# AD4GD semantic interoperability approach for a Green Deal Data Space

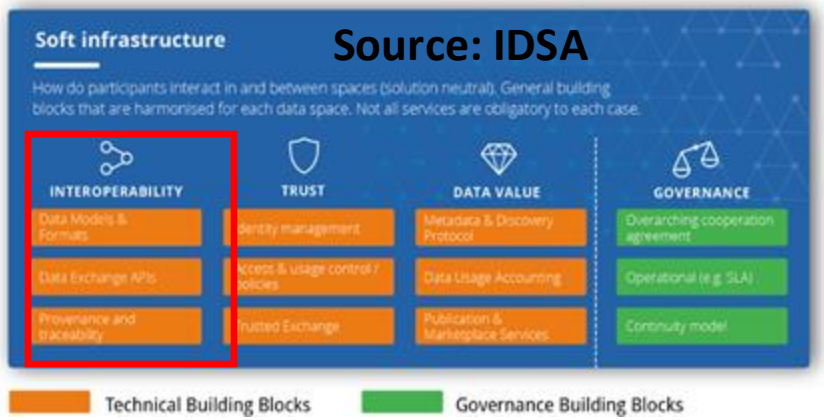
Semantic Interoperability in Data Spaces
IDSA Workshop
01/10/2024

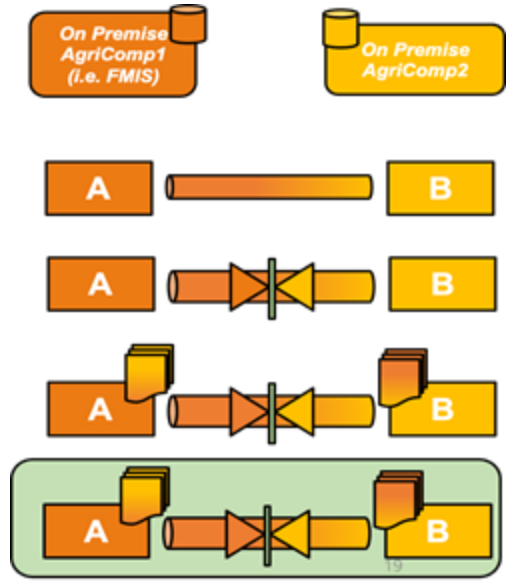
Raul Palma (PSNC)
Bogusz Janiak, Rob Atkinson, Piotr Zaborowski, Alejandro Villar, Francesca Noardo, Ivette Serral, Alba Brobia Ansoleaga, Joan Maso Pau, Lucy Bastin


Co-funded by the European Union, Switzerland and the United Kingdom

# AD4GD context

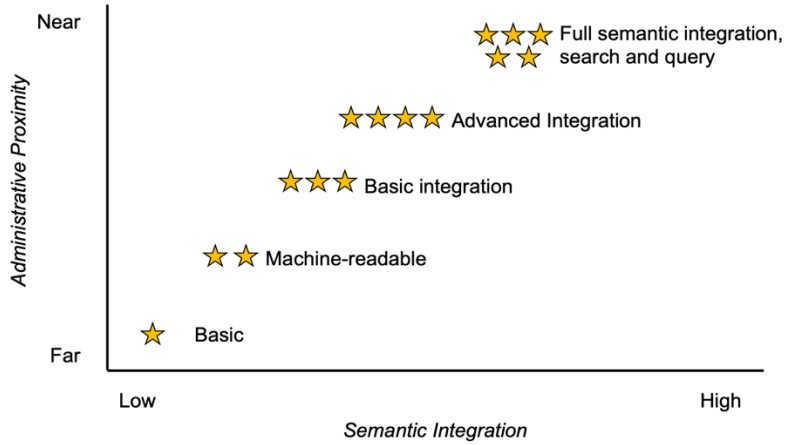
- Aims to deliver the building blocks enabling semantic interoperability in the GDDS
  - Addressing the growing set of heterogenous data and services available in the areas addressed by the GD
  - enabling an unambiguous exchange data between different systems,
  - enabling integrated data access for advanced data analytics
  - and in alignment with existing recommendations and ongoing initiatives

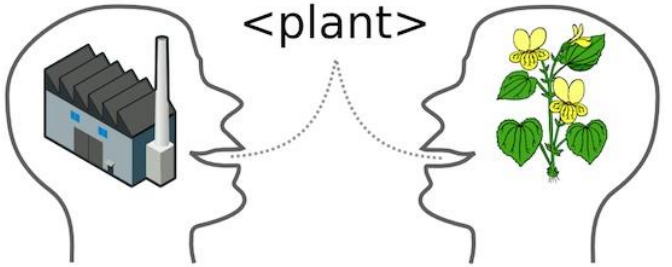


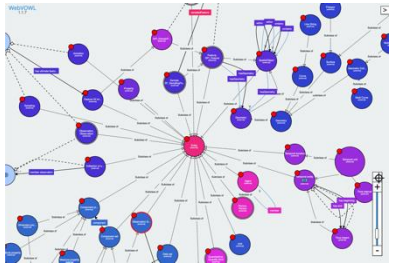



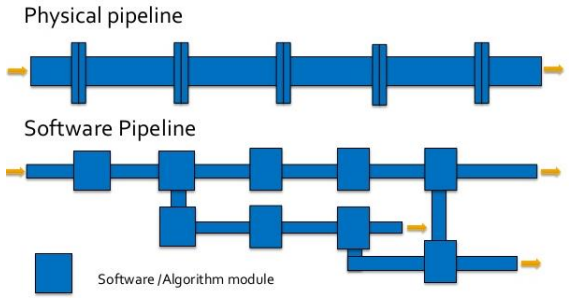
# AD4GD approach

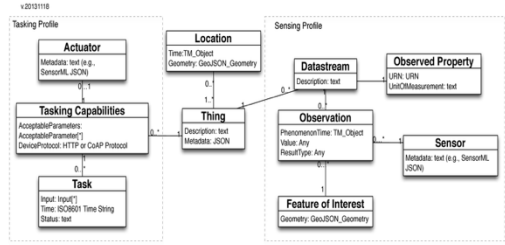
- Follows an incremental model towards a full semantic integration based on the pay-as-you-go (PAYG) data management approach
  - Inspired by the 5-star rating scheme defined for Linked Data by Tim Berners-Lee
- Involves:
  - Common information models (ontologies, taxonomies, or CVs)
    - Green Deal Information Model (GDIM)
  - Interoperability mechanisms to produce/consume the exchanged data
    - Data harmonization pipelines
  - And the provision of integrated data access interfaces
    - via standardized APIs



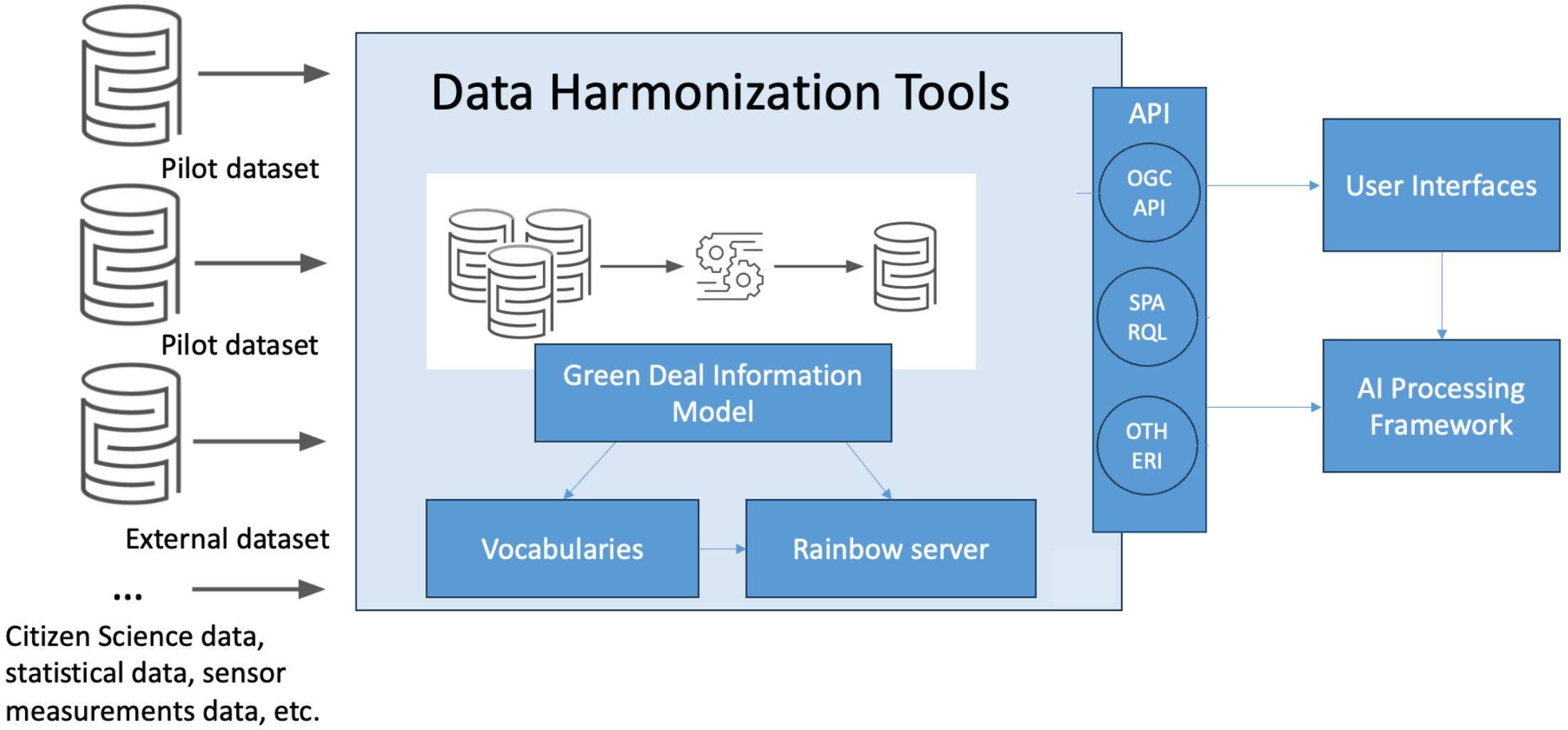





# High level view

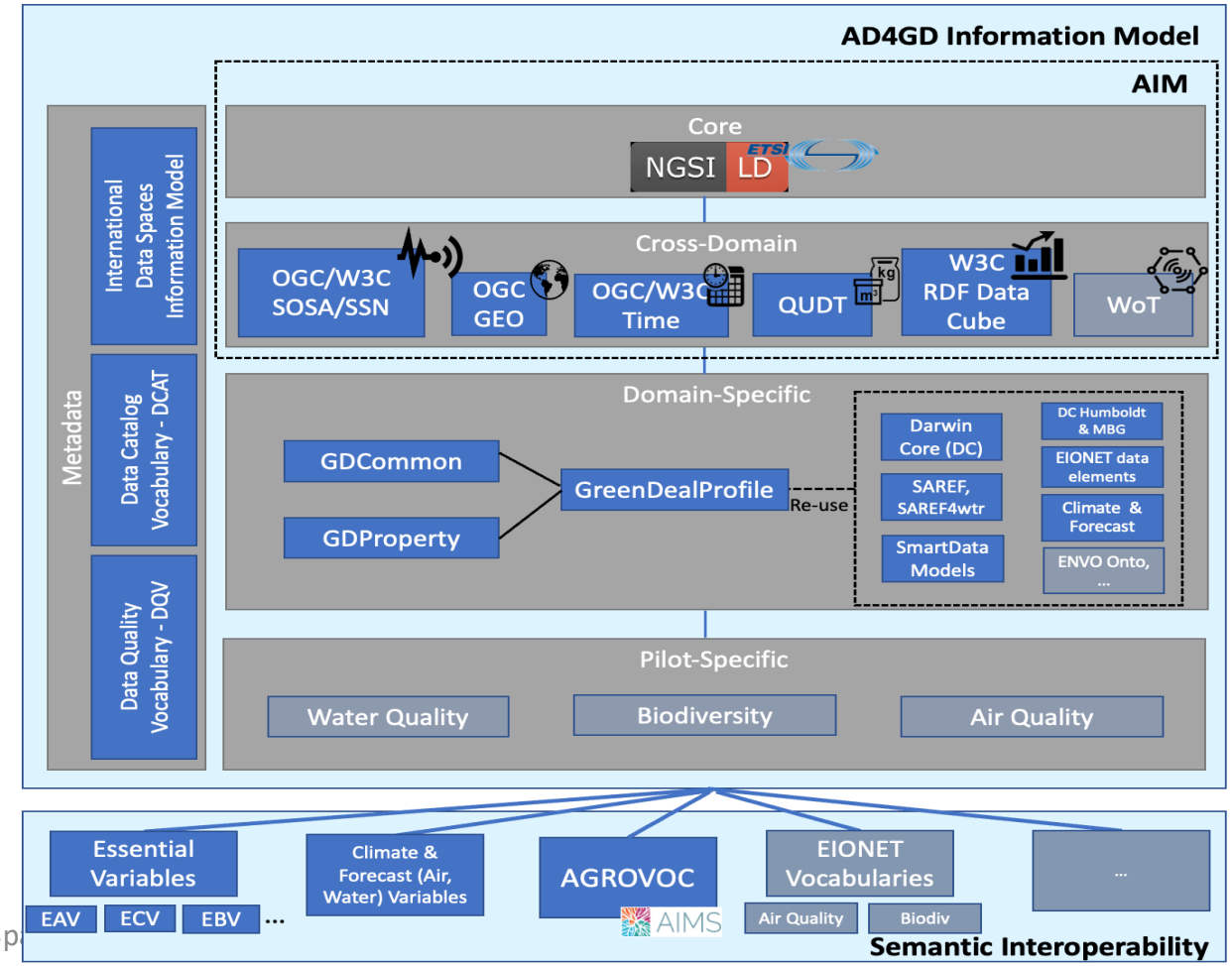


# GDIM – Green Deal information model

## Building on previous experiences….

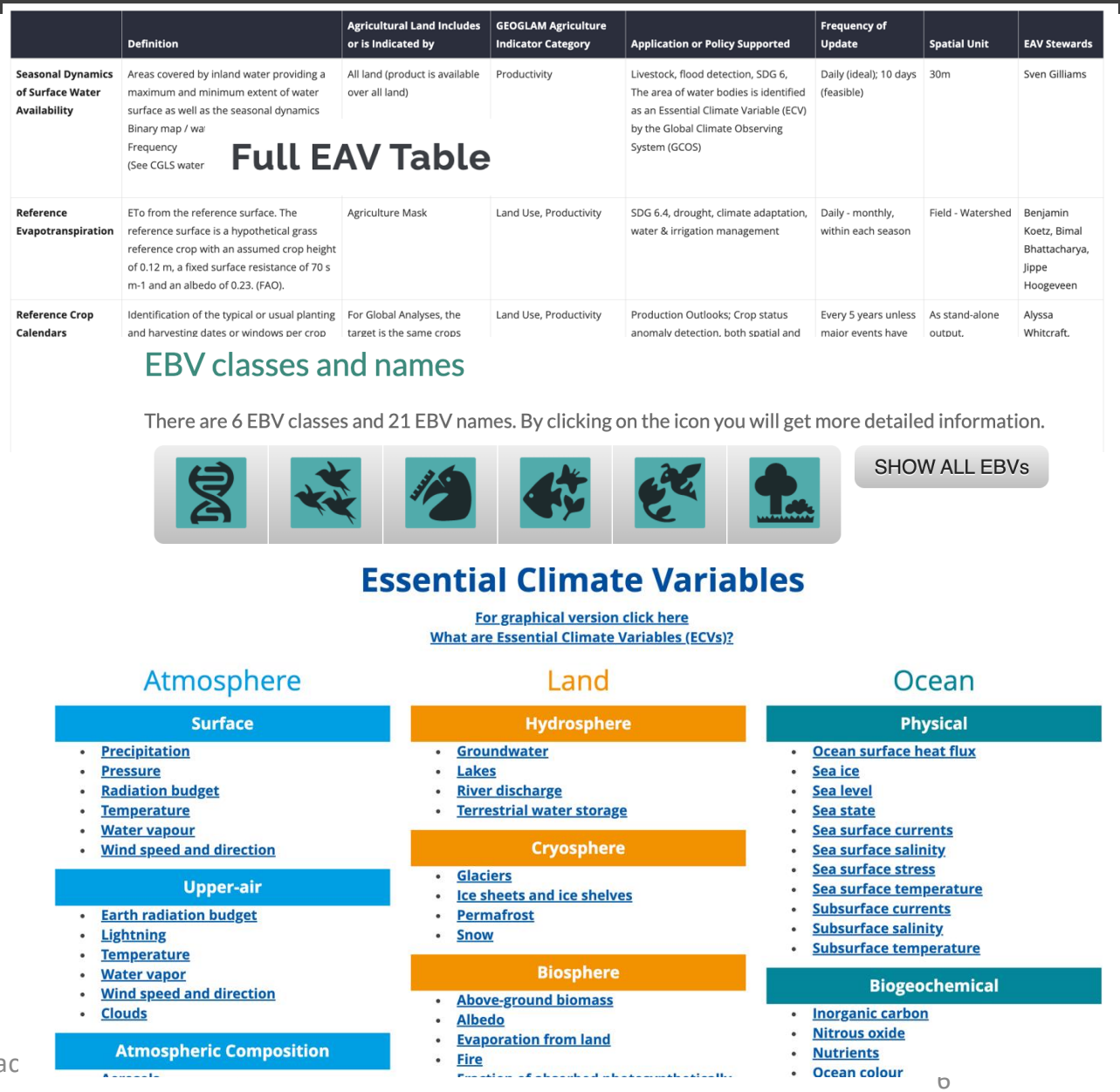
- GDIM aims to establish the basis of a common green deal data space, enable the interoperability of different systems, and the analysis of data produced by those systems in an integrated manner

- The information model is being:
  - designed following a **modular** approach in a **layered** architecture, taking into account data requirements in AD4GD
  - realized as a **suite of ontologies** and corresponding **JSON-LD contexts**, published and accessible via PIDs (e.g., w3id.org)
  - implemented in line with **best practices**, **reusing existing standards** and well-scoped models
  - **establishing alignments** with other relevant models to enable their interoperability and the integration of existing data



# Alignment with vocabularies

- Essential Variables
  - EAV, EBV, ECV, …
  - ObservableProperty axioms
- Climate & Forecast (variables)
  - ontology representing the climatic data variables for use with climate and forecast data, in the atmosphere, surface and ocean domains.
  - ObservableProperty axioms
- Agrovoc
  - controlled vocabulary for food, nutrition, agriculture, fisheries, forestry and env.
  - SmartDataModels property “agroVocConcept”



# Essential Variables in OGC Rainbow Server

## Vocabulary

Example codelist ebv
http://w3id.org/ad4gd/ev/ebv

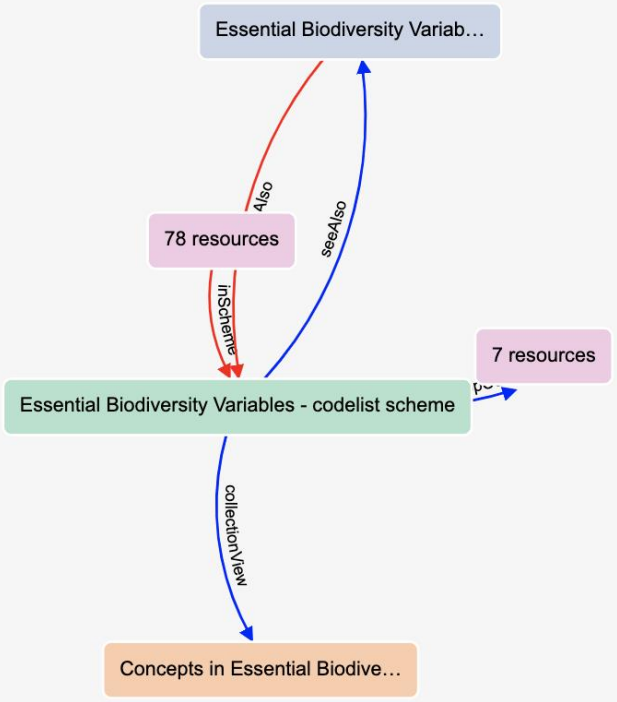
Example variable
http://w3id.org/ad4gd/ev/ebv/Spp_SP_abn_bird_FW_Count

**Preferred Label**

### Essential Biodiversity Variables - codelist scheme

**URI**

http://w3id.org/ad4gd/ev/ebv ↗




Collection Class

**Definition**

Essential Biodiversity Variables (EBVs) are defined as a minimum set of measurements, complementary to one another, that can capture major dimensions of biodiversity change. EBVs are organized in six classes (Genetic composition, Species populations, Species traits, Community composition, Ecosystem functioning, Ecosystem structure) and cover the three realms (Marine/coastal, Terrestrial and Freshwater)

**Search**

Your search term

Search

**Concept Hierarchy**

expand all, click '+' to expand individually

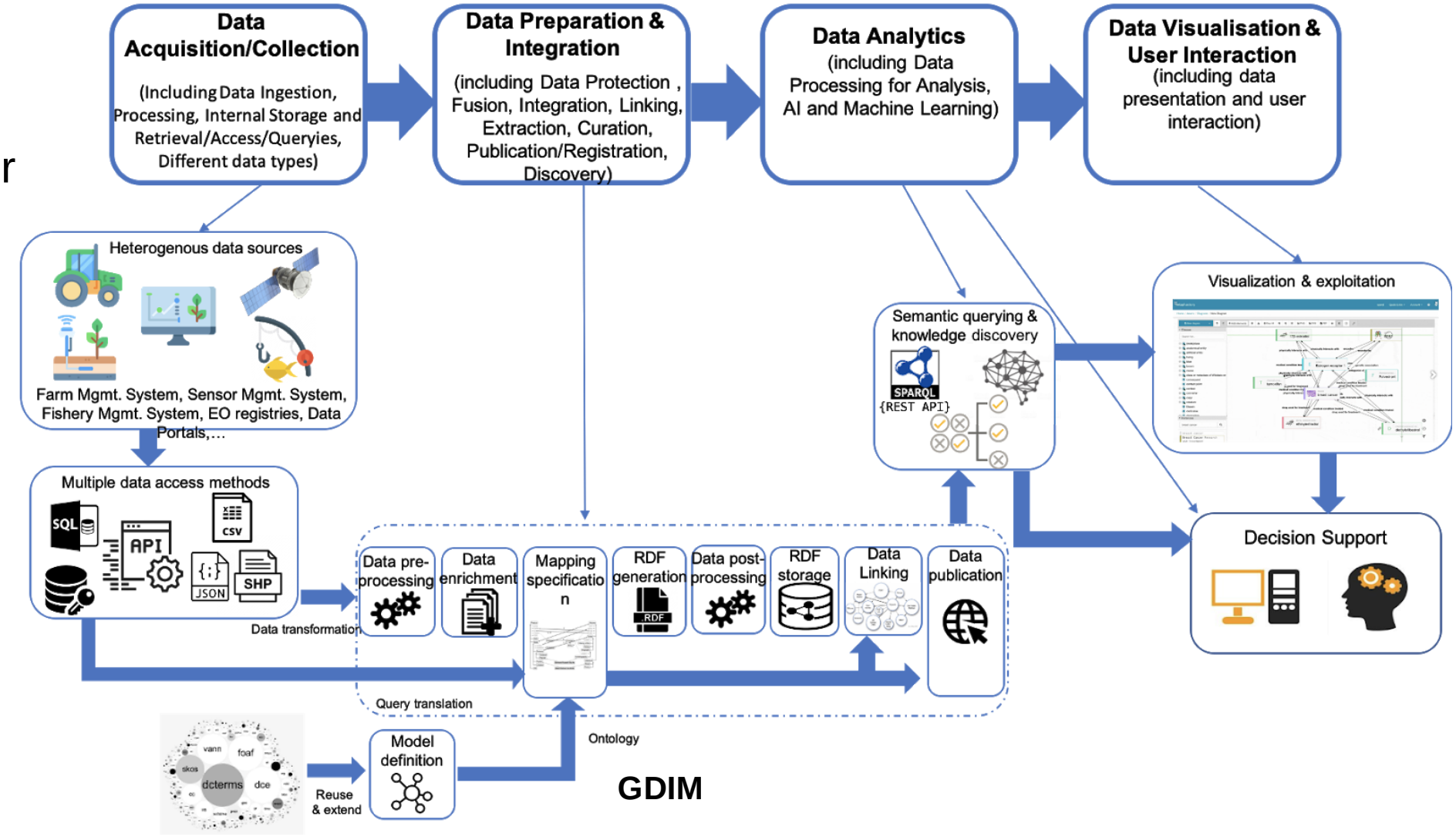
- − Community composition
  - Aerial biomass of migrating birds, bats and insects
  - Community abundance and taxonomic diversity of pollinator insects
  - Community biomass of selected functional groups of terrestrial arthropods (e.g. predator, decomposer)
  - Community biomass of soil microbes
  - Ecological Quality Ratio (EQR) of benthic freshwater invertebrates
  - Ecological Quality Ratio (EQR) of freshwater fish
  - Ecological Quality Ratio (EQR) of freshwater macrophytes
  - Ecological Quality Ratio (EQR) of freshwater phytobenthos
  - Ecological Quality Ratio (EQR) of freshwater zooplankton
  - Ecological Quality Ratio (EQR) of phytoplankton in lakes
  - Functional composition of marine phyto/zooplankton (based on traits)
  - Functional composition of soil biota
- + Ecosystem function
- + Ecosystem structure
- + Genetic composition
- + Species populations
- + Species traits

Aligned with I-ADOPT

**Description** Essential Biodiversity Variables (EBVs) are defined as a minimum set of measurements, complementary to one another, that can capture major dimensions of biodiversity change. EBVs are organized in six classes (Genetic composition, Species populations, Species traits, Community composition, Ecosystem functioning, Ecosystem structure) and cover the three realms (Marine/coastal, Terrestrial and Freshwater)

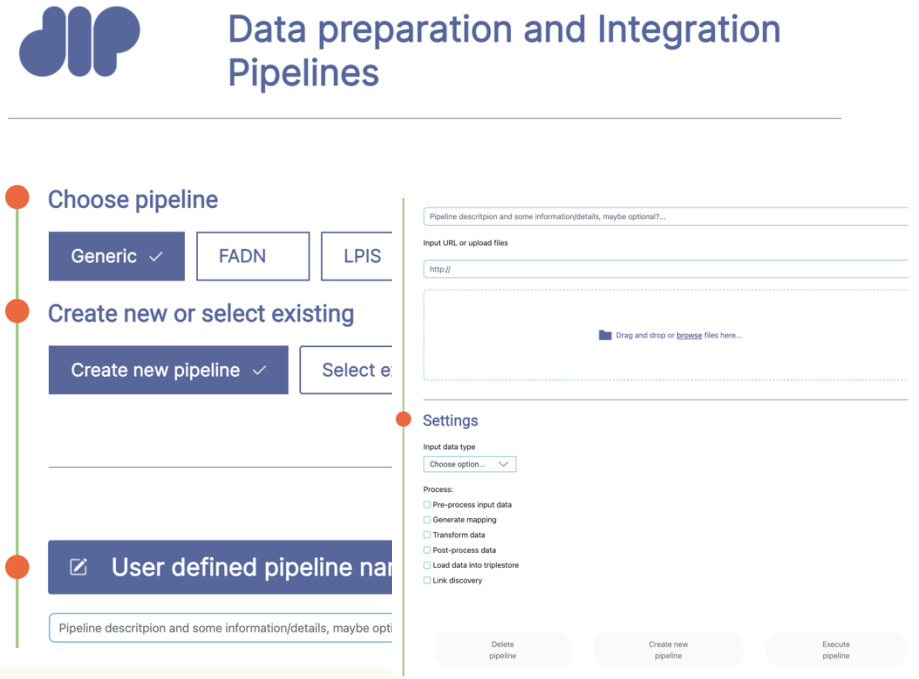
# Data harmonization approach

- Methods and tools for preparing, processing, and integrating incoming datasets
- Reuse, adapt and extend Data Preparation and Integration pipelines (DPI) tools based on linked data

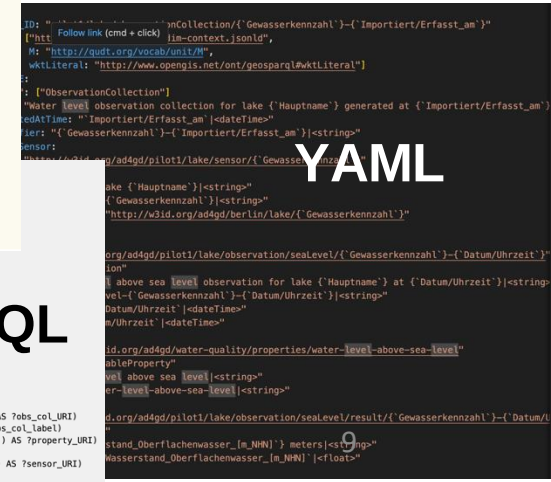


# DPI in practice

- The DPI is an ETL software toolkit
  - Executes pipelines including, data collection, preprocessing, transformation, post-processing, linking and loading of the generated linked data into a triple store.
  - Re-uses several existing tools, for particular tasks, providing a unified interface over them and connecting them in sequences to implement full pipelines
- The DPI is implemented in Python, and available via
  - CLI tool distributed under MIT license with a docker file and a singularity file to simplify deployment in different environment, including HPC.
    - https://gitlab.pcss.pl/daisd-public/dpi-pipelines/pipelines
    - Docker image available in DockerHub: https://hub.docker.com/r/montanaz0r/demeter-pipelines
  - RESTful Web service available via:
    - https://dpi-enabler-dpi-enabler.apps.dcw1.paas.psnc.pl/api/
    - https://dpi-enabler-dpi-enabler.apps.dcw1.paas.psnc.pl/api/swagger
  - Web client application
    - https://dpi-enabler-ui-test.apps.paas-dev.psnc.pl/

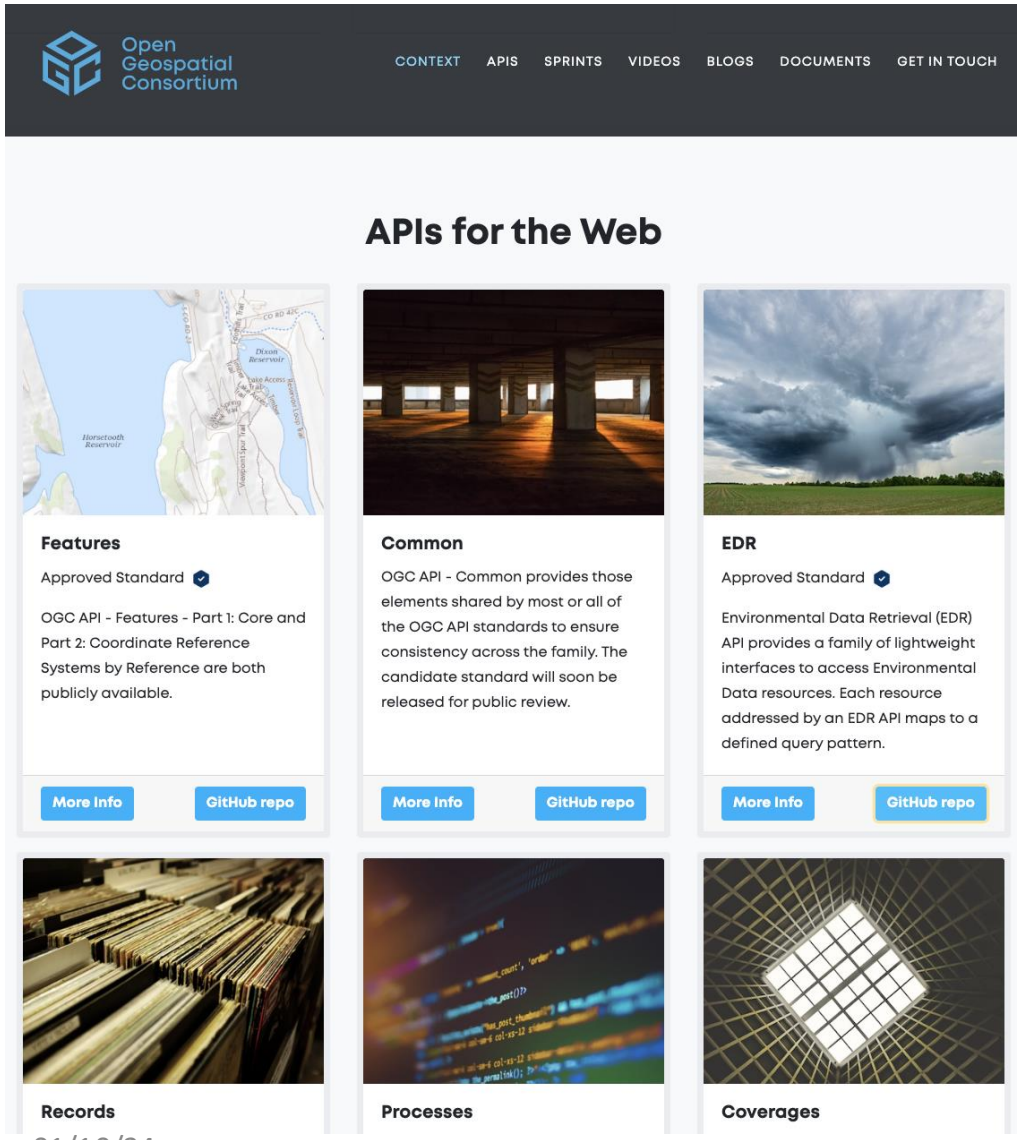




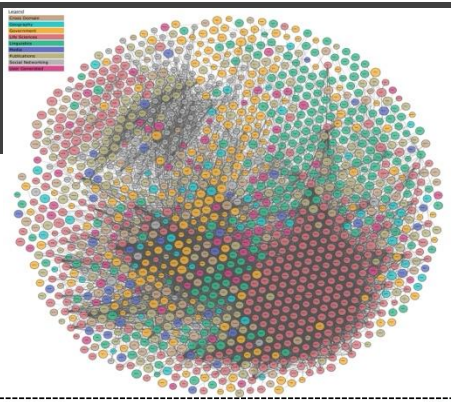



# GDDS open APIs

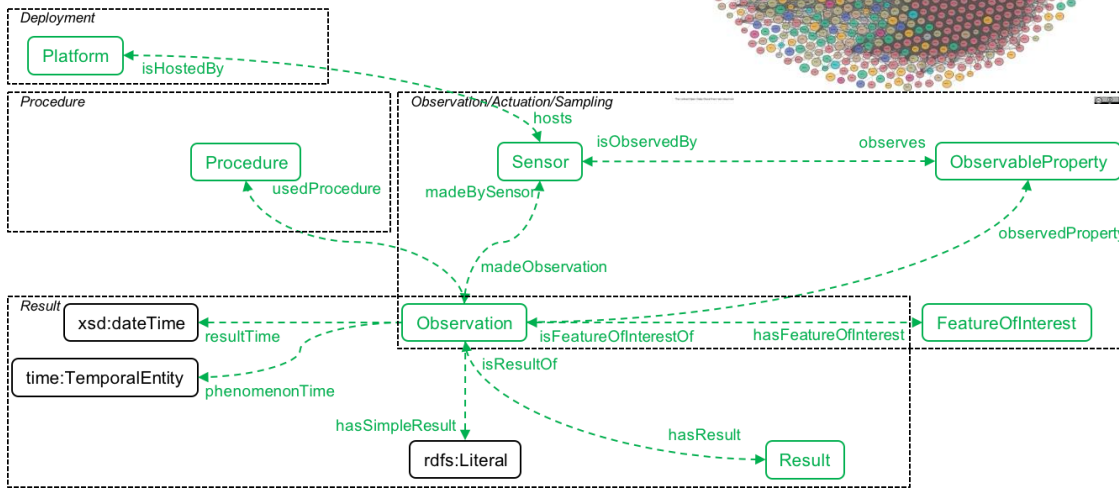
- GDIM provides the common language (lingua franca) to represent data, with explicit semantics, so that different components can understand and validate it
- The DPI allows the harmonization of data according to those models in order to enable an integrated view over different data source
- However, different components normally implement different APIs that expose or consume the data
  - Specify the APIs needed to provide the core services of the AD4GD ecosystem
    - Enabling to build FAIR science services
    - Based on GDIM and aligned with identified requirements
    - aligned with international standardization efforts

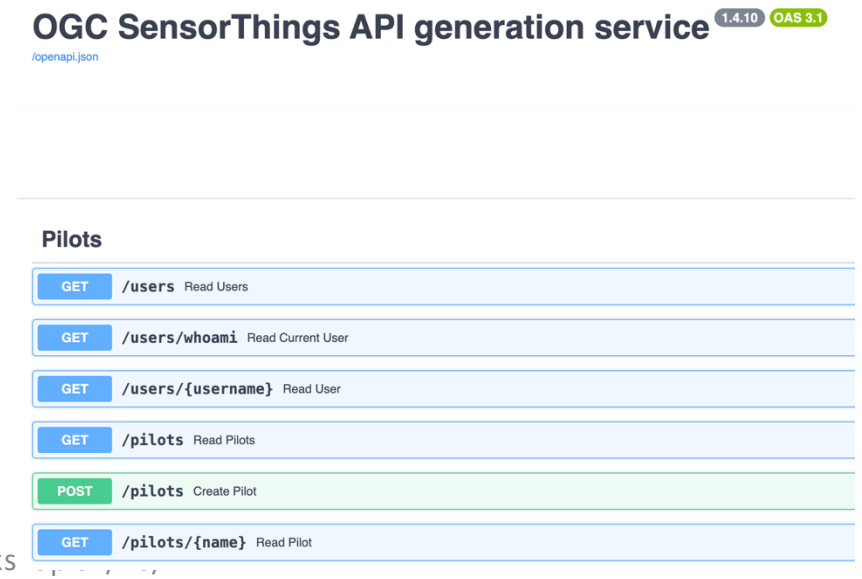


# STA generation service



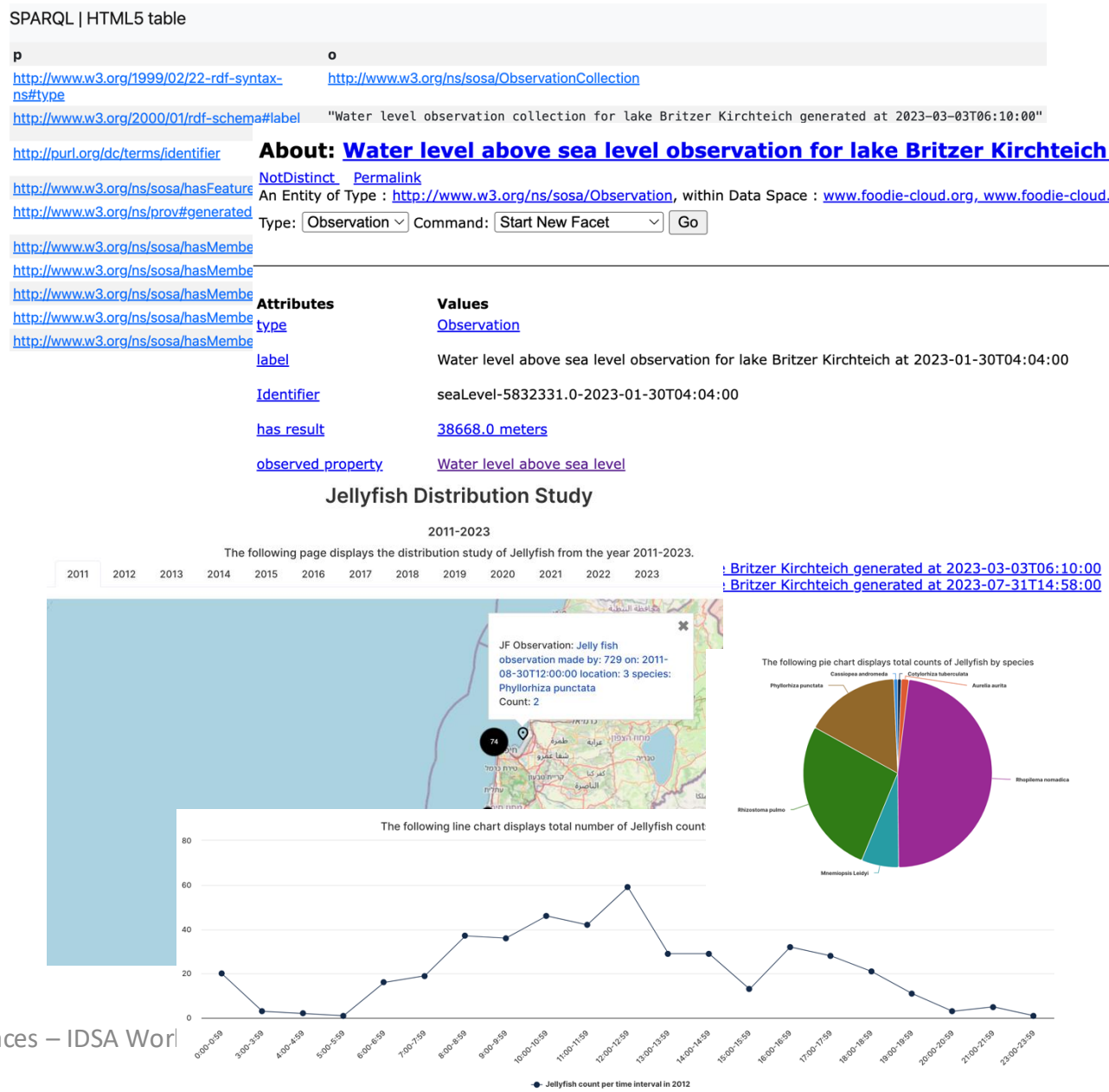
- Enables  access to Linked Data about observations that is stored in a triplestore, e.g., generated by pipelines, via an STA.
- The endpoint is created automatically from the specified dataset(s), identified as graph(s) in the triplestore.
- The dataset(s) should represent observations represented following the OGC SOSA/SSN standard model (base of GDIM).
- the service returns URL of STA and its corresponding Swagger interface
- The generated STA endpoint can be public or private and supports content negotiation for JSON or JSON-LD



# Other interfaces to access the harmonized data

- SPARQL endpoint
  - https://www.foodie-cloud.org/sparql
  - Query and/or download data
  - Also possible to generate on the fly (virtual sparql endpoint) directly over a database or existing REST API
- Faceted browser
  - https://www.foodie-cloud.org/fct/
- Custom REST API
  - https://grlc-dpi-enabler-demeter.apps.paas-dev.psnc.pl/
  - Generated on-the-fly from SPARQL query
  - Queries can be provided via GitHub repository or an specification file
- Web client apps, e.g.,
  - Metaphactory



# Discussion and questions

Raul Palma (PSNC)
rpalma@man.poznan.pl

Co-funded by the European Union, Switzerland and the United Kingdom