
DTU

DOI: 10.5281/zenodo.3865225

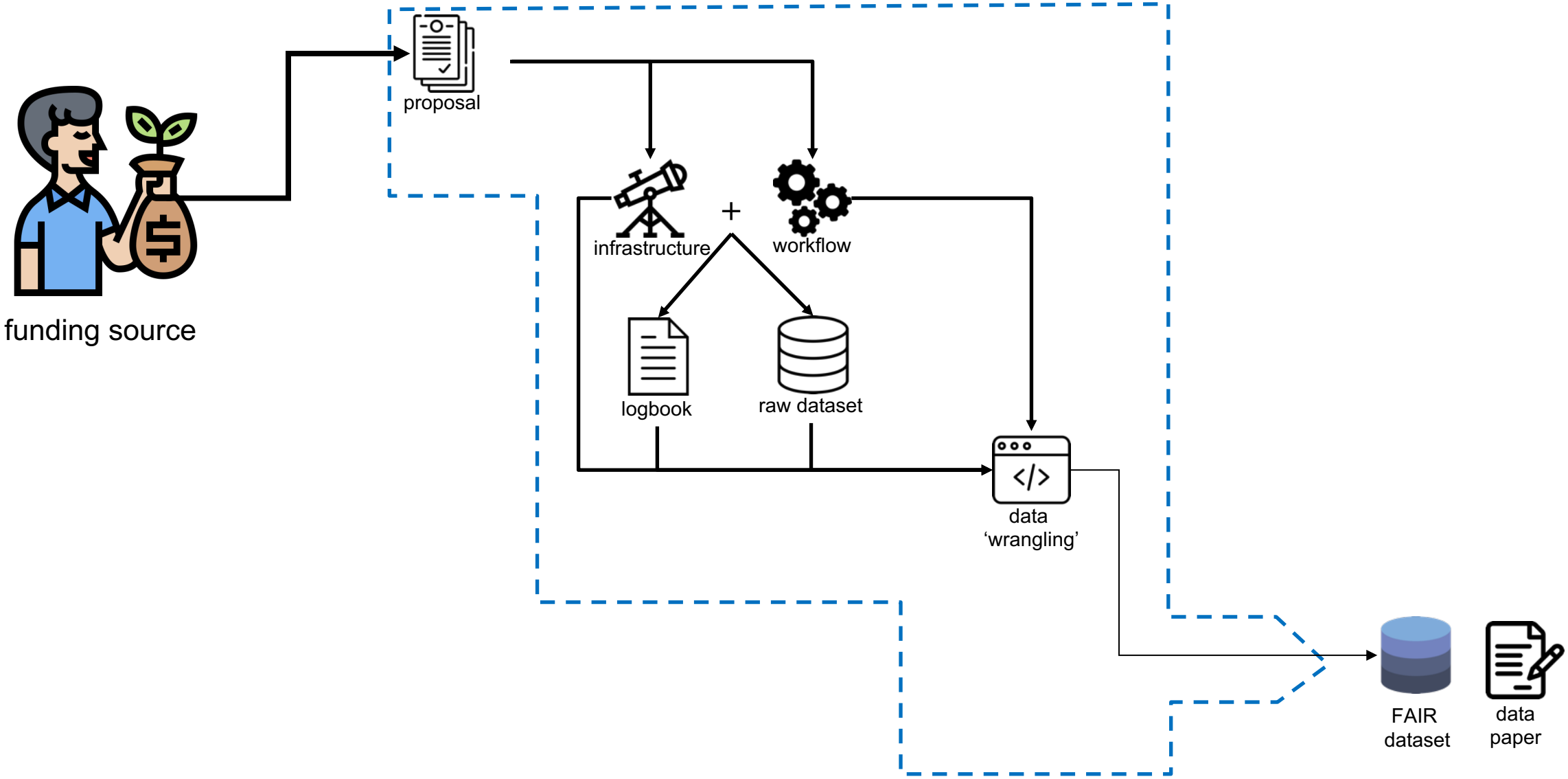
29/05/2020

Theme Day on Research in a FAIR Data Perspective

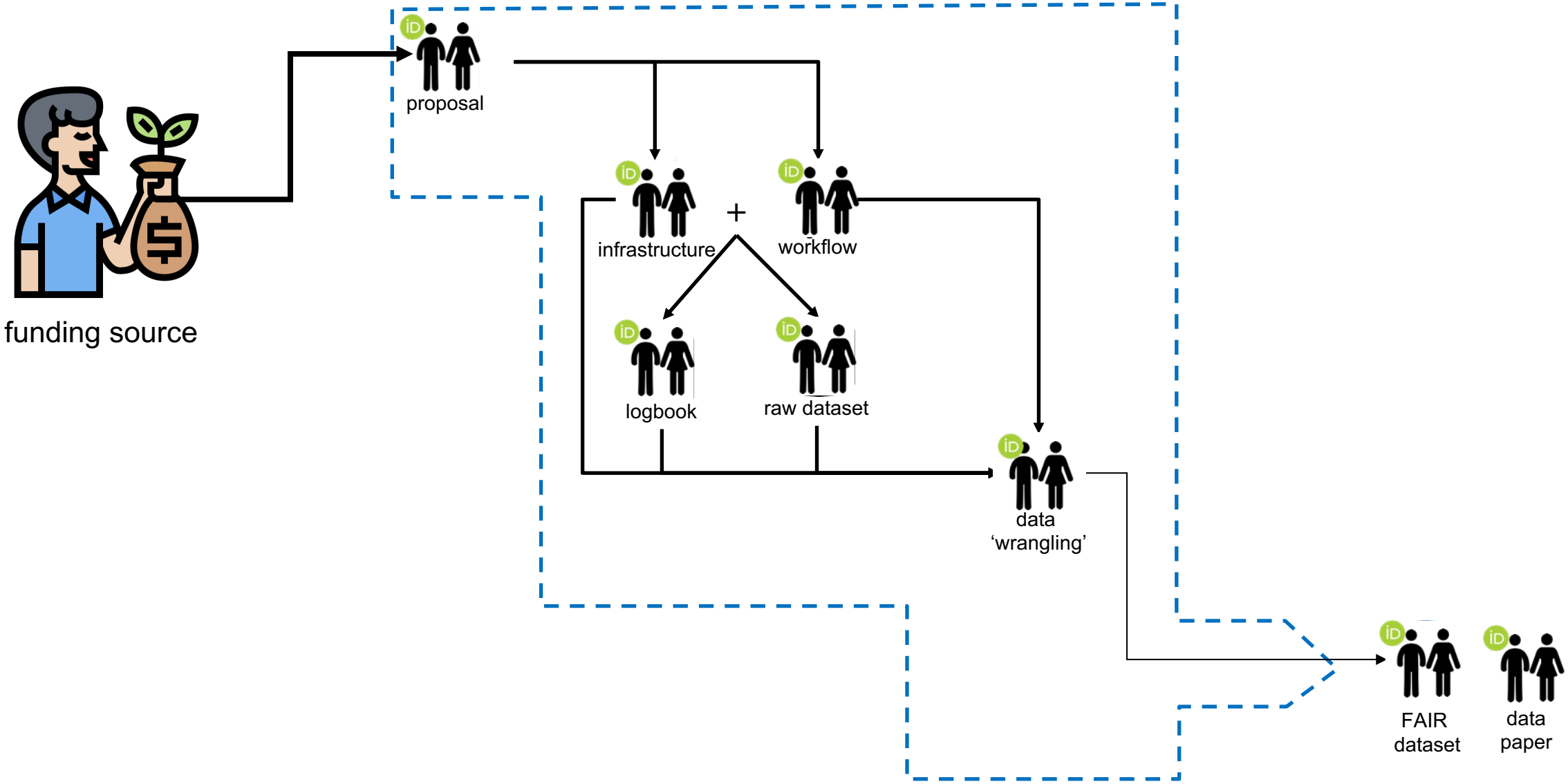
Nikola Vasiljevic

Wind Energy goes FAIR – current status and future perspective

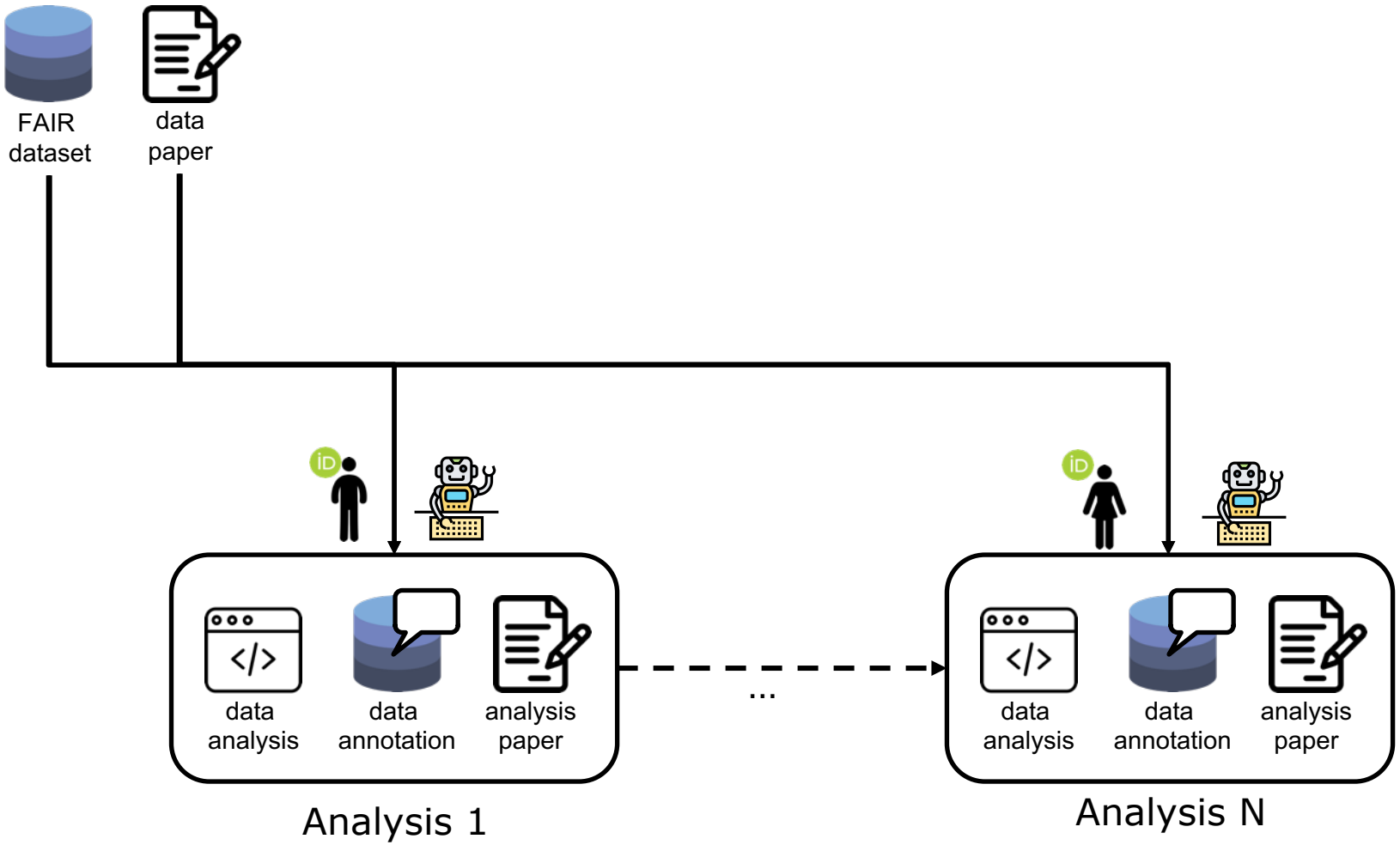
Blueprint of idealized data lifecycle



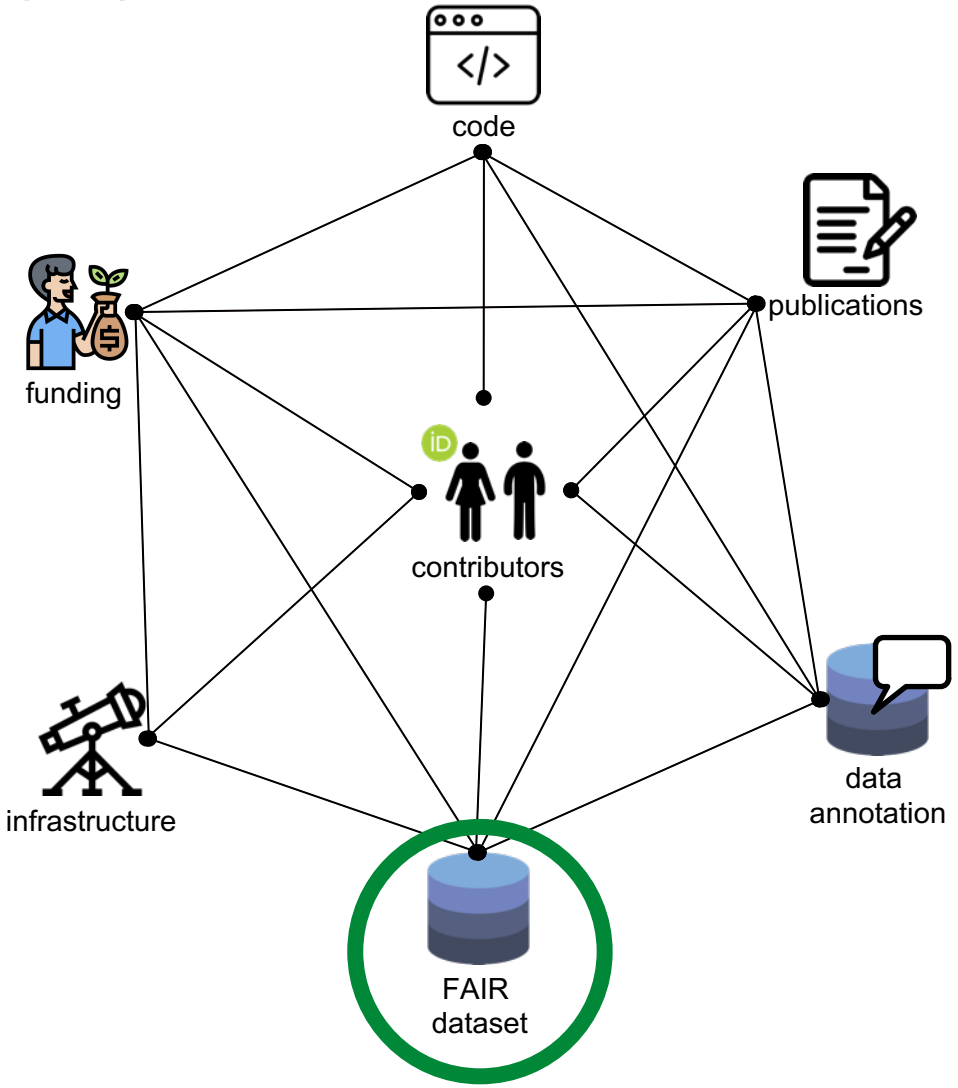
Blueprint of idealized data lifecycle



Blueprint of idealized data lifecycle



Relations



Full chain of custody

Researcher at center

It is a graph!

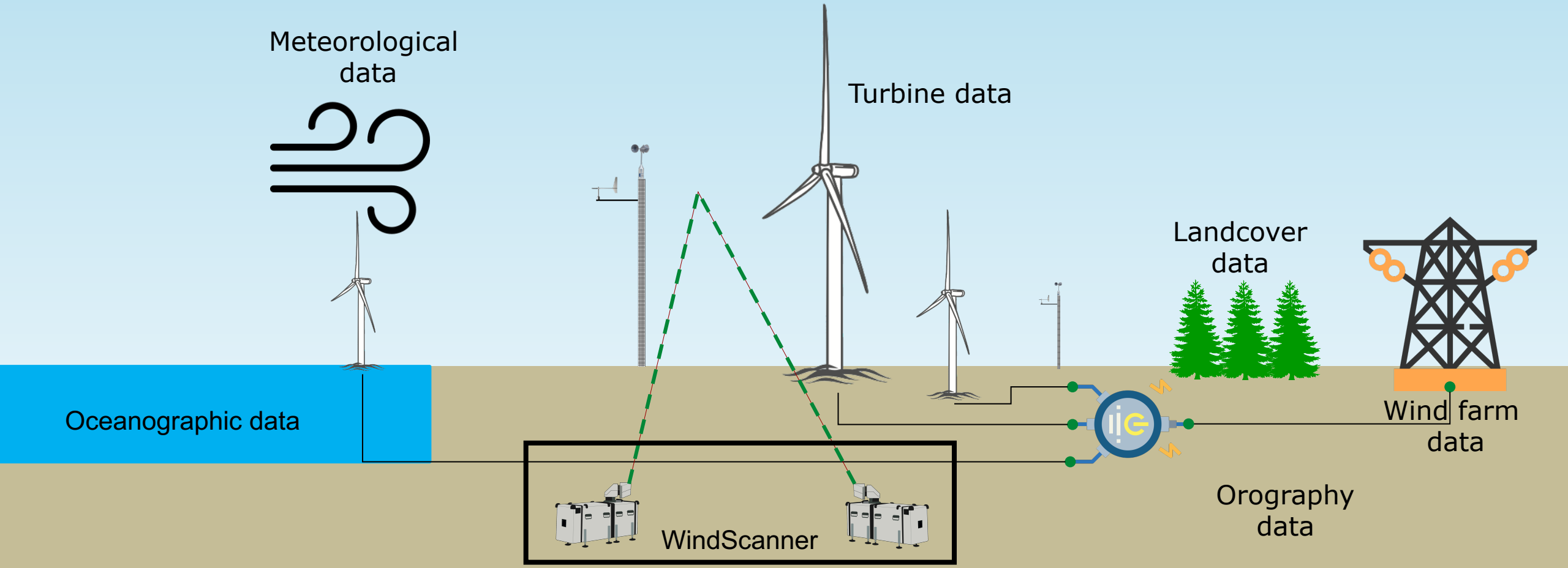
Wind Energy domain

WindScanner is just a small piece in the whole wind energy puzzle

Large majority of our data are spatio-temporal data

All sort of spatio-temporal scales

Open and confidential data

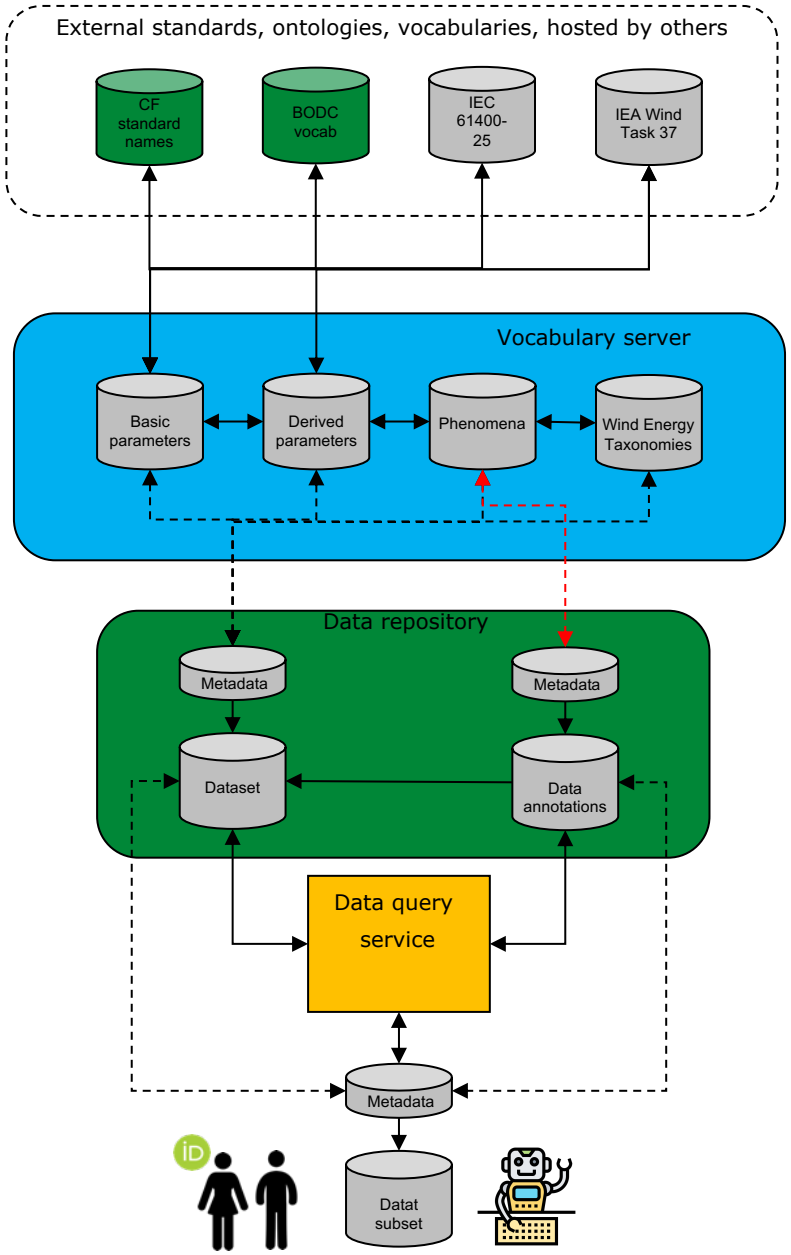


Physical infrastructures DTU Wind Energy operates

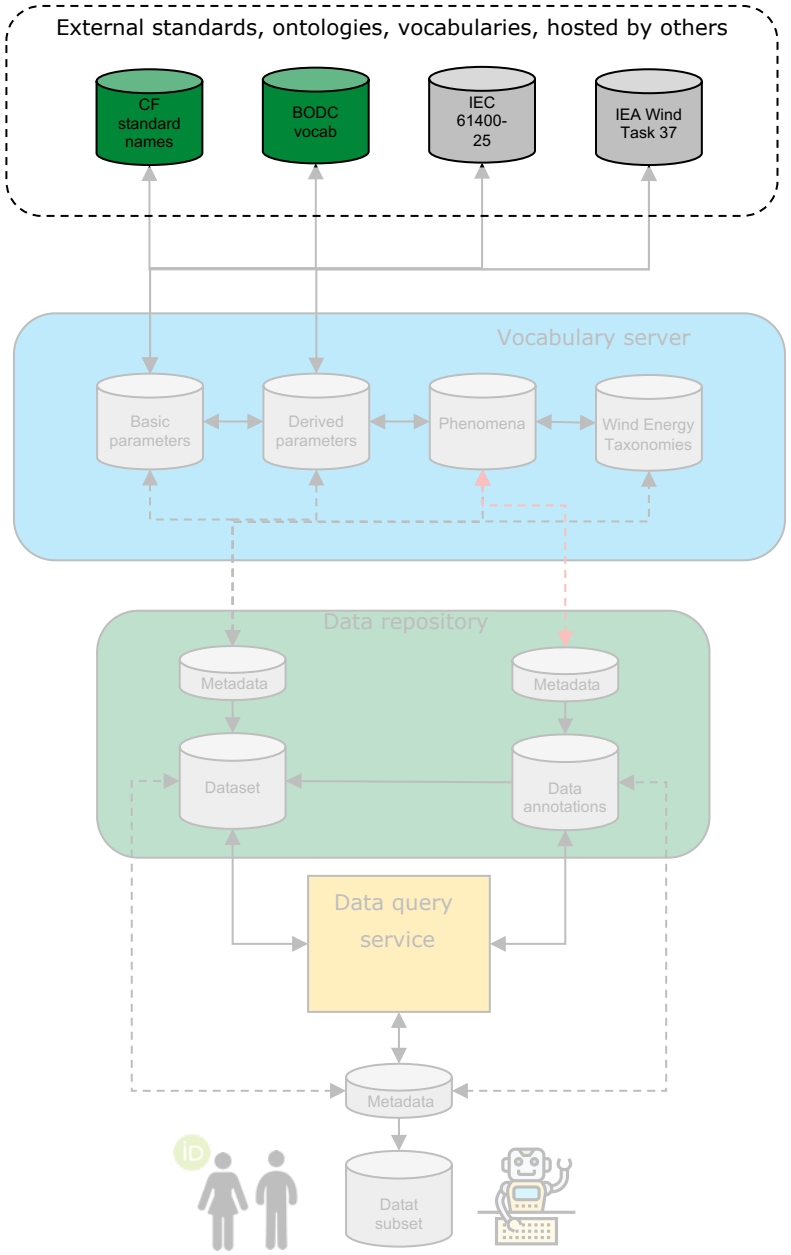
- Test Centre for large wind turbines Høvsøre
- Test Centre for large wind turbines Østerild
- Research Facilities:
  - Composite Laboratories
  - Drivetrain
  - Large Scale Facility
  - Poul la Cour Wind Tunnel
  - Research Wind Turbine V52 (Risø test station)
  - Windscanner

  …

  (+ a number of virtual infrastructures)

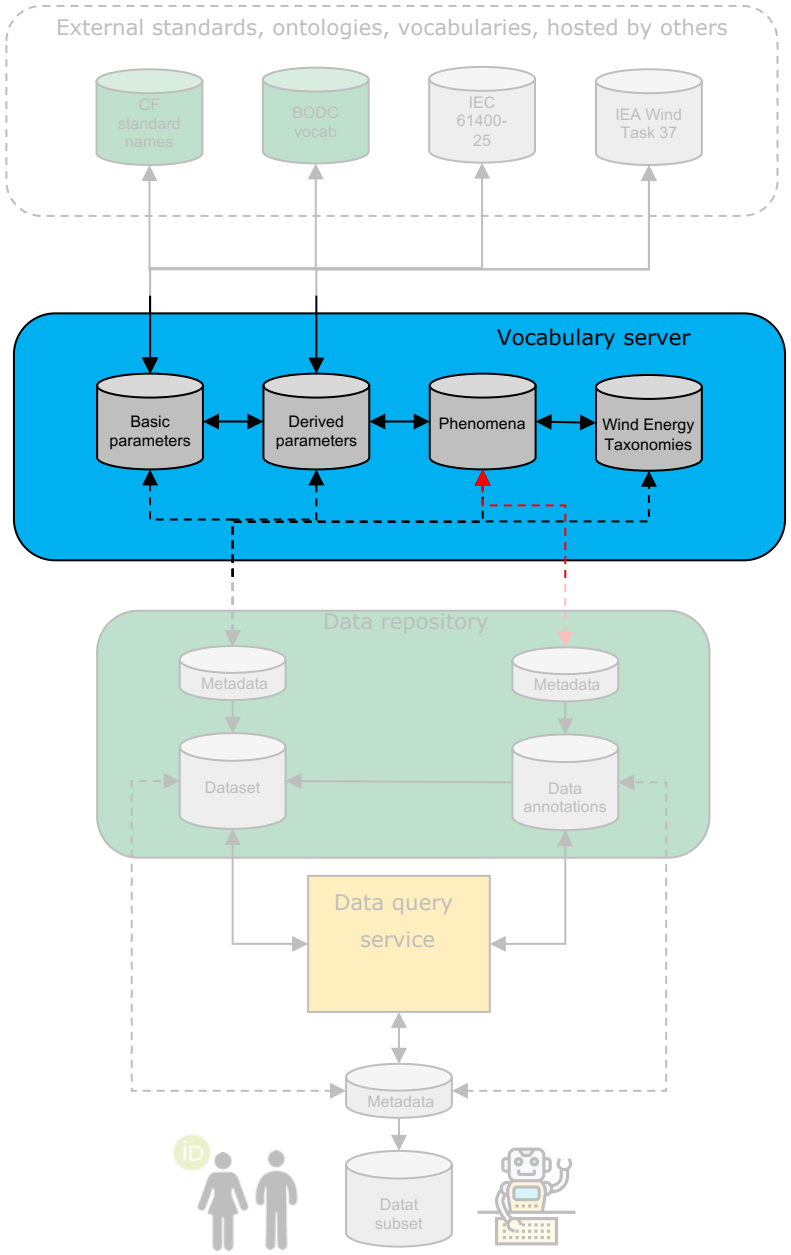


clipart from www.flaticon.com

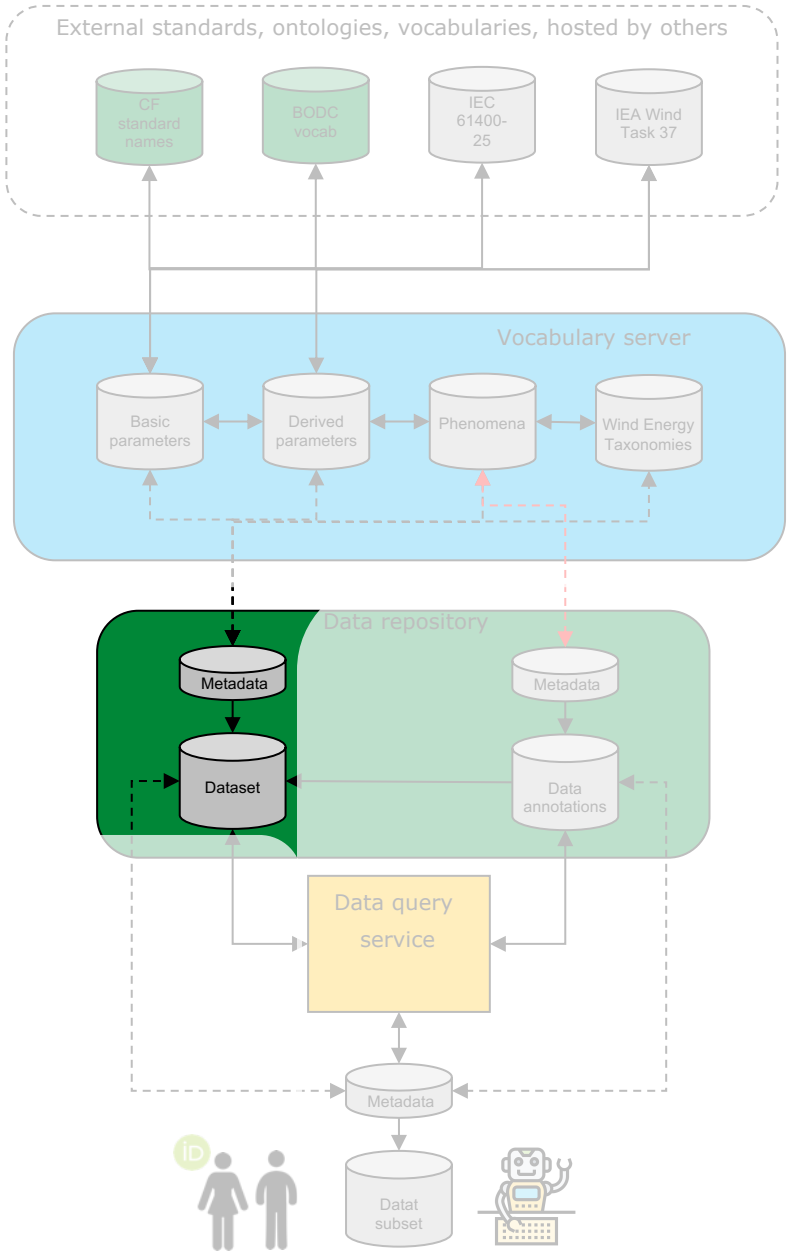


- There are number of external vocabularies and ontologies which can be reused to build domain specific ontologies
- Some of them are in good format (i.e. RDF) like those hosted by BODC server
- Some are simple human-readable web pages
- Also, there a number of domain specific standards, some as PDFs, while some in JSON which can be used
- First step would be to scrap them and put them in a common format. Aftewards, ‘cherry-pick’ those which are useful for the domain.
- For any of the above work resources are needed

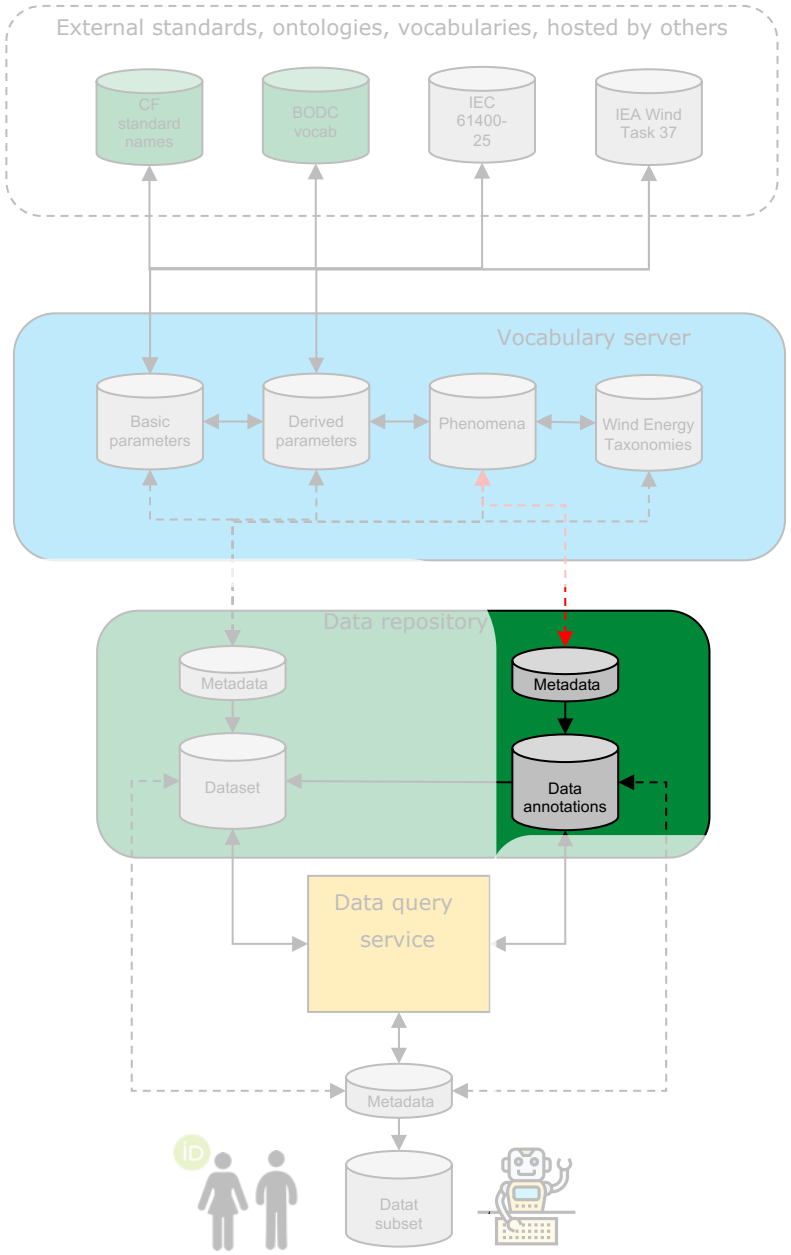
clipart from www.flaticon.com



- A work has been done to build domain specific taxonomies for ‘tagging’ of wind energy data
- These are: taxonomy of research topics, activities, external conditions, …
- Taxonomies were initial provided as PDF, then converted to JSON, and are missing definition of their terms (some can be sourced from IEC Electropedia)
- They need to be converted to RDF format and deployed on a dedicated vocabularly server
- However, we need a good RDF schema for encoding ontologies as well resources for this work
- Probably we will intially host the server at DTU Wind Energy, however …



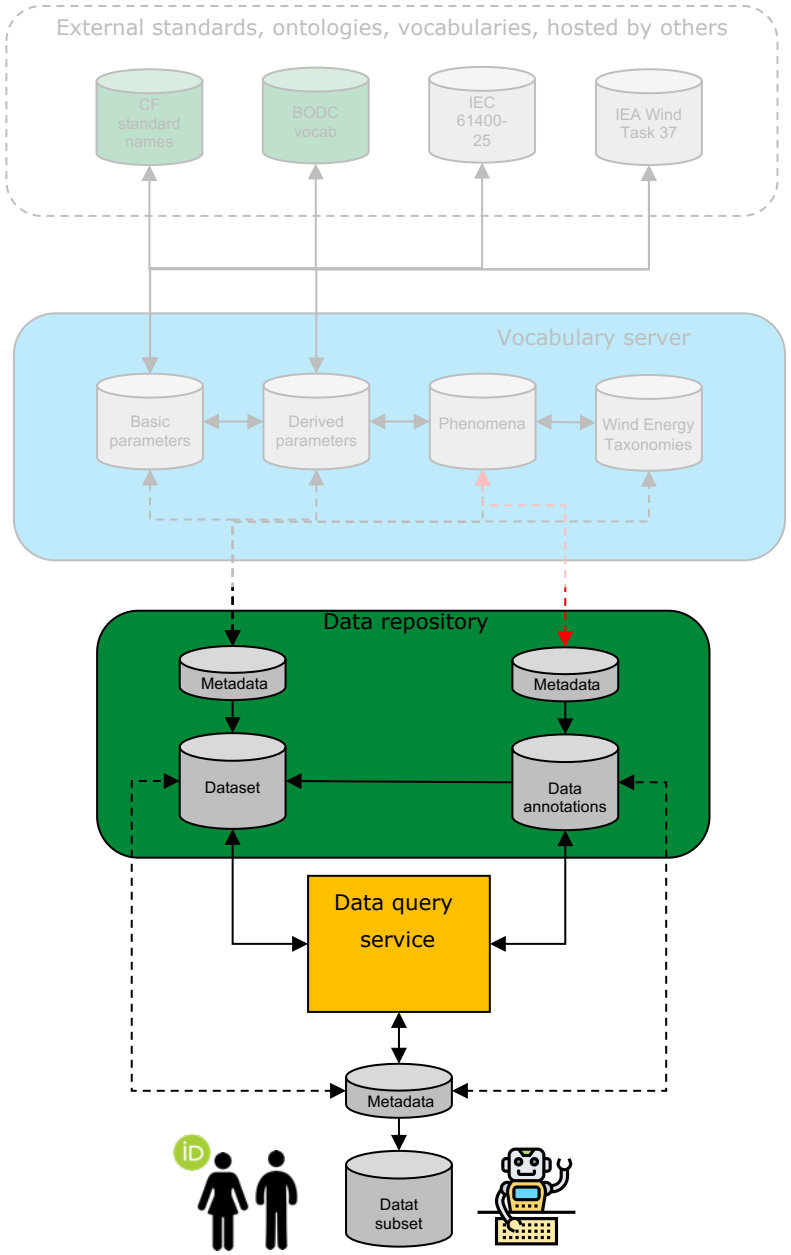
- Currently developing detailed schema for data streams and datasets
- In process of organizing Metadata4Machines workshop/hackathon with GO-FAIR (DEIC, DTU, KU)
- Strong advocacy for NetCDF, a binary data format which fuses both data and metadata



- Early proof-of-concept of data annotation and data annotation related services built as a demonstrator in colaboration with the chair of the RDA Preserving Scientific Annotation Working Group

- Based on the proof-of-concept application sent to DFF for funding to build open and domain agnostic platform for Rapid Generation and Preservation of Knowledge and Information from data (proposal was rejected)

- Work currently on pause, however a potential schema for data annotation was build by extending the RDA WG Dynamic Data Citation output

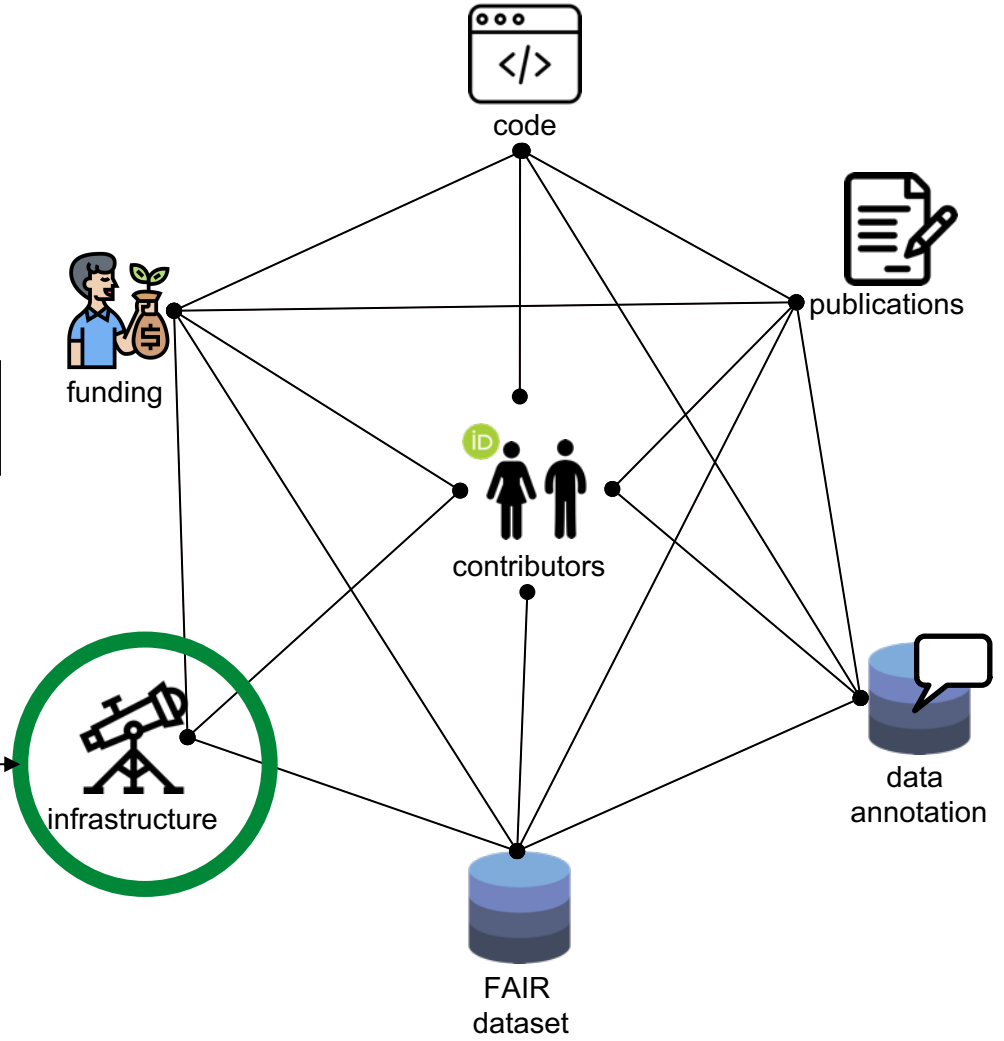


- We need a good solution for data repository which will host both datasets as well data annotations which help in an efficient dataset subsetting
- Also, we need a solution which allows simple integration of domain specific onotologies in metadata
- We have succesfuly demonstrated implementation of wind energy taxonomies in DTU Data, which is an instance of figshare
- However, no generic data repository provides data query service for subsetting datasets in accordance to certain criterias (e.g., time-space range)
- Also generic repositories offer very limited metadata for datasets and are not keen on extending metadata schema

Beyond data

Testing PIDINST schema

Proposed by the RDA WG PID Instruments



Future perspective

- Joint work with domains with whom wind energy has a large overlap (Earth science, engineering, etc.) on building:
  - common metadata schemas
  - linking ontologies
  - data query services
  - …

- We need (meta)data hackathons focused on tech since we need to move from a ‘talk’ to the ‘actual work’

- Hopefully, funding agencies will fund this type of work, since chipping and scrapping for resources from scientific projects would only ‘take away’ money from researchers. However, it is still viable, but slow, approach in becoming FAIR.

- We applied for National Energy System Transition Facilities (NEST) where we allocated resources for establishing a pool of data stewards which could substantially lift our work

Future perspective

- Joint work

(meta)data hackathons tech

funding

FAIR

establishing a pool of data stewards which could substantially lift our work

Thank you for your attention
Questions?

niva@dtu.dk