

AD4GD semantic interoperability approach for a Green Deal Data Space

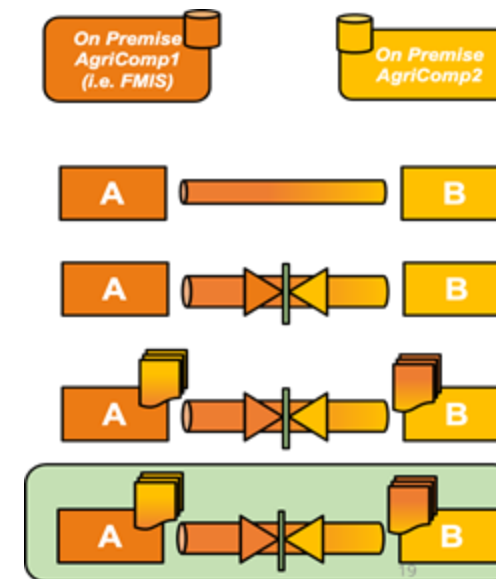
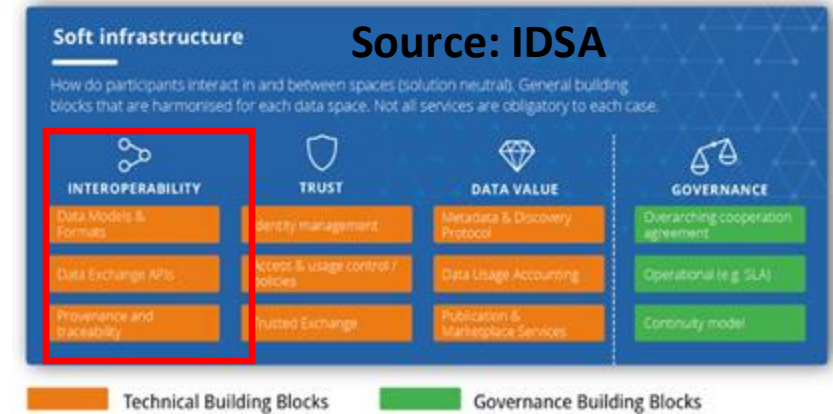
Semantic Interoperability in Data Spaces
IDSA Workshop
01/10/2024

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



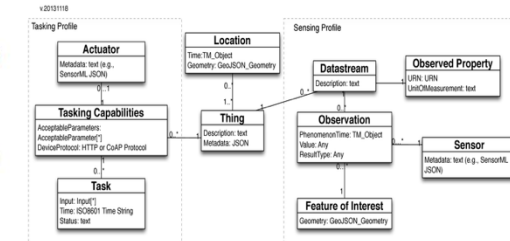
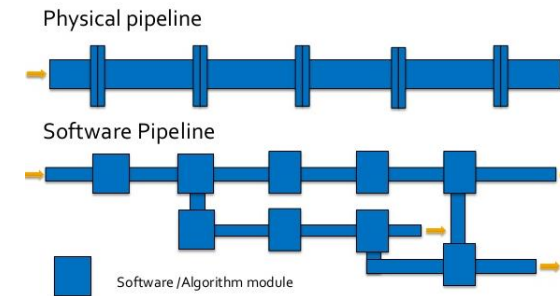
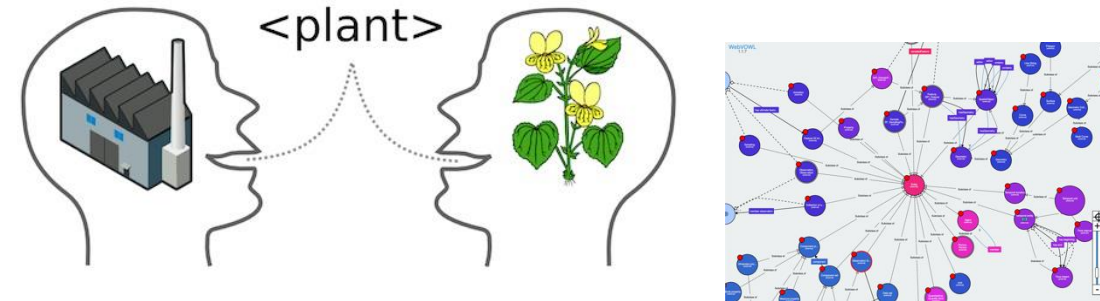
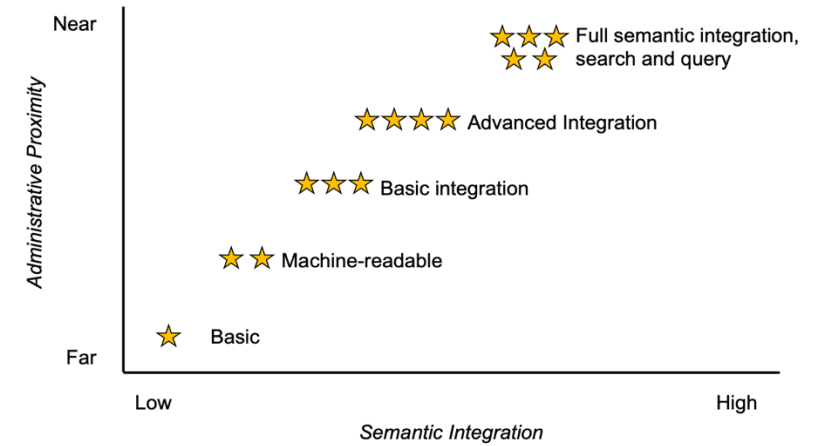
Co-funded by the European Union, Switzerland and the United Kingdom

- Aims to deliver the building blocks enabling semantic interoperability in the GDDS
 - Addressing the growing set of heterogeneous 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

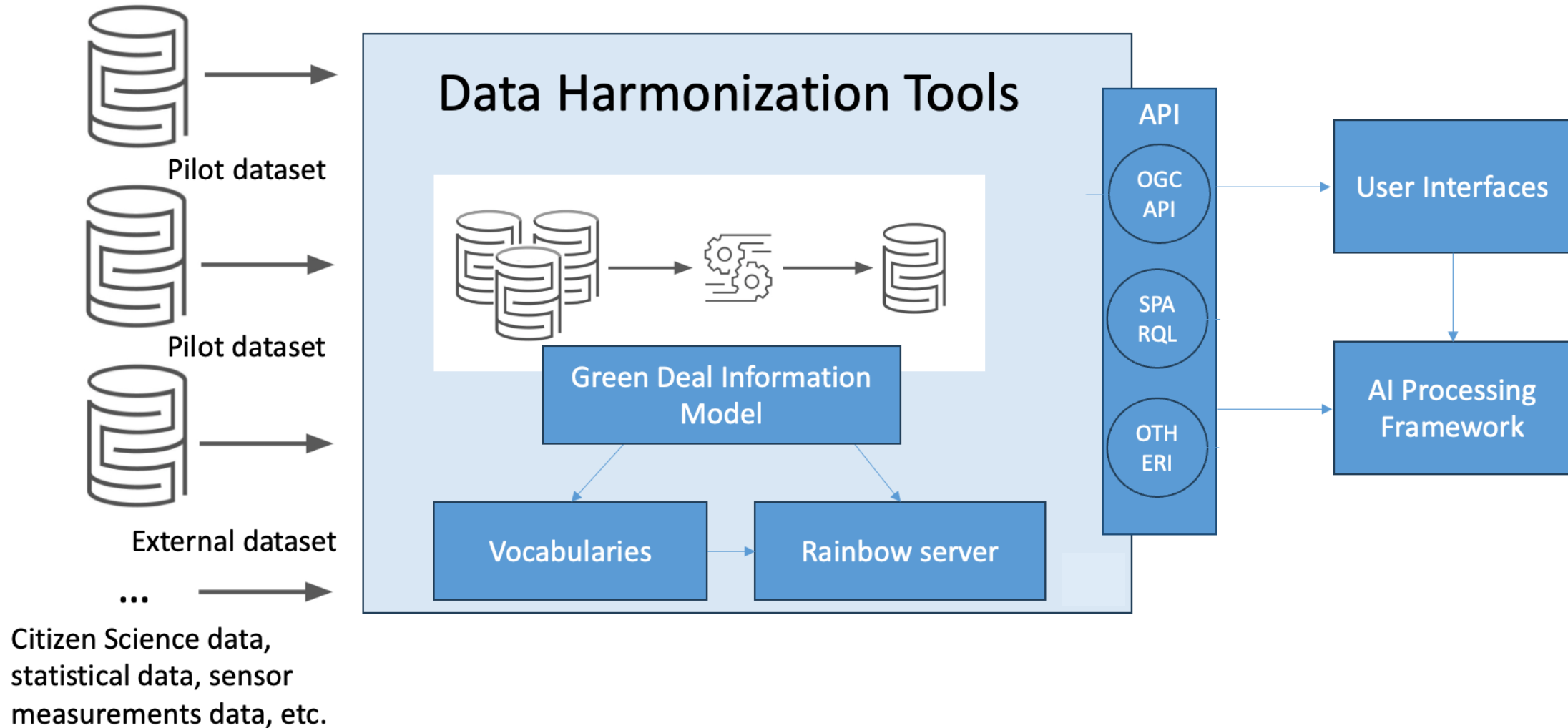


AD4GD approach

- 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

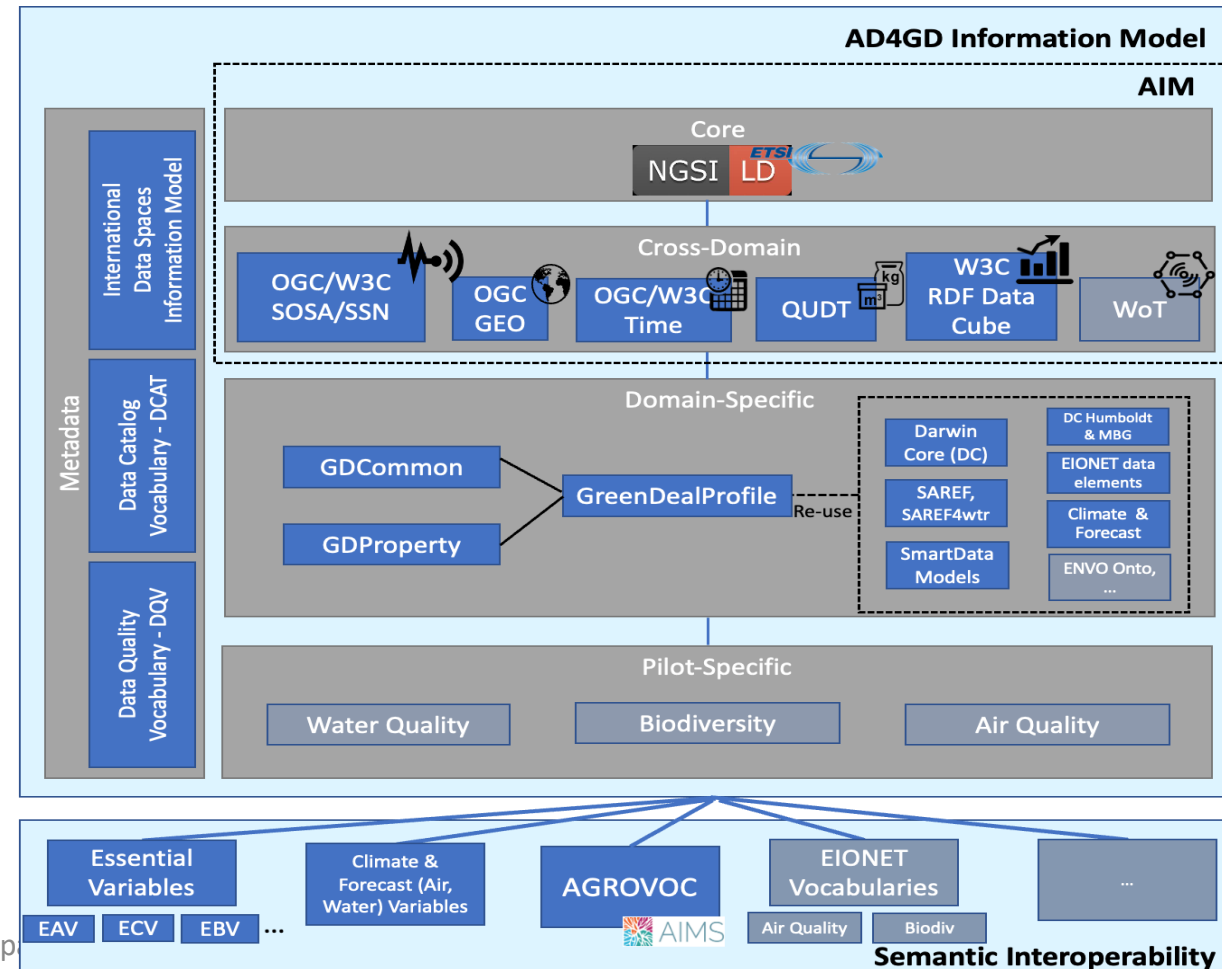


GDIM – Green Deal information model

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

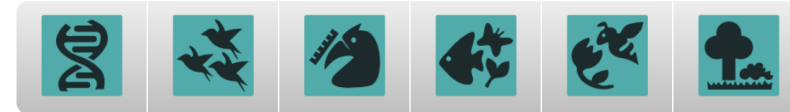
- **Essential Variables**
 - EAV, EBV, ECV, ...
 - ObservableProperty axioms
- **Climate & Forecast (variables)**
 - ontology representing the climatic data variables for use with climate and forecast data, in the atmosphere, surface and ocean domains.
 - ObservableProperty axioms
- **Agrovoc**
 - controlled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

	Definition	Agricultural Land Includes or is Indicated by	GEOGLAM Agriculture Indicator Category	Application or Policy Supported	Frequency of Update	Spatial Unit	EAV Stewards
Seasonal Dynamics of Surface Water Availability	Areas covered by inland water providing a maximum and minimum extent of water surface as well as the seasonal dynamics Binary map / wa Frequency (See CGLS water	All land (product is available over all land)	Productivity	Livestock, flood detection, SDG 6, The area of water bodies is identified as an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS)	Daily (ideal); 10 days (feasible)	30m	Sven Gilliams
Reference Evapotranspiration	ET _o from the reference surface. The reference surface is a hypothetical grass reference crop with an assumed crop height of 0.12 m, a fixed surface resistance of 70 s m ⁻¹ and an albedo of 0.23. (FAO).	Agriculture Mask	Land Use, Productivity	SDG 6.4, drought, climate adaptation, water & irrigation management	Daily - monthly, within each season	Field - Watershed	Benjamin Koetz, Bimal Bhattacharya, Jippe Hoogeveen
Reference Crop Calendars	Identification of the typical or usual planting and harvesting dates or windows per crop	For Global Analyses, the target is the same crops	Land Use, Productivity	Production Outlooks; Crop status anomaly detection, both spatial and	Every 5 years unless major events have	As stand-alone output.	Alyssa Whitcraft.

Full EAV Table

EBV classes and names

There are 6 EBV classes and 21 EBV names. By clicking on the icon you will get more detailed information.



SHOW ALL EBVs

Essential Climate Variables

[For graphical version click here](#)
[What are Essential Climate Variables \(ECVs\)?](#)

Atmosphere

- Surface**
 - [Precipitation](#)
 - [Pressure](#)
 - [Radiation budget](#)
 - [Temperature](#)
 - [Water vapour](#)
 - [Wind speed and direction](#)
- Upper-air**
 - [Earth radiation budget](#)
 - [Lightning](#)
 - [Temperature](#)
 - [Water vapor](#)
 - [Wind speed and direction](#)
 - [Clouds](#)
- Atmospheric Composition**

Land

- Hydrosphere**
 - [Groundwater](#)
 - [Lakes](#)
 - [River discharge](#)
 - [Terrestrial water storage](#)
- Cryosphere**
 - [Glaciers](#)
 - [Ice sheets and ice shelves](#)
 - [Permafrost](#)
 - [Snow](#)
- Biosphere**
 - [Above-ground biomass](#)
 - [Albedo](#)
 - [Evaporation from land](#)
 - [Fire](#)

Ocean

- Physical**
 - [Ocean surface heat flux](#)
 - [Sea ice](#)
 - [Sea level](#)
 - [Sea state](#)
 - [Sea surface currents](#)
 - [Sea surface salinity](#)
 - [Sea surface stress](#)
 - [Sea surface temperature](#)
 - [Subsurface currents](#)
 - [Subsurface salinity](#)
 - [Subsurface temperature](#)
- Biogeochemical**
 - [Inorganic carbon](#)
 - [Nitrous oxide](#)
 - [Nutrients](#)
 - [Ocean colour](#)

Essential Variables in OGC Rainbow Server

Vocabulary

Example codelist ebv

<http://w3id.org/ad4gd/ev/ebv>

Example variable

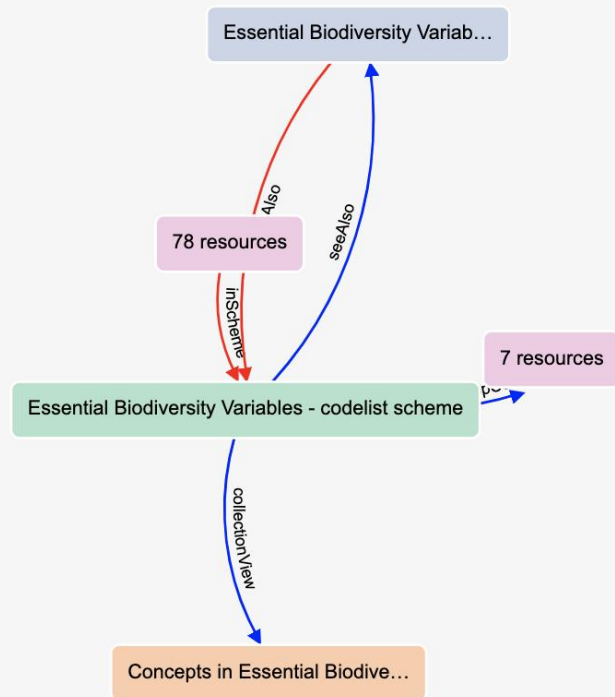
http://w3id.org/ad4gd/ev/ebv/Spp_SP_abn_bird_FW_Count

Preferred Label

Essential Biodiversity Variables - codelist scheme

URI

<http://w3id.org/ad4gd/ev/ebv>



Semantic Interop

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)

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 phytoplankton
 - 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

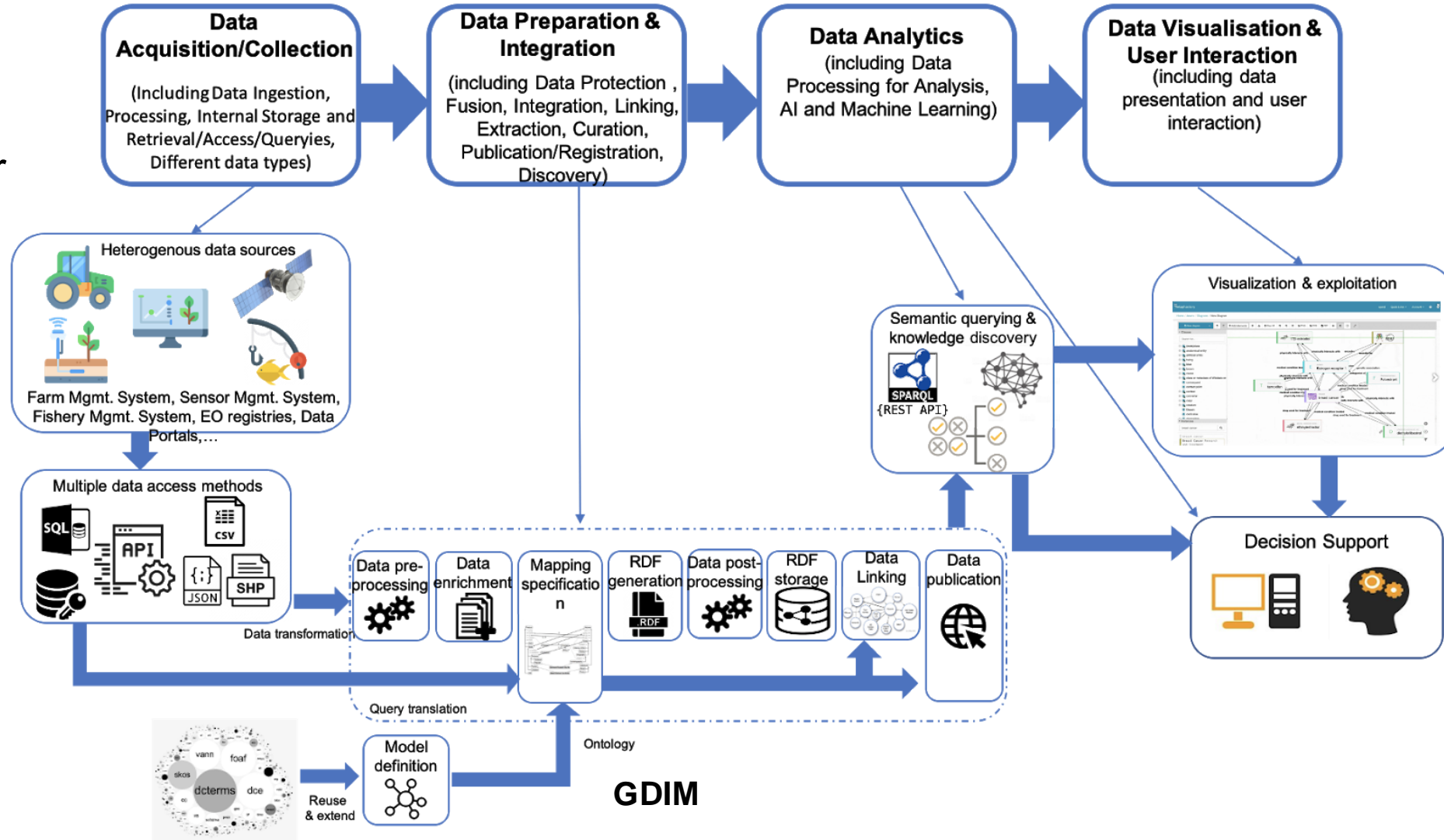
Essential Biodiversity Variables (EBVs) are defined as a minimum set of measurements, complementary to one another, that can capture major dimensions of biodiversity change. EBVs are organized in six classes (Genetic composition, Species populations, Species traits, Community composition, Ecosystem functioning,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

Search

Your
search term

Data harmonization approach

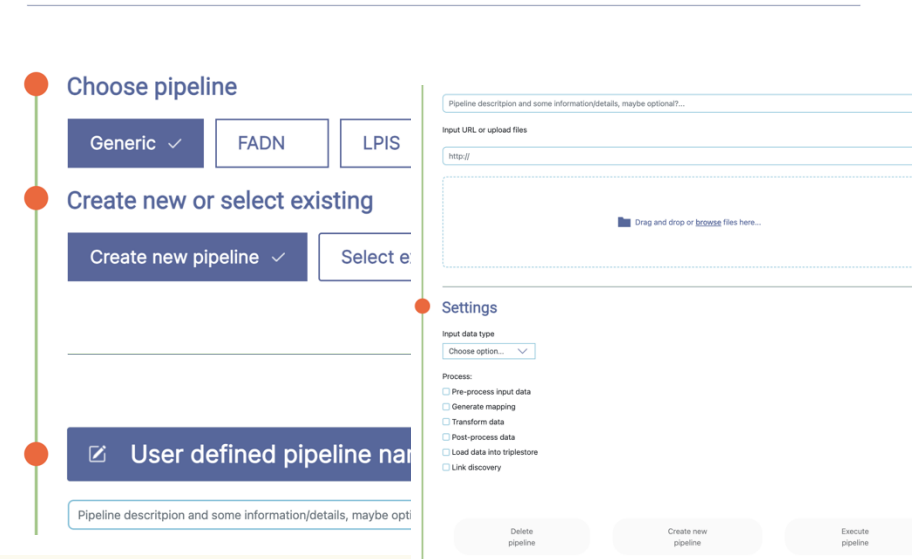
- Methods and tools for preparing, processing, and integrating incoming datasets
- Reuse, adapt and extend Data Preparation and Integration pipelines (DPI) tools based on linked data



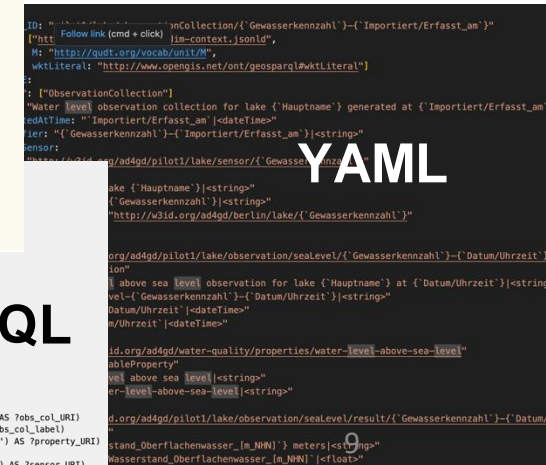
DPI in practice



- 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.pcass.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.psnk.pl/api/>
 - <https://dpi-enabler-dpi-enabler.apps.dcw1.paas.psnk.pl/api/swagger>
 - Web client application
 - <https://dpi-enabler-ui-test.apps.paas-dev.psnk.pl/>



```
mappings:
person:
s: http://example.org/${id}
po:
- [ex:name, John]
- [ex:firstName, John]
- [ex:city, Ghent]
```



```
FROM <file:jf-pilot-sample.csv>
WHERE {
  BIND (MDS(?obsID) AS ?identifier)
  BIND (COALESCE(xsd:date(?eventDate), ?eventDate) AS ?property_date)
  BIND (URI(CONCAT("https://w3id.org/iliad/jellyfish/collection/", ?identifier)) AS ?obs_col_URI)
  BIND (STR(LANG(CONCAT("Jelly fish observation collection #", ?obsID), "en")) AS ?obs_col_label)
  BIND (URI("https://w3id.org/iliad/jellyfish/properties/jellyfishAbundanceProperty") AS ?property_URI)
  BIND (MDS(?recordedBy) AS ?identifierSensor)
  BIND (URI(CONCAT("https://w3id.org/iliad/jellyfish/sensor/", ?identifierSensor)) AS ?sensor_URI)
  BIND (MDS(CONCAT(?obsID, "-", ?locationZone)) AS ?identifierFeature)
  BIND (URI(CONCAT("https://w3id.org/iliad/jellyfish/feature/", ?identifierFeature)) AS ?feature_URI)
  BIND (MDS(REPLACE(STR(?quantificationMethod), " ", "%20")) AS ?identifierMethod)
}
```

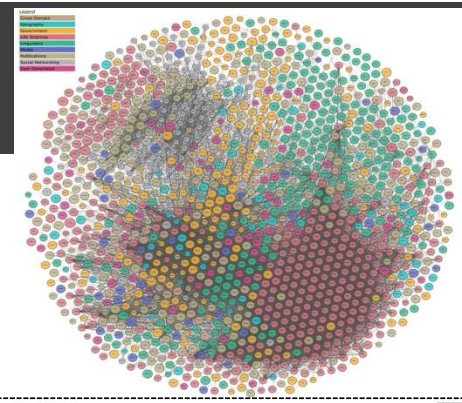
SPARQL

- 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

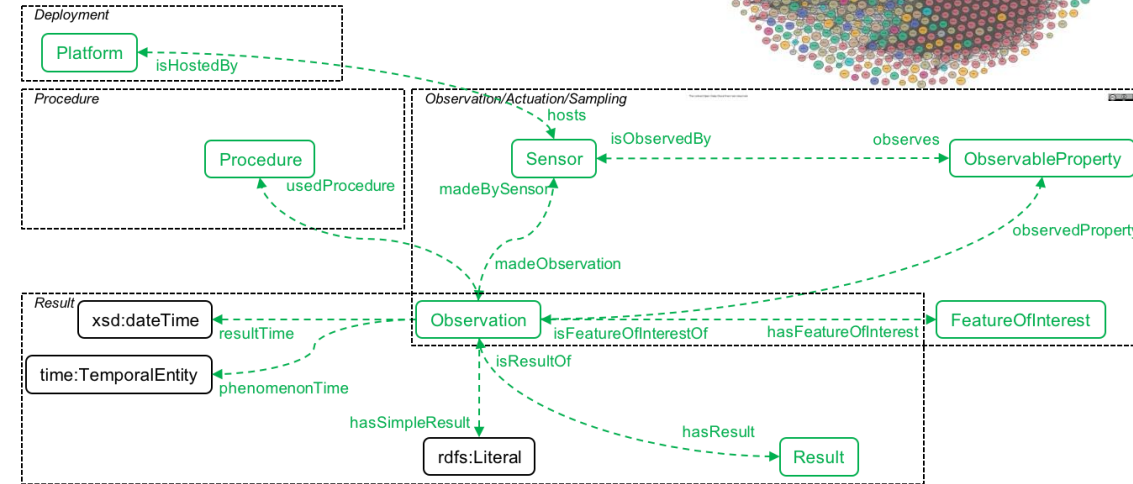
The screenshot shows the Open Geospatial Consortium website. The header includes the logo and navigation links: CONTEXT, APIS, SPRINTS, VIDEOS, BLOGS, DOCUMENTS, GET IN TOUCH. The main content area is titled "APIs for the Web" and displays a grid of API cards. Each card features a representative image, a title, a status (e.g., "Approved Standard"), a brief description, and buttons for "More info" and "GitHub repo".

API Name	Status	Description
Features	Approved Standard	OGC API - Features - Part 1: Core and Part 2: Coordinate Reference Systems by Reference are both publicly available.
Common	Common	OGC API - Common provides those elements shared by most or all of the OGC API standards to ensure consistency across the family. The candidate standard will soon be released for public review.
EDR	Approved Standard	Environmental Data Retrieval (EDR) API provides a family of lightweight interfaces to access Environmental Data resources. Each resource addressed by an EDR API maps to a defined query pattern.
Records		
Processes		
Coverages		

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



OGC SensorThings API generation service 1.4.10 OAS 3.1
/openapi.json

Pilots	
GET	/users Read Users
GET	/users/whoami Read Current User
GET	/users/{username} Read User
GET	/pilots Read Pilots
POST	/pilots Create Pilot
GET	/pilots/{name} Read Pilot

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

SPARQL | HTML5 table

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/sosa/ObservationCollection>

<http://www.w3.org/2000/01/rdf-schema#label> "Water level observation collection for lake Britzer Kirchteich generated at 2023-03-03T06:10:00"

<http://purl.org/dc/terms/identifier> **About:** [Water level above sea level observation for lake Britzer Kirchteich](#)

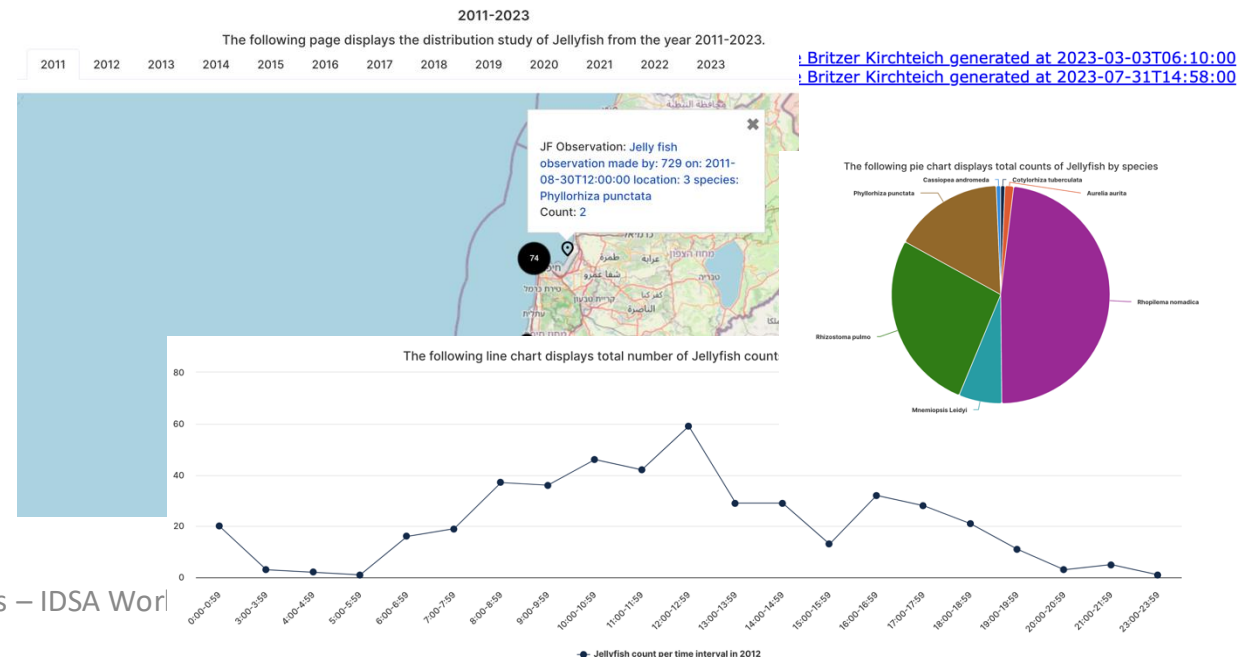
<http://www.w3.org/ns/sosa/hasFeature> [NotDistinct](#) [Permalink](#)

<http://www.w3.org/ns/sosa/hasMember> An Entity of Type : <http://www.w3.org/ns/sosa/Observation>, within Data Space : www.foodie-cloud.org, www.foodie-cloud.org.

Type: Command:

Attributes type	Values
label	Observation
Identifier	Water level above sea level observation for lake Britzer Kirchteich at 2023-01-30T04:04:00
has result	seaLevel-5832331.0-2023-01-30T04:04:00
observed property	38668.0 meters
	Water level above sea level

Jellyfish Distribution Study



Discussion and questions



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



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

FOLLOW OUR PROGRESS

*Subscribe to our
newsletter*



Visit our website on www.ad4gd.eu



Always as [@ad4gd_project](https://www.instagram.com/ad4gd_project)