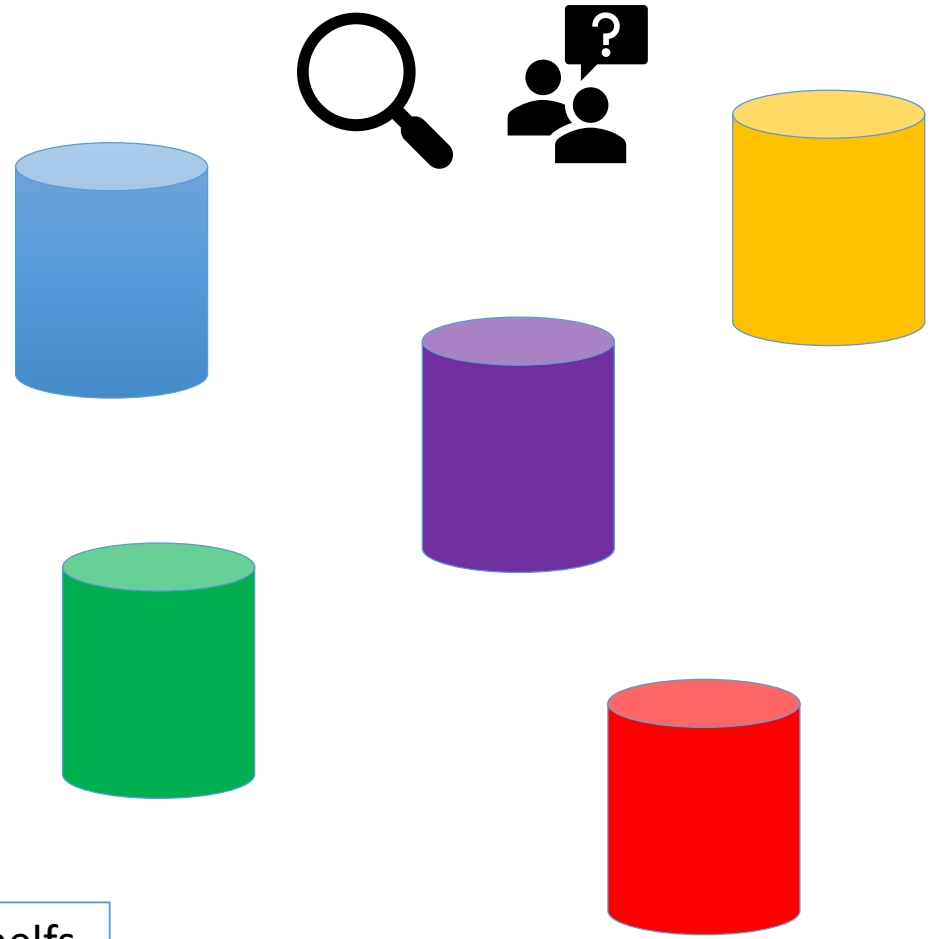
Research Grant 03EI1005A-G

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

Challenges I: Finding and Accessing Data

- Many data bases exist, each in its own flavor
  - 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

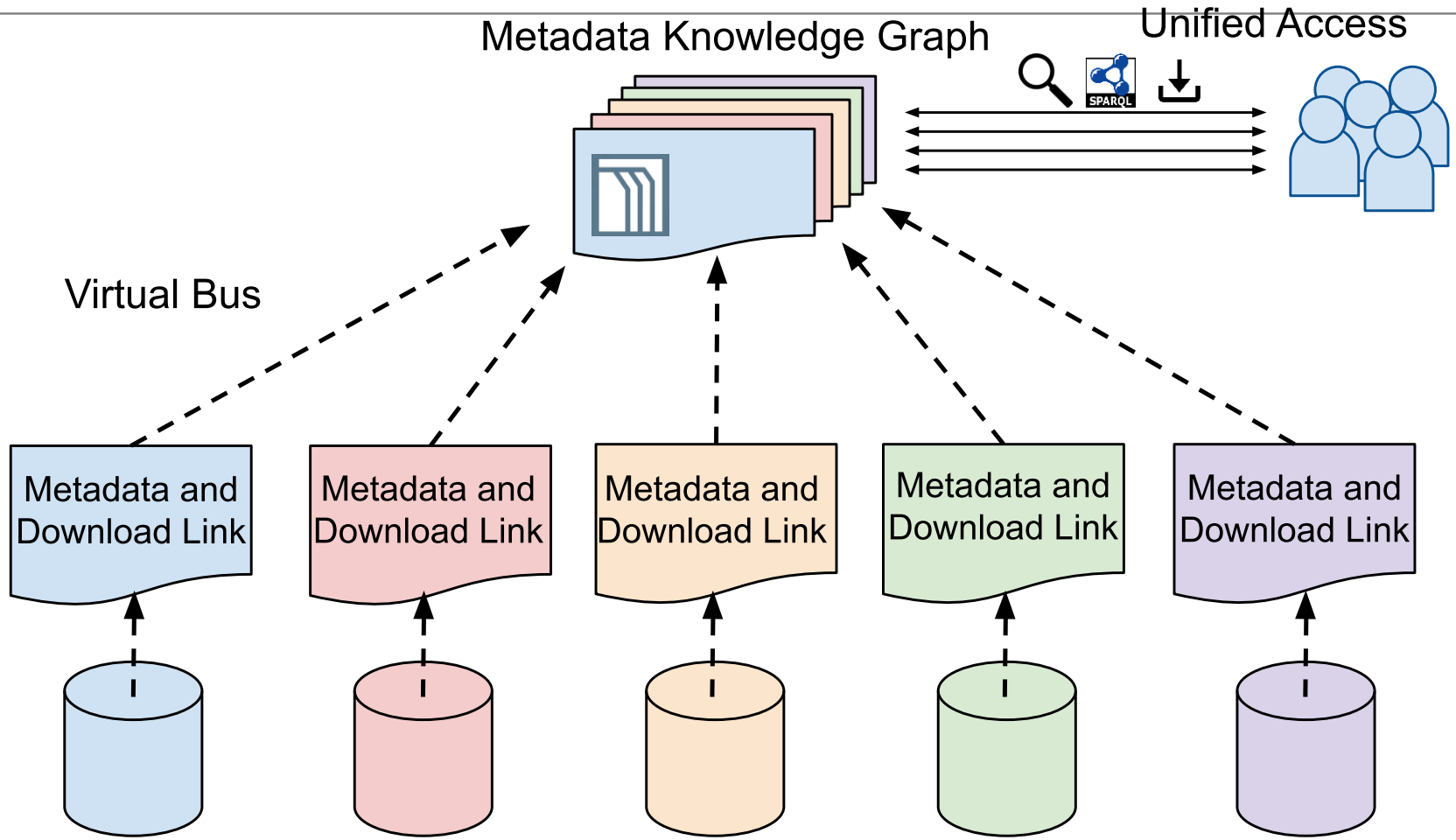


Imagine a library without a catalog and systematic numbers on the shelfs

Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

Solution I: A 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
- In case of data bases possibly also an API/Interface description



The domain uses the OEP Metadata string: https://openenergy-platform.org/tutorials/jupyter/OEMetadata/

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

Findable: OEMetadata

LOD-GEOSS

- 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 (v1.5.1) is “ontology ready”
- Target: 5-star Linked Open Data

Categories

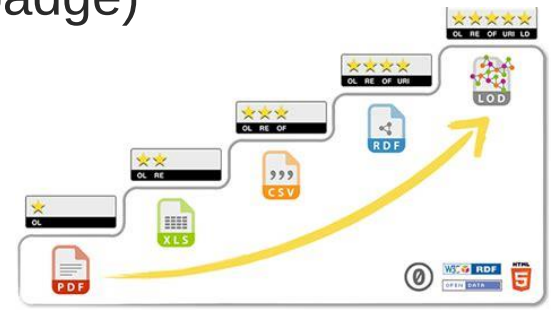
- **General** (name, title, description)
- **Context** (homepage, funding, contact)
- **Spatial** (location, extent, resolution)
- **Temporal** (referenceDate, timeseries)
- **Source** (origin, licenses)
- **Provenience** (contributors)
- **Resource** (schema, fields, type, description)
- **Review**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

OEMetadata

FRICTIONLESS DATA

DataCite

Findable Accessible Interoperable Reusable



https://github.com/OpenEnergyPlatform/oemetadata

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

Research Grant 03EI1005A-G

DLR Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center

REINER LEMOINE INSTITUT

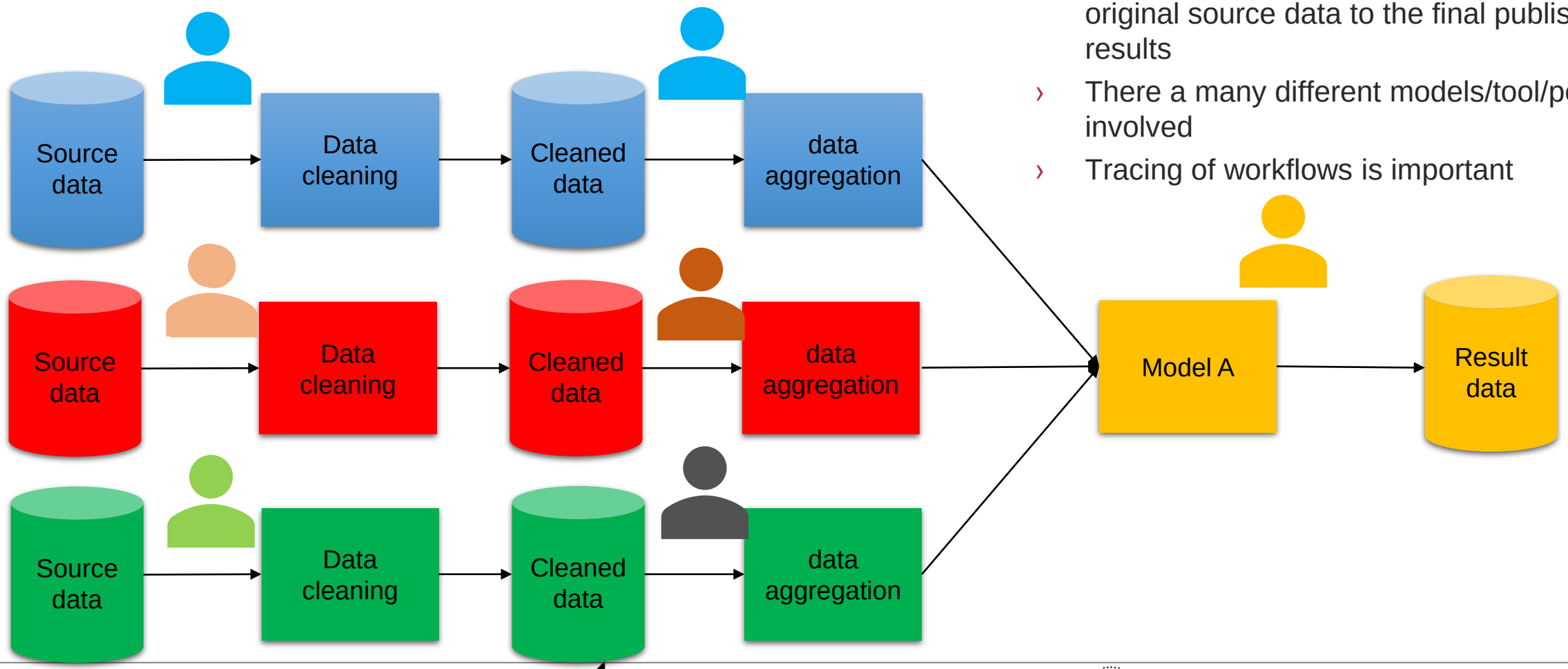
InfAI Institute for Applied Informatics

Universität Stuttgart IER Institut für Energiewirtschaft und Rationelle Energieanwendung

PIK Potsdam Institute for Climate Impact Research

JÜLICH Forschungszentrum

Challenge II: Who has done what to the data?



- There are many processing steps from the original source data to the final published results
- There a many different models/tool/persons involved
- Tracing of workflows is important

Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

PROV-O as a W3 Standard

LOD-GEOSS

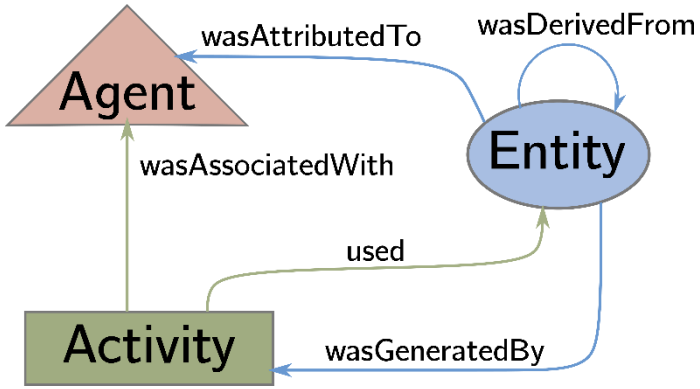
Entities	Digital, physical or conceptional (.e.g. data sets or files)
Activities	Use or generate entities, e.g. algorithms or workflows
Agents	Have control over activities or own entities, e.g. persons, organizations, software

Entities, **Activities** and **Agents** are related through **relations** with each other.

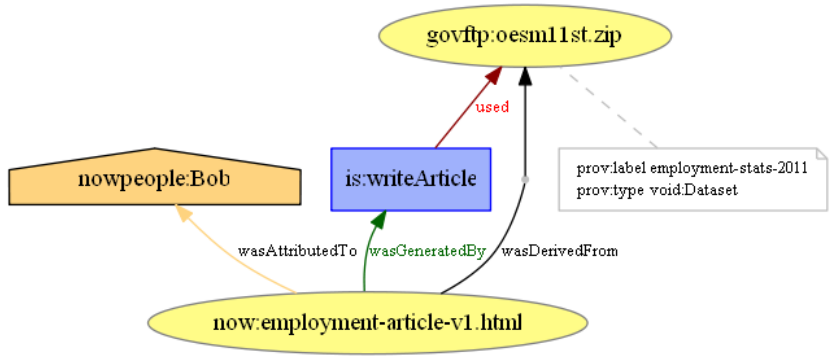
Provenance Information can be collected automatically and visualized in graphs, e.g. with ProvStore

A record **of who (Agent)** processed with which **Tools (Activity)** which **data (Entities, input and output)**

Schematic:



Example:

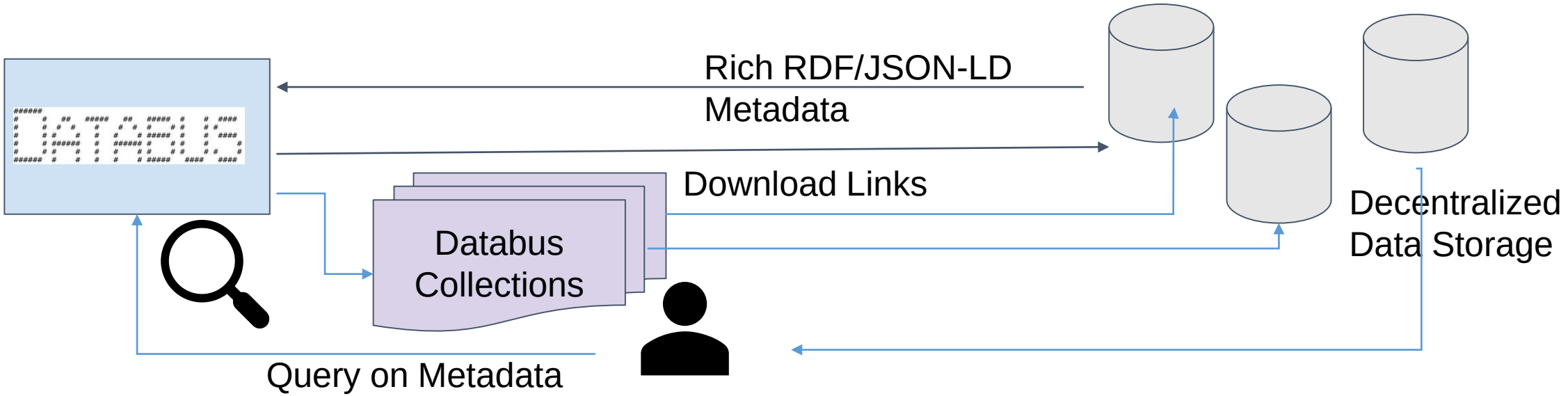


Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag

Research Grant 03EI1005A-G

Findable & Accessible: The Open Energy Databus

Databus is a virtual bus. It can address files on the web and coordinate dataflows based on DataID metadata. No actual data is uploaded to the bus.

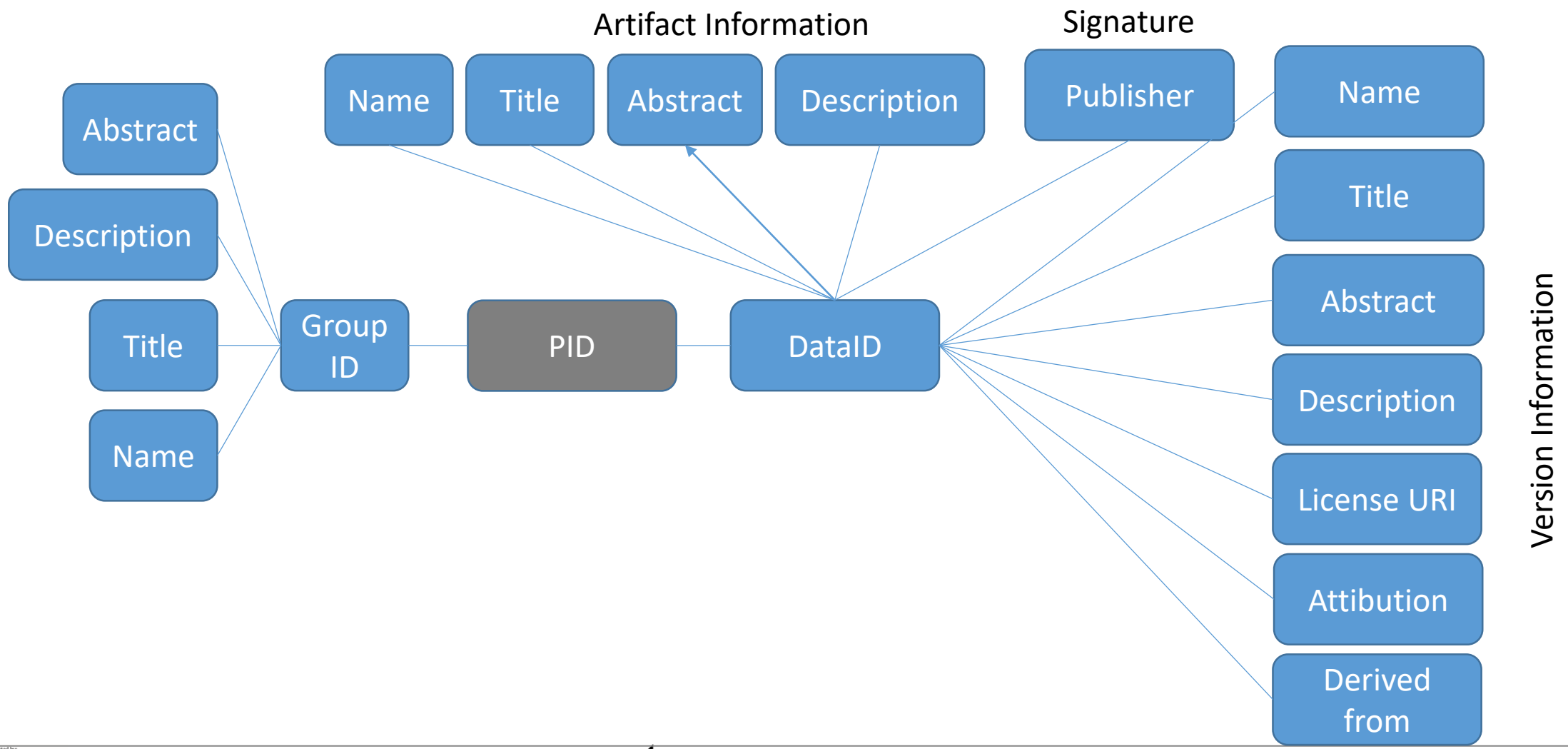


- Unique and persistent data identifiers (PID) are created by the Databus https://databus.openenergyplatform.org/account_name/group/artifact/version/
- Data sets are linked to their source data through the data ids
- Incremental modifications to data (e.g. people can reuse cleanings or aggregations someone else has done before)

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

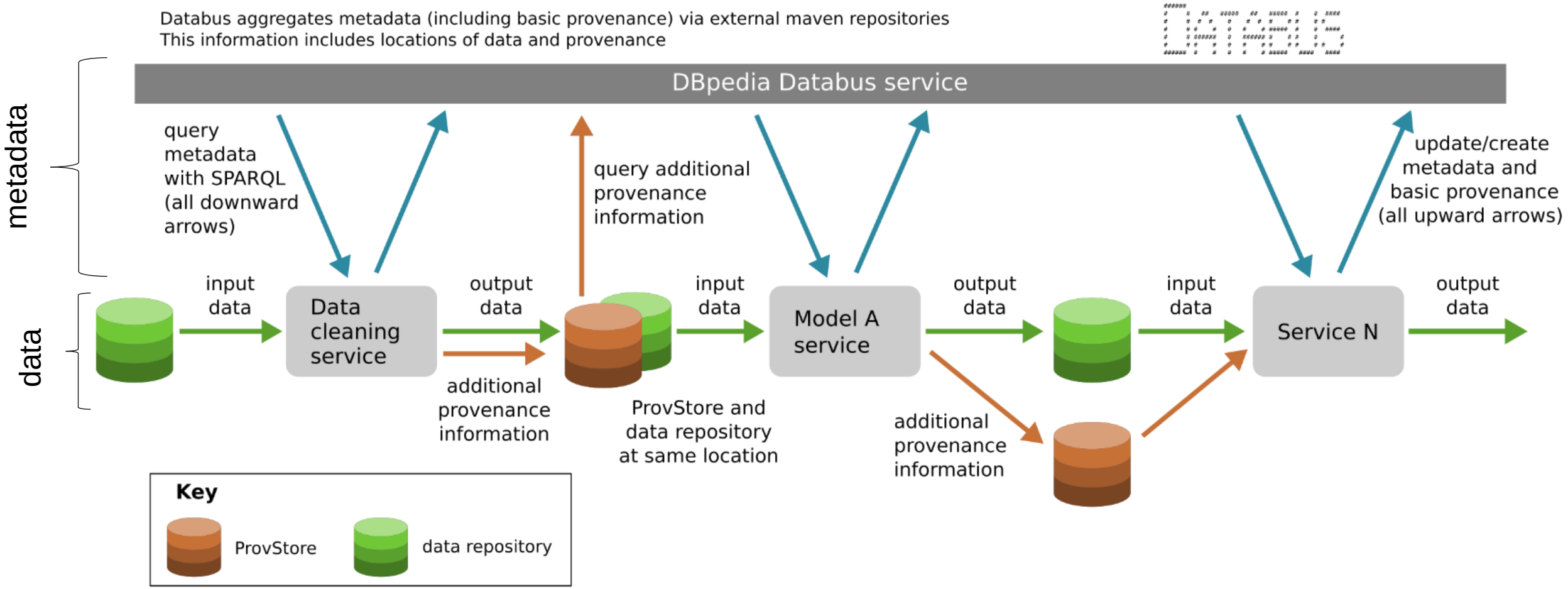
Research Grant 03EI1005A-G

Databus Metadata



Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

Findable & Accessible: Architecture Concept

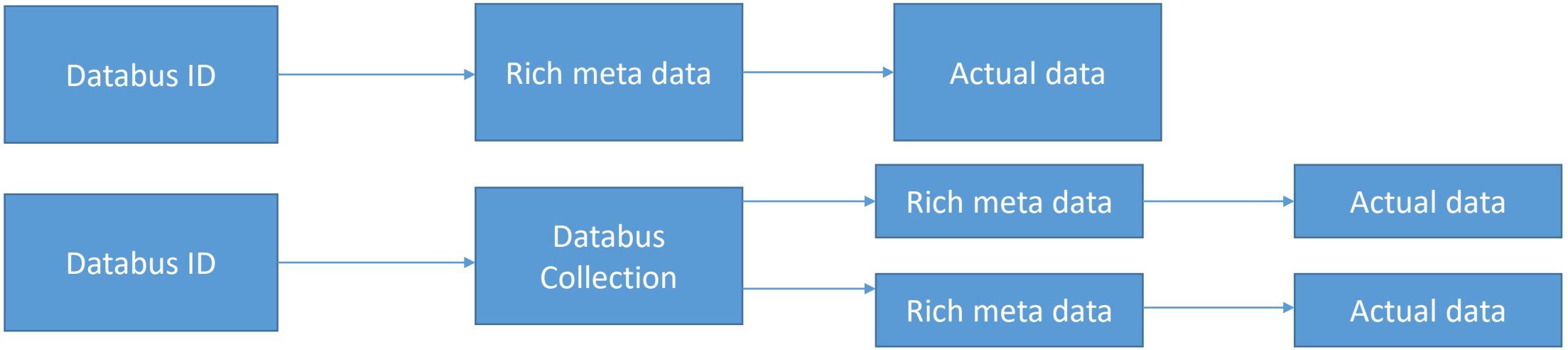


Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

Databus as PID / Fair Digital Object (FDO) provider

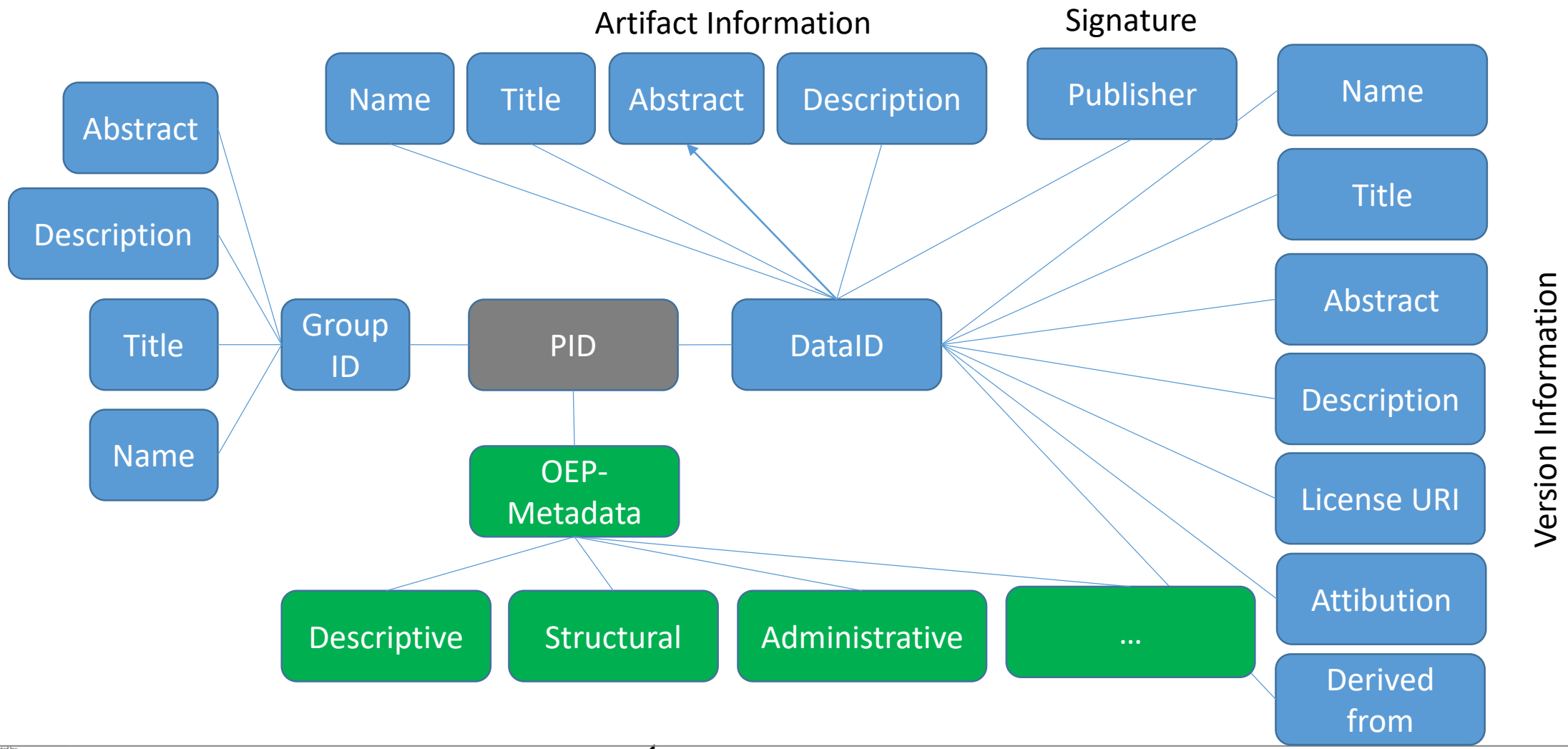
LOD-GEOSS

- Uploading meta data to the Databus creates a unique ID for the data set which can be used as a persistent data identifier (PID) to cite and share the data
- Databus URI (PID): https://databus.openenergyplatform.org/account_name/group/artifact/version/
  - Databus collections can be used to group data
  - The actual data is linked within the meta data
- Databus ID can be used to access all necessary information
- Databus ID is a pointer to a (FAIR) digital object
- **Persistent ID to make data citable**



Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

Enhancing the databus graph with OEMetadata



Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

Interoperable: The Open Energy Ontology

LOD-GEOSS

- Each data source comes with its own annotation

- Example from solar meteorology:
  - **GHI:** Global Horizontal Irradiation (Energy) or Irradiance (Power)
  - **Global:** Could also be Global Horizontal Irradiation or Irradiance
  - **Surface downward irradiation:** The usual term in climate science for what we usually call GHI

- **Taxonomies** or **ontologies** create a **data language** to annotate data
- Ontologies can describe relations: *direct radiation* **is a part** of the *global irradiation reaching the surface*
- Ontologies make data interpretable, also by machines and algorithms

- **Good ontologies are created on a consensus building and open development process within the community.**
  - We use the 'Open Energy Ontology', https://openenergy-platform.org/ontology/, https://doi.org/10.1016/j.egyai.2021.100074

- Data sets can be annotated with the "**Subject"** tag, individual columns in the resource section with "**Is about"** within the JASON-LD metadata

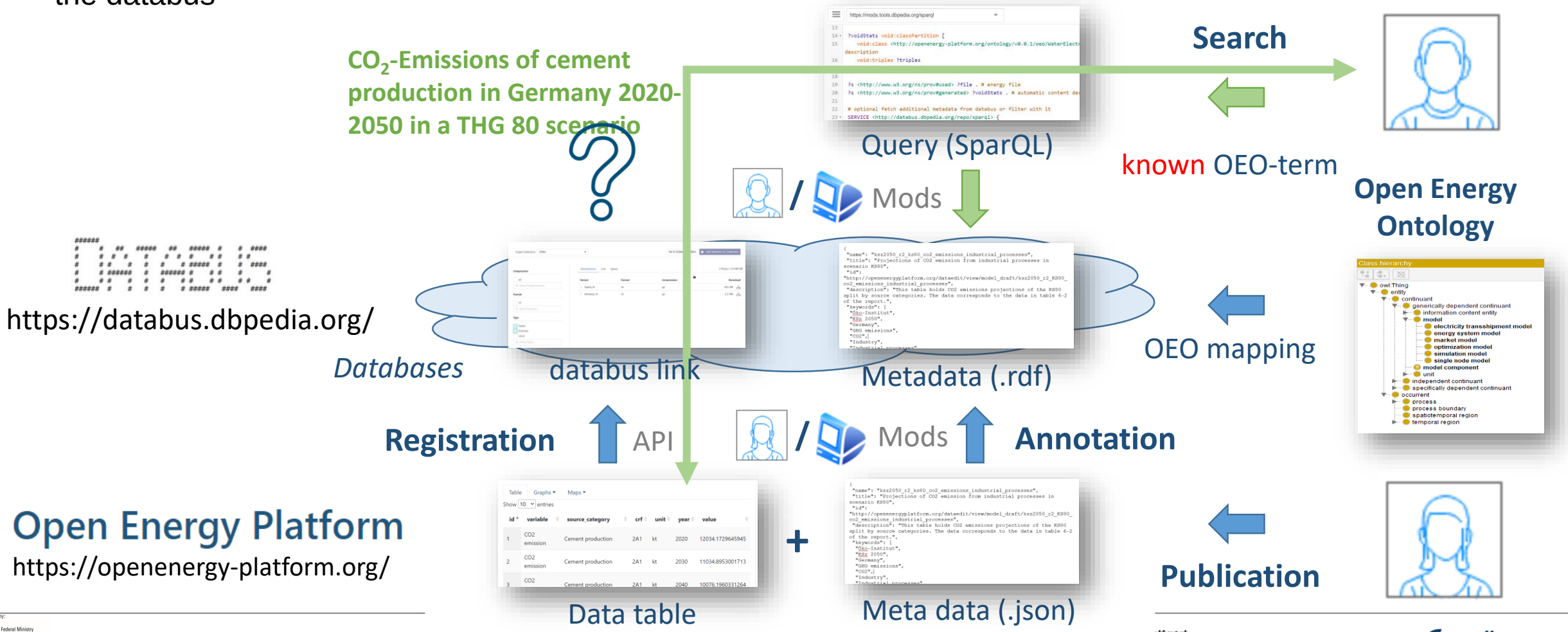
Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

Research Grant 03EI1005A-G

Demonstrator: Publication of a Data Set Using the databus

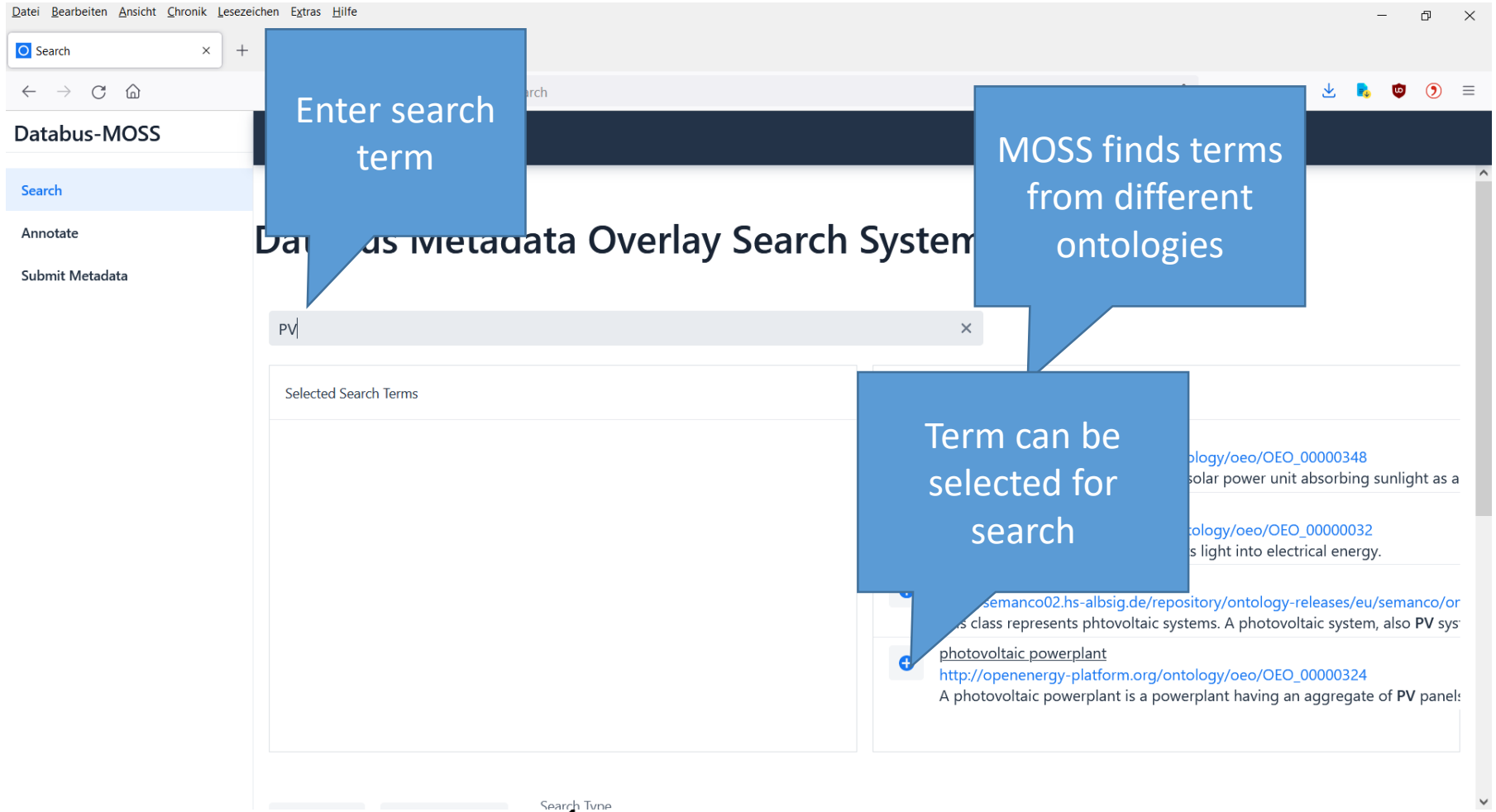
LOD-GEOSS

› **Goal**: Demonstration of the improved visibility and improved discovery of a data set through the registration in the databus



Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

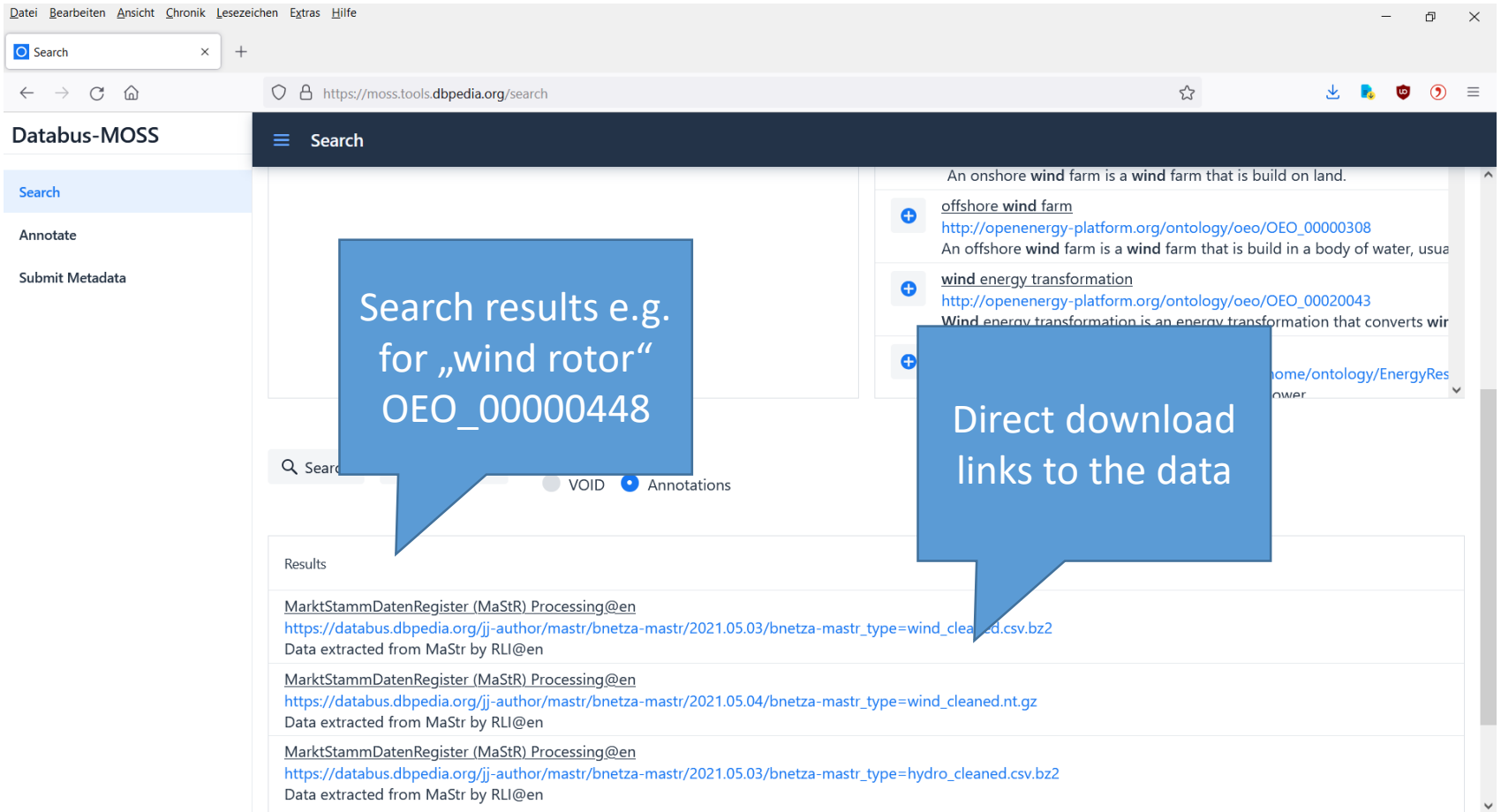
Searching data with the MOSS (Metadata Overlay Search System)



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

Research Grant 03EI1005A-G

Searching with MOSS



Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag
Research Grant 03EI1005A-G

Conclusion

- The databus offers
  - a service to manage and search registered metadata
  - Persistent identifies for tracing data processing and citing data
  - Databus as pointers to digital objects
- **Reusable:** Data licenses are an obligatory part of the DataID and OEMetadata and are linked to dalicc.net to be machine actionable.
- The databus supports the implementation of FAIR principles in the Domain of Energy Systems Analysis
- The developed architecture in conjunction with the use of the Open Energy Ontology enables semantic searches for data in the domain of energy systems analysis
- **The developed architecture with distributed repositories, common metadata and schema descriptions, an ontology and a data catalog already forms some kind of open data cloud with the domain of energy systems analysis.**

- Further resources:
  - https://zenodo.org/communities/lod-geoss
    - Recommendations on Data Licensing
    - Demonstration and Best Practices
  - https://databus.openenergyplatform.org/
  - hhttps://github.com/OpenEnergyPlatform/oemetadata
  - https://openenergy-platform.org/ontology/
  - https://doi.org/10.1016/j.egyai.2021.100074
  - https://moss.tools.dbpedia.org/search
  - https://github.com/LOD-GEOSS/databus-snippets

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

Research Grant 03EI1005A-G

LOD-GEOSS

Contact us

German Aerospace Center

Energy Systems Analysis

Carsten Hoyer-Klick

Email: carsten.hoyer-klick@dlr.de

Supported by:

Federal Ministry for Economic Affairs and Energy

on the basis of a decision by the German Bundestag

License

Please cite as:

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

Supported by:

Federal Ministry for Economic Affairs and Energy

on the basis of a decision by the German Bundestag

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

PIK Potsdam Institute for Climate Impact Research

JÜLICH Forschungszentrum