# Implementing FAIR through a distributed data infrastructure

Carsten Hoyer-Klick, Johannes Frey, Ulrich Frey, Hedda Gardian, Anastasis Giannousakis, Jan Göpfert, Tobias Hecking, Christian Hofmann, Sophie Jentzsch, Kevin Knosala, Leander Kotzur, Stefan Kronshage, Patrick Kuckertz, Christoph Muschner, Michaja Pehl, Vera Sehn, Detlef Stolten
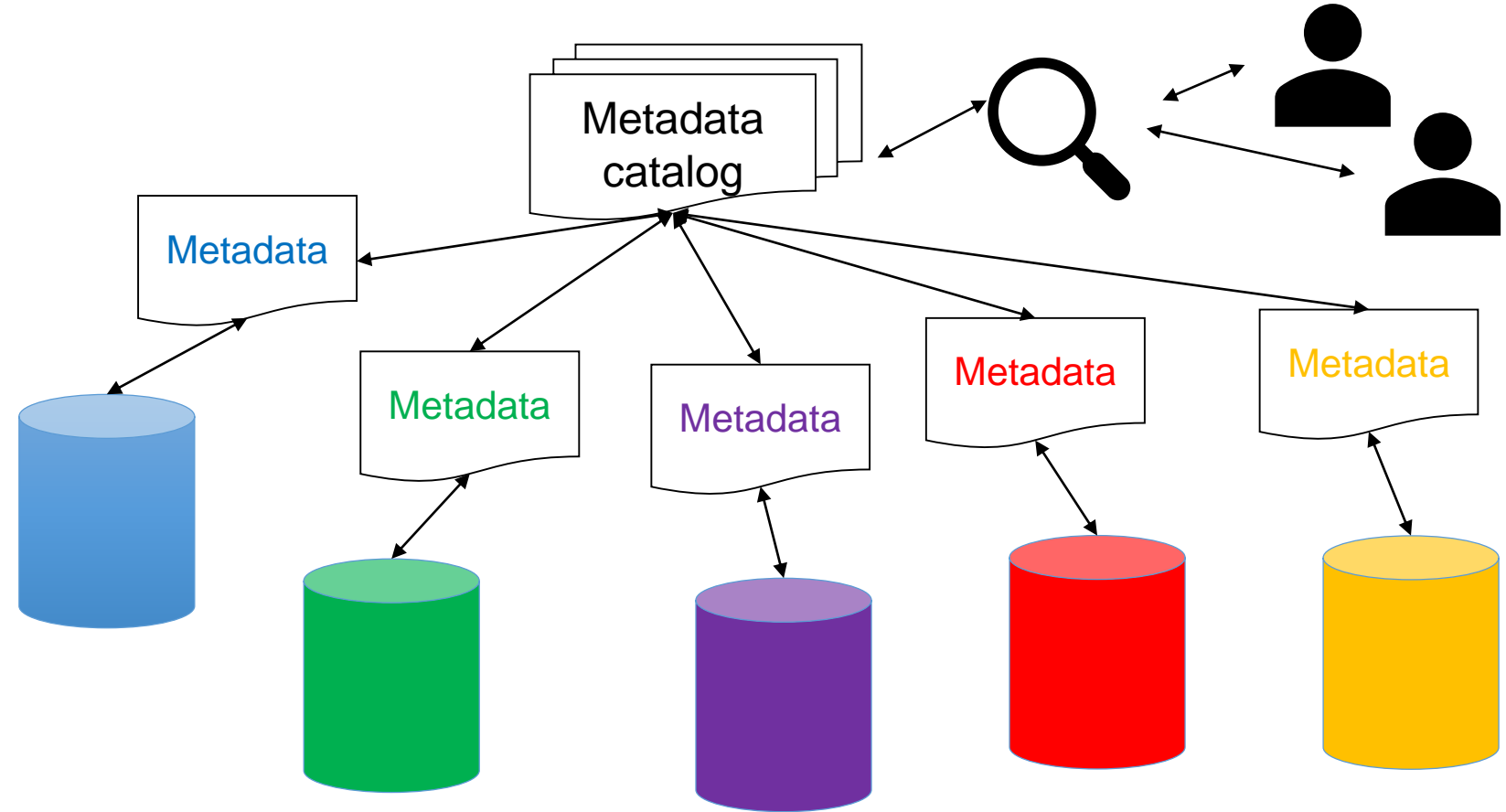
Challenges I:

› As stated in the opening: energy system modelling remains important for policy development, but

› 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
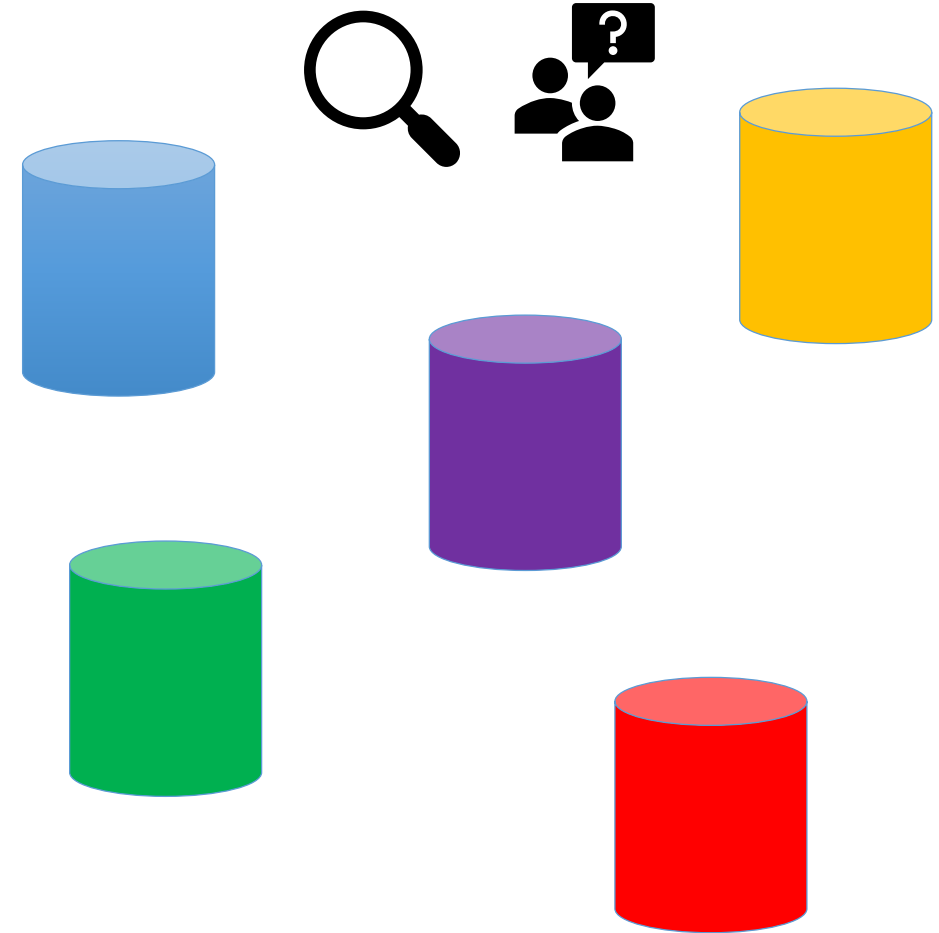
Solution I: A Metadata Catalog

› 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/
JSON-LD extension is currently under way, to be released maybe this week, at least within October 2021.
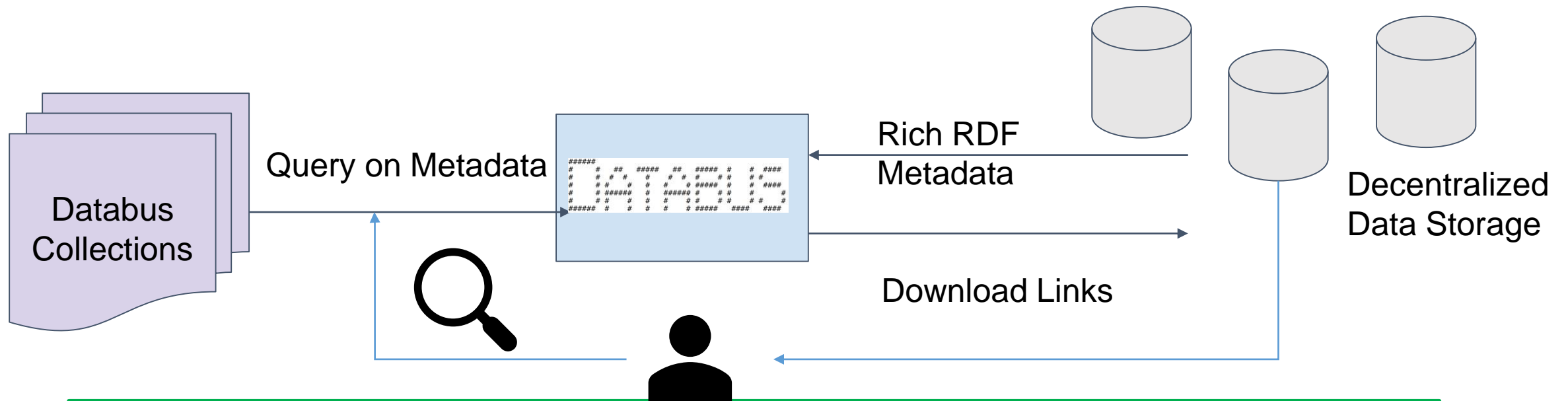
Challenge II: Who has done what to the data?
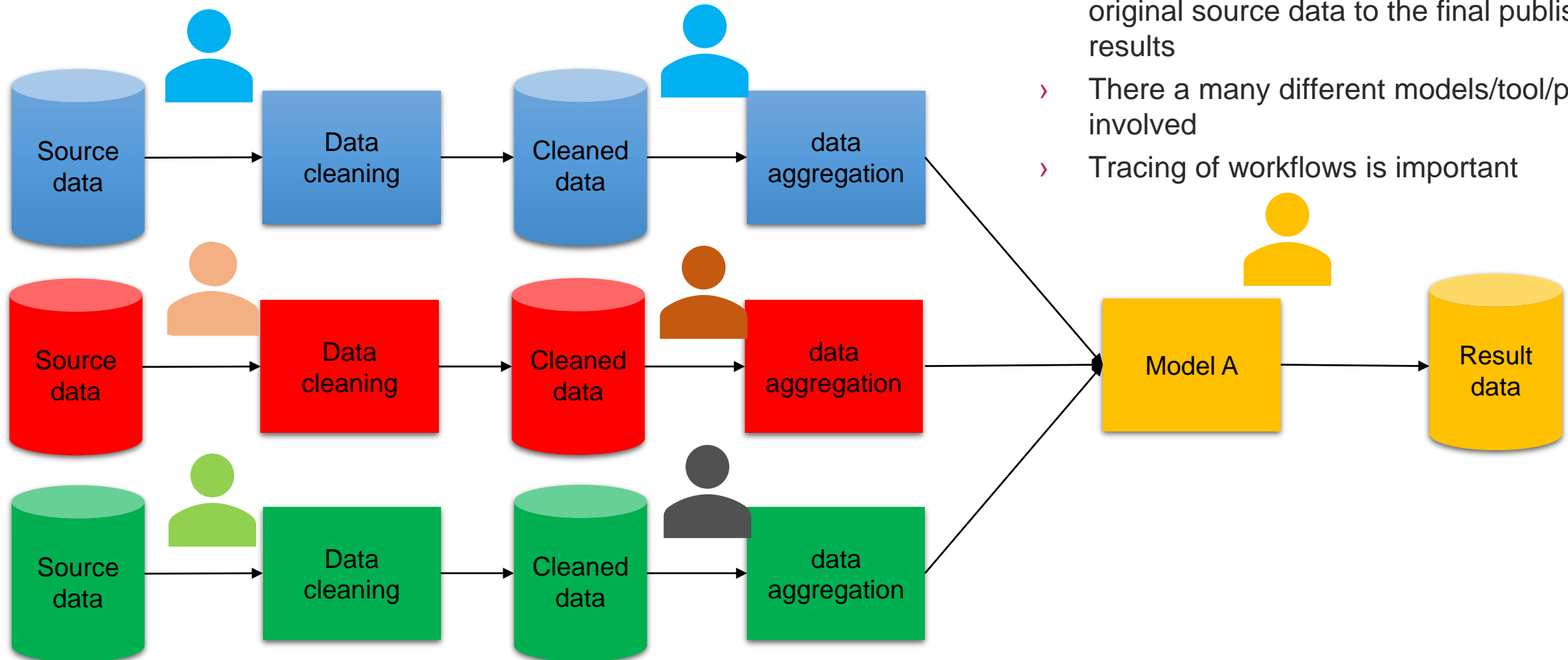
- › There are many processing steps from the original source data to the final published results
- › There a many different models/tool/persons involved
- › Tracing of workflows is important

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.

Query on Metadata

Rich RDF Metadata

Databus Collections

Decentralized Data Storage

Download Links

- 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 else has done before)

Architecture Concept
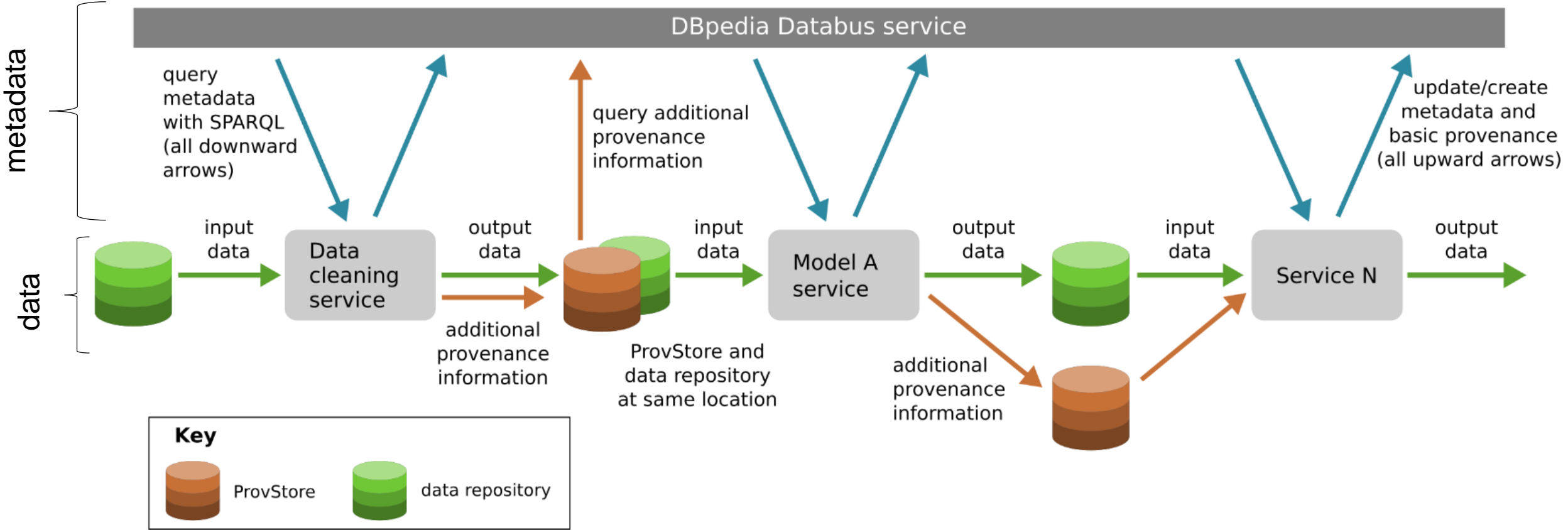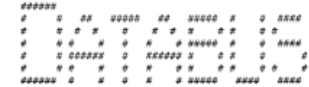
Challenge III: What is the data about?

› Each data source comes with its own annotation

› Example from solar meteorology:
  › **GHI:** Global Horizontal Irradiation
  › **Global:** Could also be Global 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
› 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

› New JASON-LD Meta data string can link meta data to an ontology, about to be released

Implementing the FAIR Principles

› **F**indable

Development and setup of the data bus as metadata catalog for data in energy systems analysis

Databus architecture and meta data

› **A**ccessible

Descriptions of the data formats and interfaces to data bases, Best Practice Guides, links to the data sources in the metadata

› **I**nteroperable

Enhancement of on Open Energy Ontology as a common data language, open data format descriptions

The open energy ontology

› **R**eusable

Improvement of data licenses, here with a special focus on GEOSS data, provenance information with the data
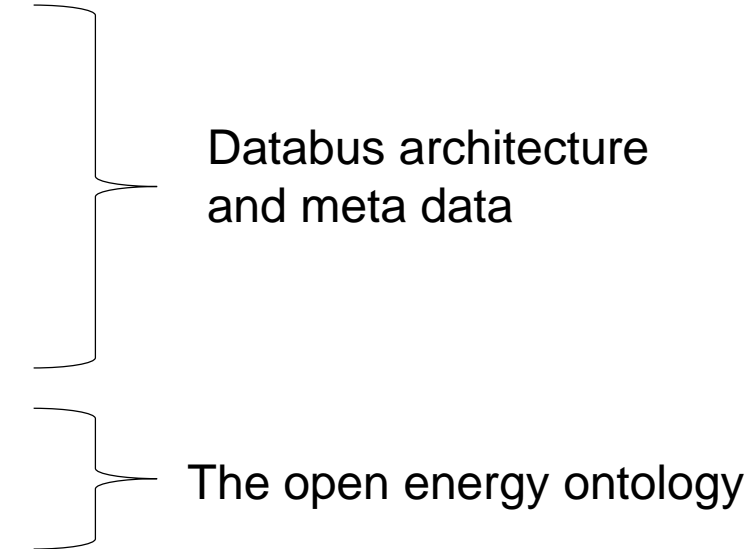
Demonstrator: Publication of a Data Set Using the databus

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

CO₂-Emissions of cement production in Germany 2020-2050 in a THG 80 scenario

https://databus.dbpedia.org/

Query (SparQL)

Search

known OEO-term

Open Energy Ontology

Databases   databus link   Metadata (.rdf)   OEO mapping

Registration   API   / Mods   **Annotation**

Open Energy Platform

https://openenergy-platform.org/

Data table   +   Meta data (.json)

Publication

Klimaschutzszenario 80 (KS80) BMU, Ökoinstitut
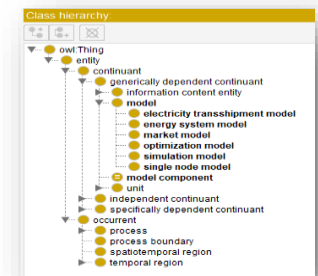
Demonstrator Model Coupling

Analysis of the refinancing of RE-technologies in Germany from 2020 to 2030

- Energy demand
- Technology data
- RE Potentials
- …

Transformation
FINE (FZJ)

Transformation
TIMES (IER)

- Installed capacities
- Power generation
- Operational cost
- …

Transformation
AMIRIS (DLR)

- Used power per technology
- Electricity price
- …
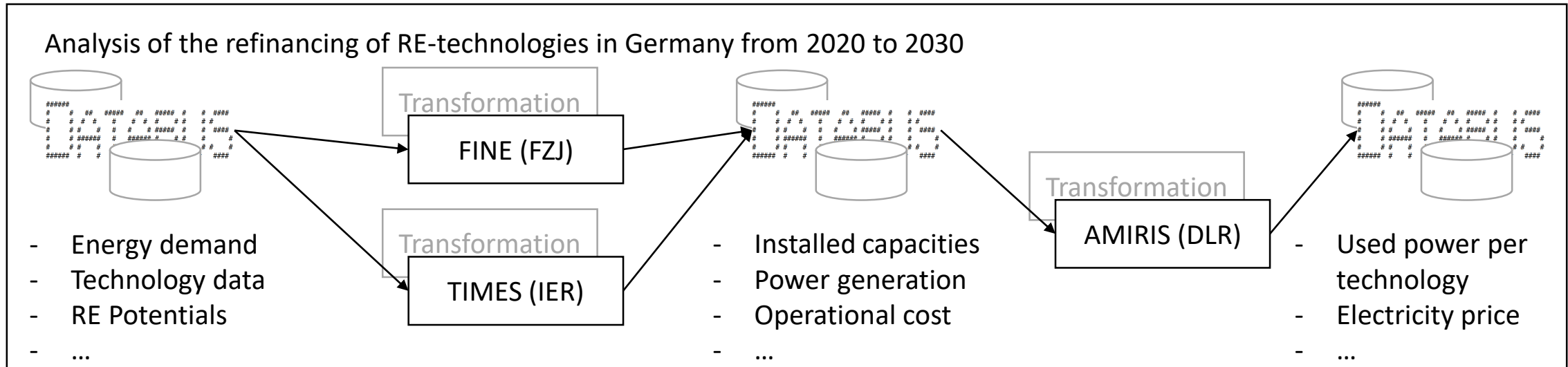
› Goal of the demonstrator
  › Provision of a holistic application example of a databus based coupling of heterogeneous energy system models for answering a research question.
  › Databus and Metadata description as a standard interface for data between models
› Challenges in the handling of data
  › Preparation and description of data content (data and metadata standards)
  › Preparation and description of data formats (open licenses, machine readable formats, …)
  › Discovery, interpretation and further processing of the data

Conclusion

LOD-GEOSS

› The databus offers

  › a service to manage and search registered metadata

  › Persistent identifies for tracing data processing

› The databus can improve the data economy and scientific efficiency as we can build better on existing previous work, improve the data instead of repeating work others have done before.

› 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 can improve open data exchange, model coupling, tracing of workflows and collaboration for better scientific results**

› Further resources:

  › https://lod-geoss.github.io

  › https://databus.dbpedia.org

  › https://openenergy-platform.org/tutorials/jupyter/OEMetadata/

  › https://openenergy-platform.org/ontology/

  › https://doi.org/10.1016/j.egyai.2021.100074

Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center

RLI 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

Contact us

German Aerospace Center

Energy Systems Analysis

Carsten Hoyer-Klick

Email: carsten.hoyer-klick@dlr.de

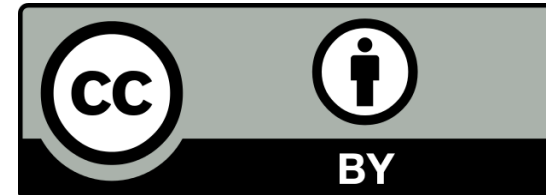Supported by:
Federal Ministry
for Economic Affairs
and Energy

on the basis of a decision
by the German Bundestag

Research Grant 03EI1005A-G

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:

„Implementing FAIR through a distributed data infrastructure" © Hoyer-Klick, et. Al. | CC BY 4.0