LOD-GEOSS

Improved Discovery and Access to Research Data in Energy Systems Analysis

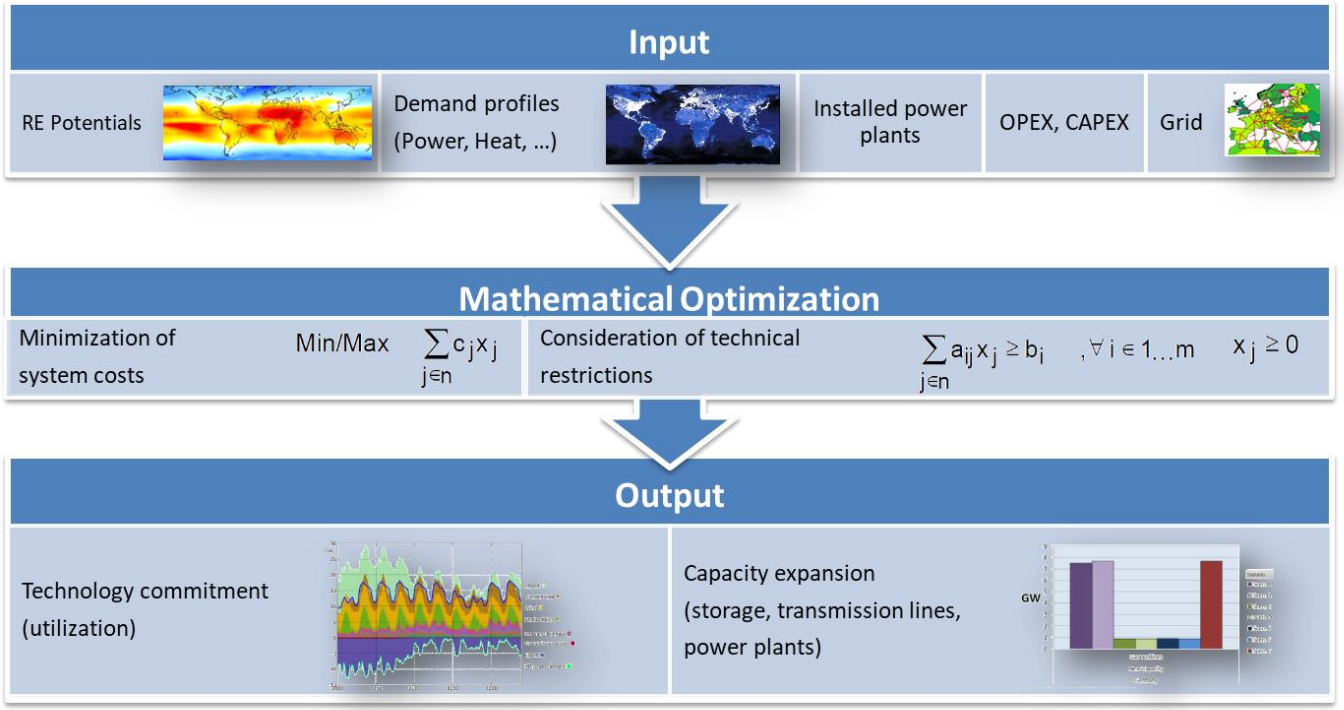
Carsten Hoyer-Klick (DLR), Johannes Frey (InfAI)

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

The Domain of Energy Systems Analysis

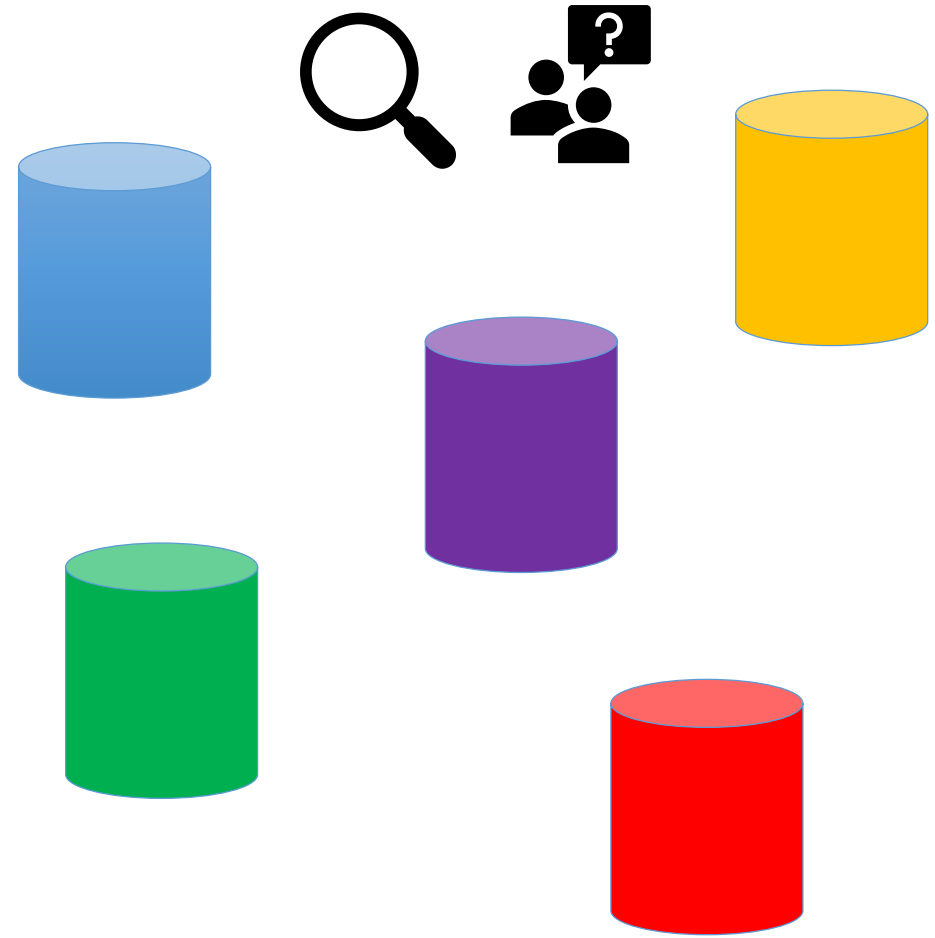
LOD-GEOSS

- Focus of the research is the simulation and optimization of future national and transnational energy systems
- Modelling is very dependent on a wide range of different input data and produces a lot of data itself
- Very heterogenous input data, e.g.
  - Geographical and meteorological data (gridded data)
  - Time series of demand, prices
  - Technology data bases
  - Parameters



Challenges I:

- Many data bases exist, each in its own flavor
  - Data access
  - Data format
  - Data licenses (if at all)
  - Sometimes hard to find
- Descriptive information missing or not well attached and therefore little information to be crawled
- Data collection is a labor intensive task
- Data cleaning is repeated by many researchers with different results
- Data quality is often unknown



Similar Problems in Earth Observation in the early 2000

LOD-GEOSS

- Many providers of Earth Observation data
- Every provider with his own access mechanisms
- Data is hard to find and to access
- Solutions:
  - Founding of GEO (Group on Earth Observation) as a G8 initiative in 2005.
  - GEOSS (Global Earth Observation Systems of Systems) as an architecture to network Earth Observation around the Globe.
  - Data and Architecture Committee to define interoperability and communication standards
  - Use of data standards of the Open Geospatial Consortium (OGC) for geographical data and webservices.

Solution: A Metadata Catalog for Search and Discovery

LOD-GEOSS

User interfaces

Data catalog

Data Provider

Solar resources A (e.g. DLR)

Data Provider

Solar resources B (e.g. NREL)

Knowledge Database

Policy data base e.g. IEA, UNEP, REN21

Live Sensor Data

Measurement stations, Satellite data

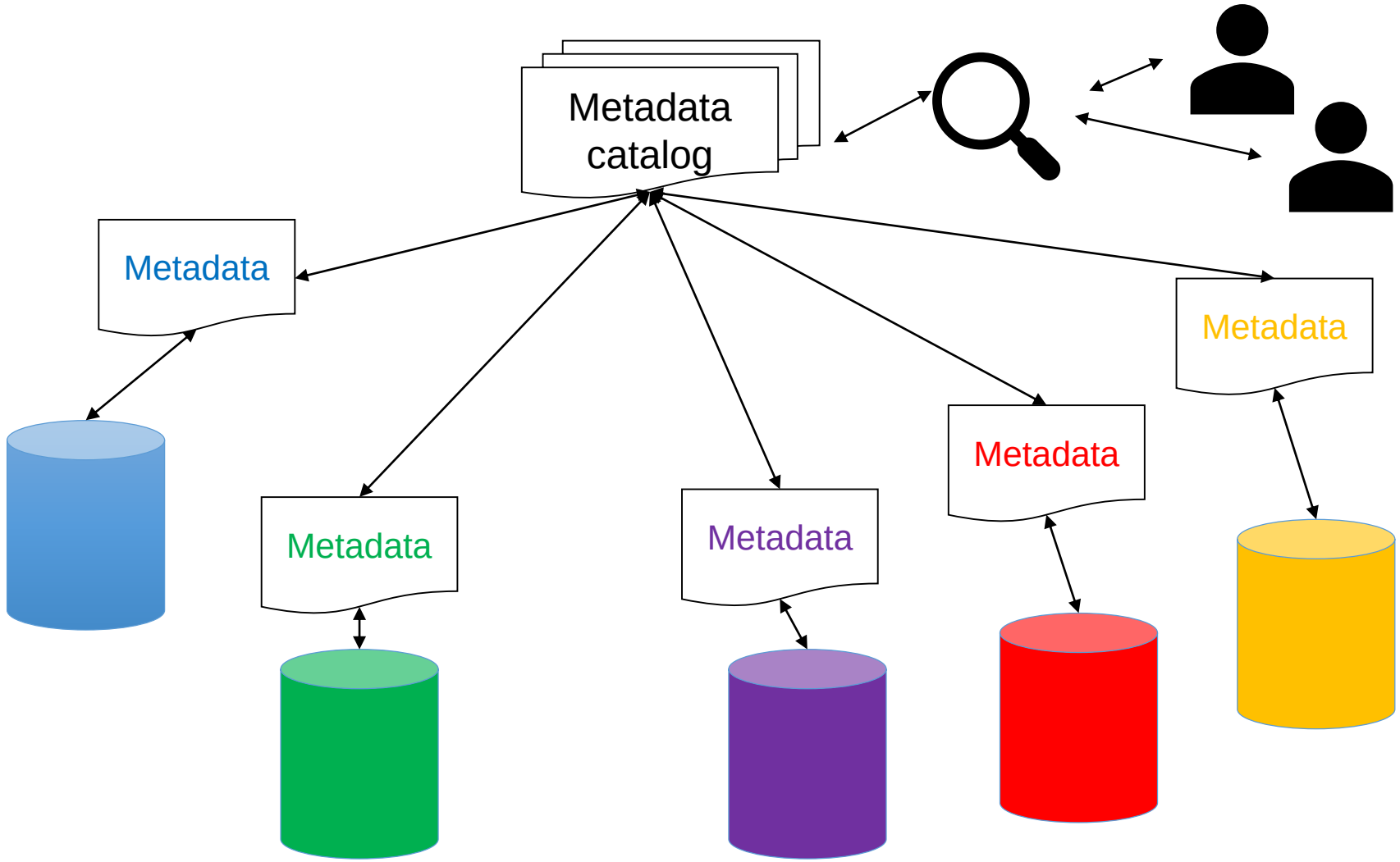
Communication with open and standardized internet protocols

Problem:
- Many different forms of meta data
- No common vocabulary

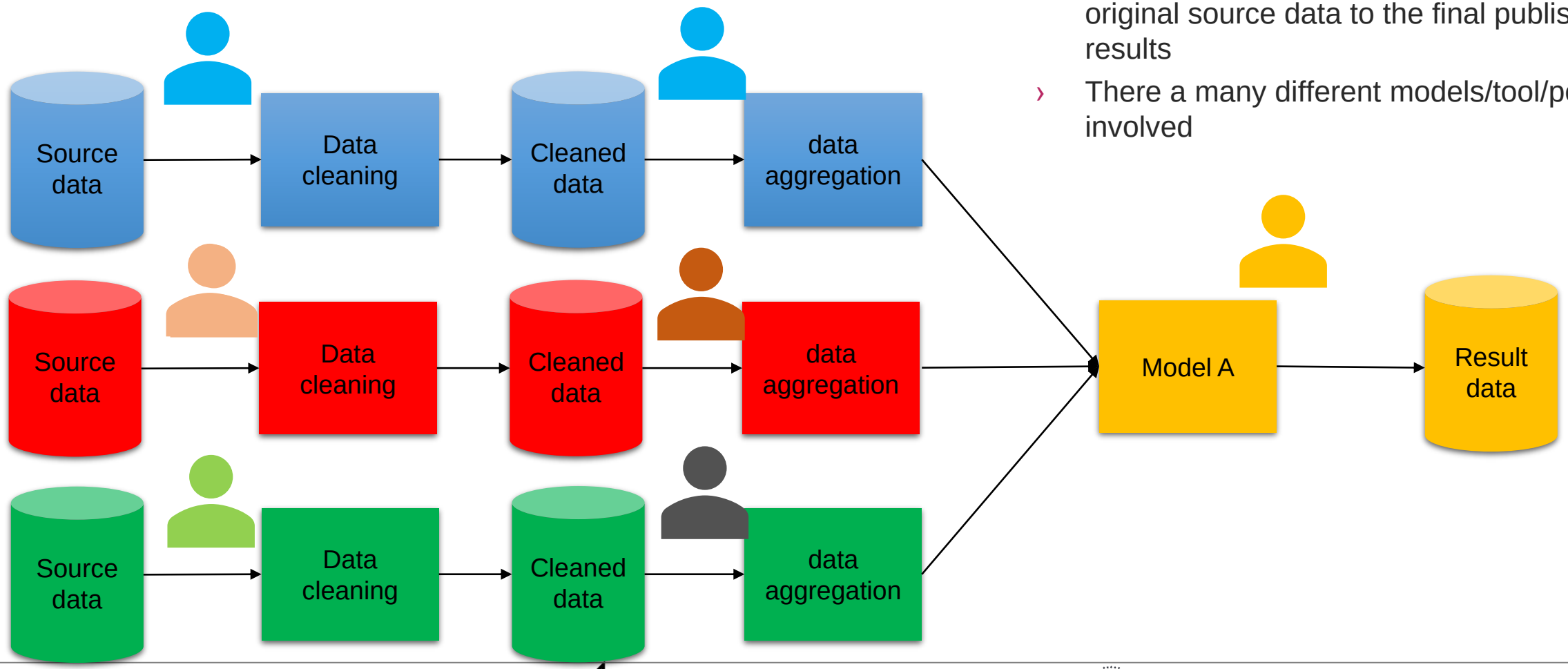
Set up of the architecture within the framework of GEOSS

Solution I: A Metadata Catalog

- A metadata catalog or index harvests the metadata from the available data sources
- The catalog can be used to discover data
- Metadata contains descriptive information
- The metadata contains a URI to the actual data
- In case of data bases possibly also an API/Interface description



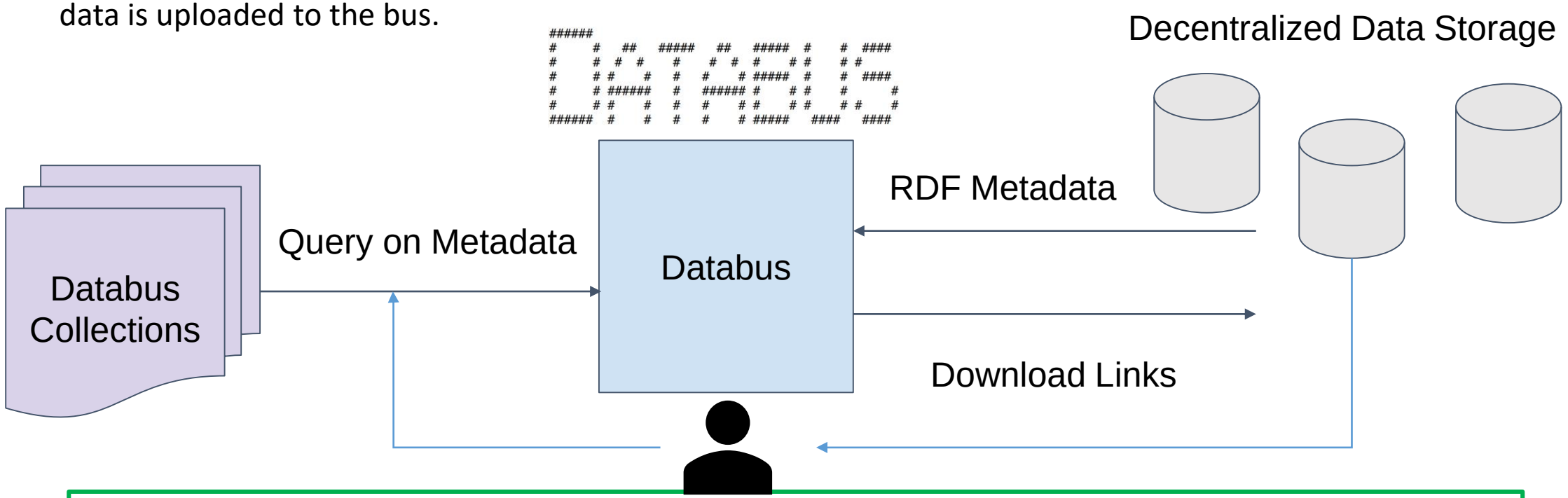
Challenge II: Who has done what to the data?



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

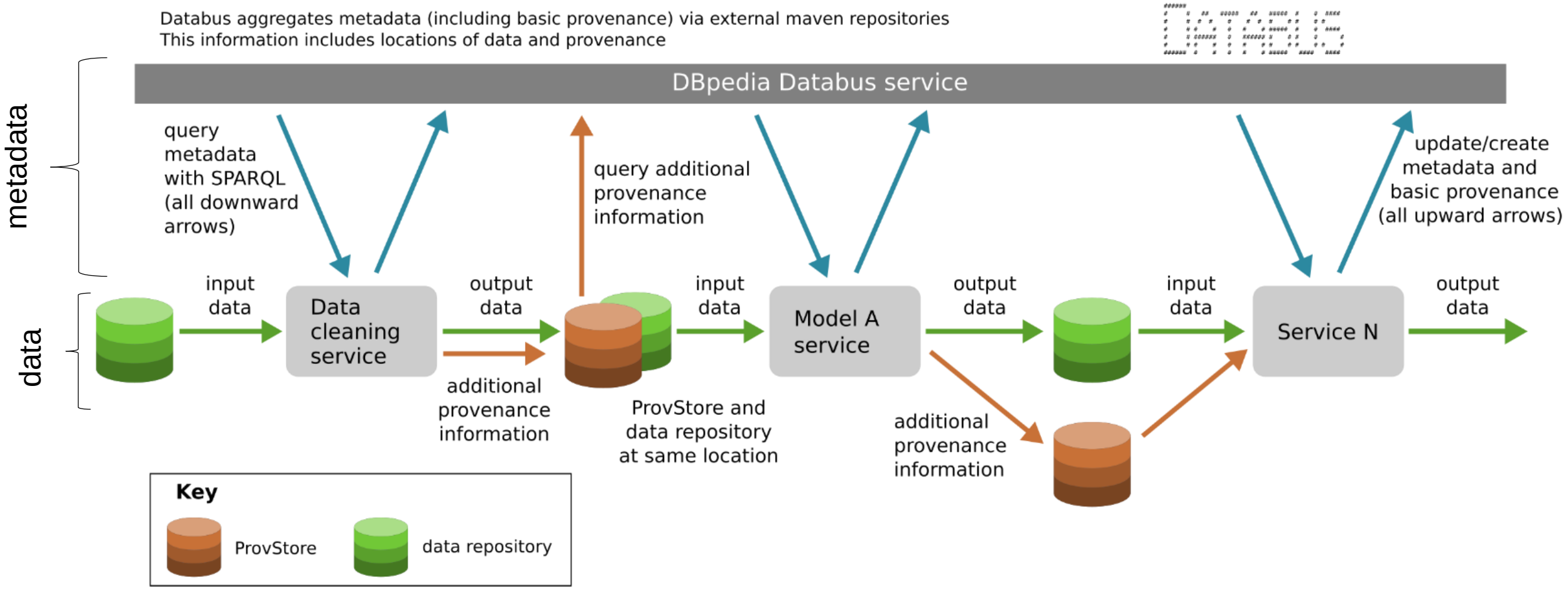
The Databus Platform

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 data identifiers are created by the Databus
- 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 has done before)

Architecture Concept



Challenge III: What is the data about?

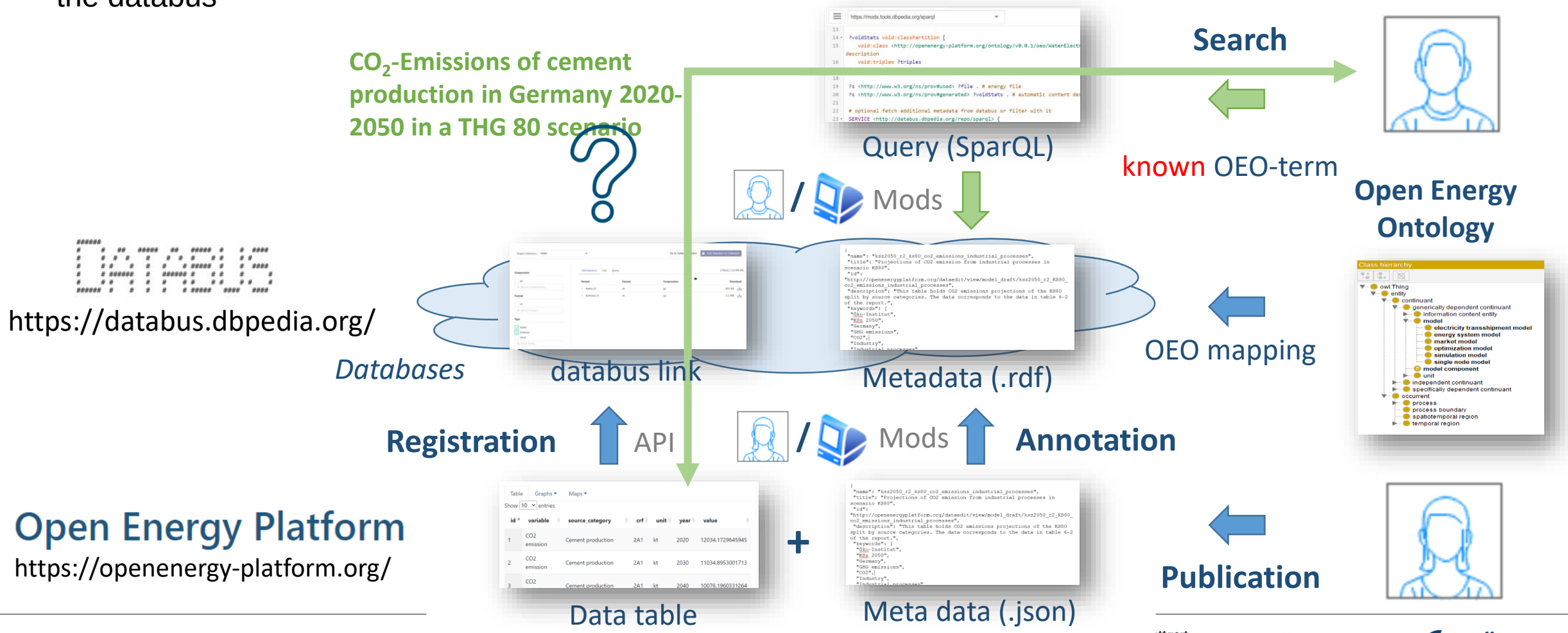
LOD-GEOSS

- Each data source comes with its own annotation
- Example from solar meteorology:
  - GHI: Global Horizontal Irradiation
  - Global: Could also be Global Horizontal Irradiation
  - Direct: Could be Direct Normal Irradiation or Direct Horizontal Irradiation
  - 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
- **Learning from Linked Open Data: Semantic annotation of the data with the Open Energy Ontology**
- The domain uses the OEP Metadata string: https://openenergy-platform.org/tutorials/jupyter/OEMetadata/
  - JSON-LD Extension is currently under way, to be released maybe this week, at least within October 2021.

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



Conclusion

- The domain of energy systems analysis works with very heterogenous data sets
- Meta data is increasingly used to describe the data
- The databus offers
  - a service to manage and search registered metadata
  - Persistent identifies for tracing data processing
- 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 meta data collection on the databus is something as a domain specific search index for data**

- Further ressources:
  - https://lod-geoss.github.io
  - https://databus.dbpedia.org
  - https://openenergy-platform.org/tutorials/jupyter/OEMetadata/
  - https://openenergy-platform.org/ontology/

LOD-GEOSS

Contact us

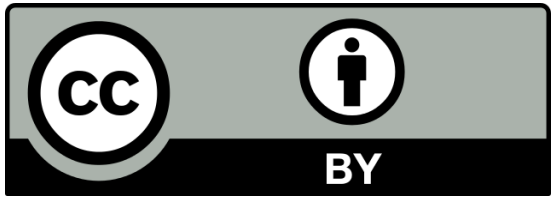
German Aerospace Center

Energy Systems Analysis

Carsten Hoyer-Klick

Email: carsten.hoyer-klick@dlr.de

License



© German Aerospace Center

Except where otherwise noted, this work and its content (texts and illustrations) are licensed under the

Attribution 4.0 International (CC BY 4.0)

See license text for further information.

Please cite as:

„Improved Discovery and Access to Research Data in Energy Systems Analysis “ © Germany Aerospace Center | CC BY 4.0

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