31. August – 2. September 2020

Geospatial Sensing Virtual 2020

Enable true multidisciplinary applications in the emerging research e-infrastructures through geospatial standards

Joan Masó, CREAF

Geospatial Sensing
Fecha: September 1st, 2020
Virtual

Cos4Cloud GRUMETS CREAF

Table of contents

- Analysis Ready Data
- Emerging e-Infrastructures
- European Open Science Cloud
- The role of geospatial standards
- Need to support the whole scientific cycle
- EOSC and Citizen Science

From row data to Analysis Ready Data

- There is an increase in demand for data that can be directly analyzed. This is particularly true for EO remote sensed data.

- To respond to this demand CEOS has defined Analysis Ready Data for Land (CARD4L):

  - Products processed to a minimum set of requirements and organized into a form that allows **immediate analysis** with a minimum of additional user effort. These products would be resampled onto a common geometric grid (for a given product) and would provide baseline data for further interoperability both through time and with other datasets.

  - Products should follow a **Product Family Specification**
    - A specification of primary geophysical measurements that can be derived from satellite instruments as ARD. A PFS provides a list of requirements for a type of product to be considered ARD. These lists of requirements are applicable to general metadata, per pixel metadata, geometric and radiometric corrections.
      - Surface Reflectance, Surface Temperature, Normalized Radar Backscatter, Polarimetric Radar
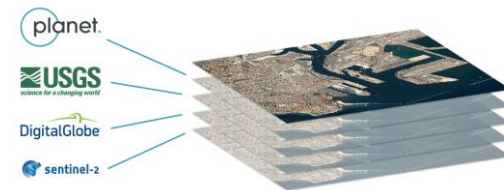
Can we generalize the concept of Analysis Ready Data to any sensor data?

- In principle **we could** for any sensor (remote or in-situ)
- Following the CEOS methodology, we need to know which Product Family Specification to apply (or define a **new one**)
- First:
  - We need a **primary geophysical measurement** *(a scientific variable)* that we can **agree on**
    - in ARD concept there is an implicit aim towards compatibility of measurements
  - Having a comprehensive **vocabulary** of primary geophysical measurements becomes essential
- Second:
  - We need to define a set of processes to apply to the data to **transform it** into ARD and document it in a clear way.
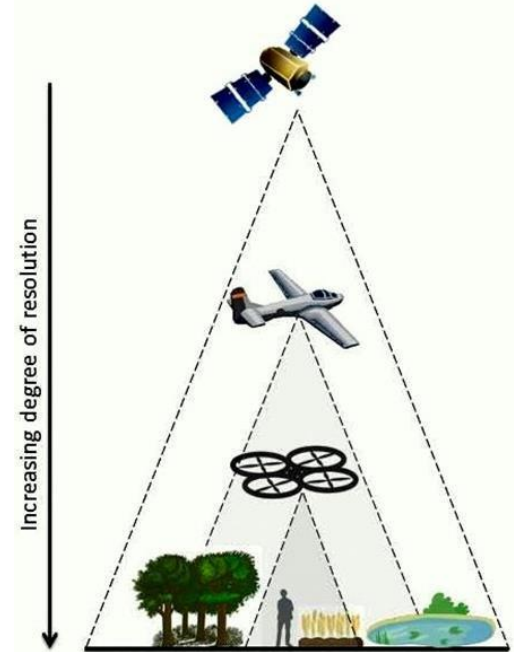
Currently we are working on this topic in the OGC Testbed 16

Increasing degree of resolution

# *Where to find a <u>primary geophysical measurement</u> vocabulary?*

- Everyone has heard of the International System of Units or SI Units.

- The International System of Quantities, or ISO 80000, is a standard that provides the foundations for understanding quantities.

- But this is not enough because the same units can be applied to different measurements
  - Example: temperature in Kelvin can be applied to:
    - Sea surface temperature
    - Atmospheric temperature

Table 1: Base Quantities of ISO 80000.

Base Quantity (ISO 80000)	Dimension Symbol	SI Unit
length	L	meter
mass	M	kilogram
time	T	second
electric current	I	ampere
thermodynamic temperature	Θ	kelvin
amount of substance	N	mole
luminous intensity	J	candela

Where to find a primary geophysical measurement vocabulary?

- There are some recent efforts:
  - **Scientific Variables Ontology**:
    - a mechanism for storing the conceptual information necessary for identifying, disambiguating, and assembling **scientific variables** (http://www.geoscienceontology.org/documentation/index.html)
  - **Cross-Domain Naming Conventions for Describing Process Models, Data Sets and Their Associated Variables**
    - (https://csdms.colorado.edu/wiki/CSN_Searchable_List-Names)

biomass

Where to put it?

CREAF
Cos4Cloud
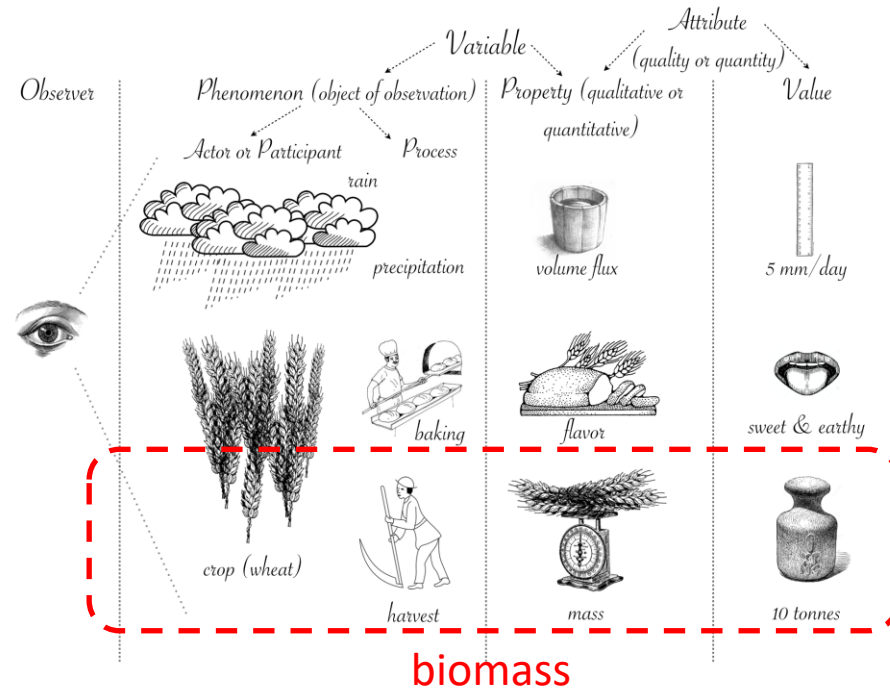
http://www.geoscienceontology.org/search.html

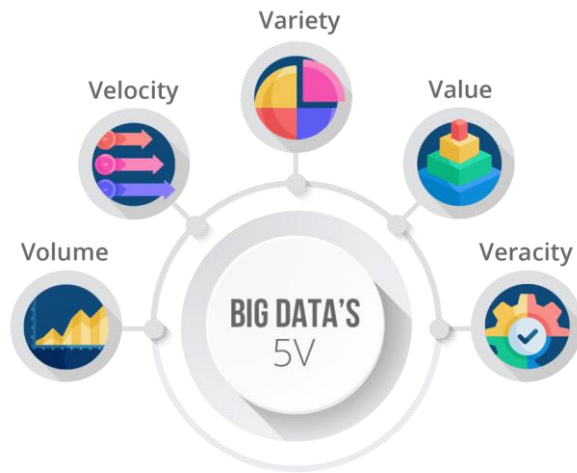<http://www.geoscienceontology.org/svo/svl/assumption#atmosphere-and-radiation_standard_temperature>

"atmosphere-and-radiation_standard_temperature"

"atmosphere-and-radiation_standard_temperature"

## Common Model – Observations and Measurements (O&M) SOS Web Service

Sensor Things API Model
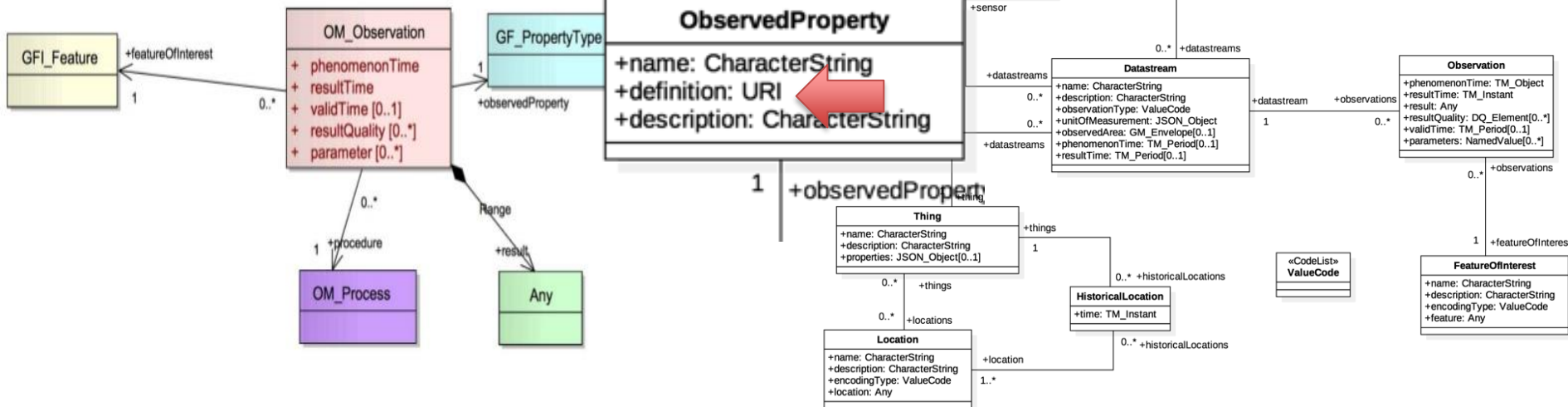
Why am I taking about ARD and variables?

- ARD and variables vocabularies sets some rules and **some order** in the big data Variety

- Still time series analysis of **long time series** (Volume) may require too much computer efforts to process

- The emerging e-Infrastructures take advantage of the Analysis Ready Data: we have available to respond questions by **analyzing** the data (processing)
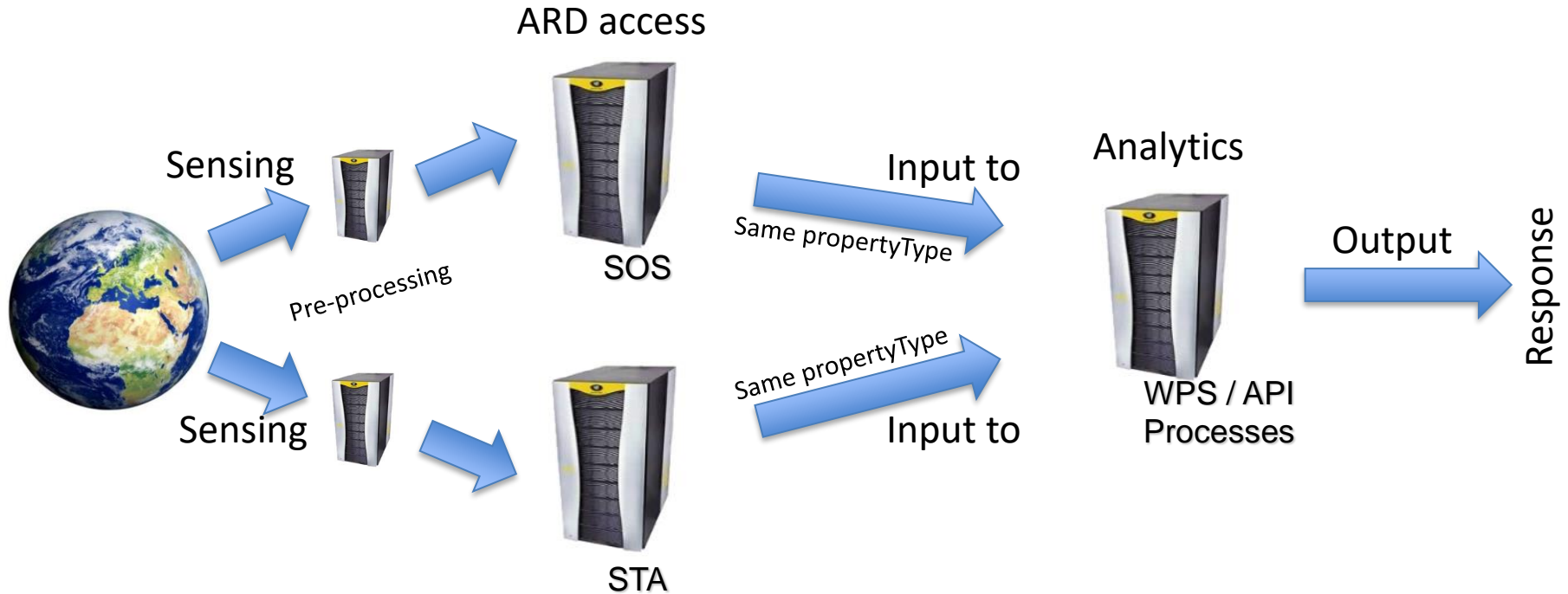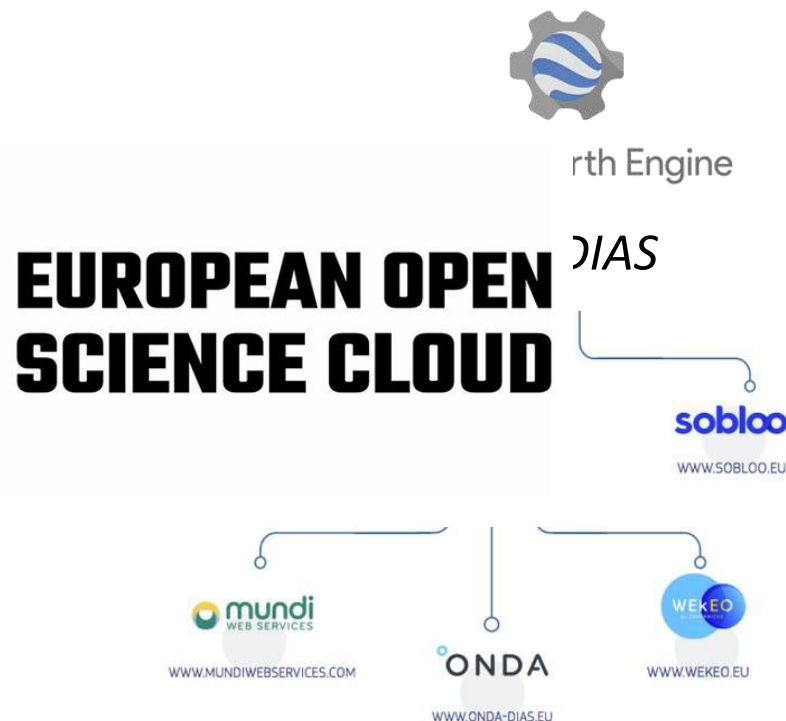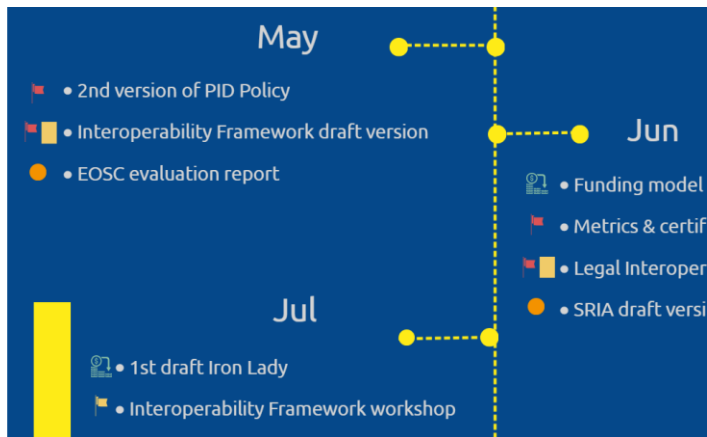
What is the EOSC

- It aims to be **a safe environment** for researchers to store, analyse and re-use data **for research, innovation and educational** purposes
- This universal entry point has the potential to help 1.7 million researchers and 70 million professionals in **science, technology, humanities and social sciences** to access a growing volume of open data
- It was launched in 2019 and it is in its **infancy** in terms of services available but it is growing fast
- The rules to participate are not completely defined

EUROPEAN OPEN SCIENCE CLOUD

BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES

A trusted, open environment for sharing scientific data

Open and seamless services to analyse and reuse research data

Linking data

Connecting across borders and scientific disciplines

Connecting scientists globally

Long term and sustainable

Improving science
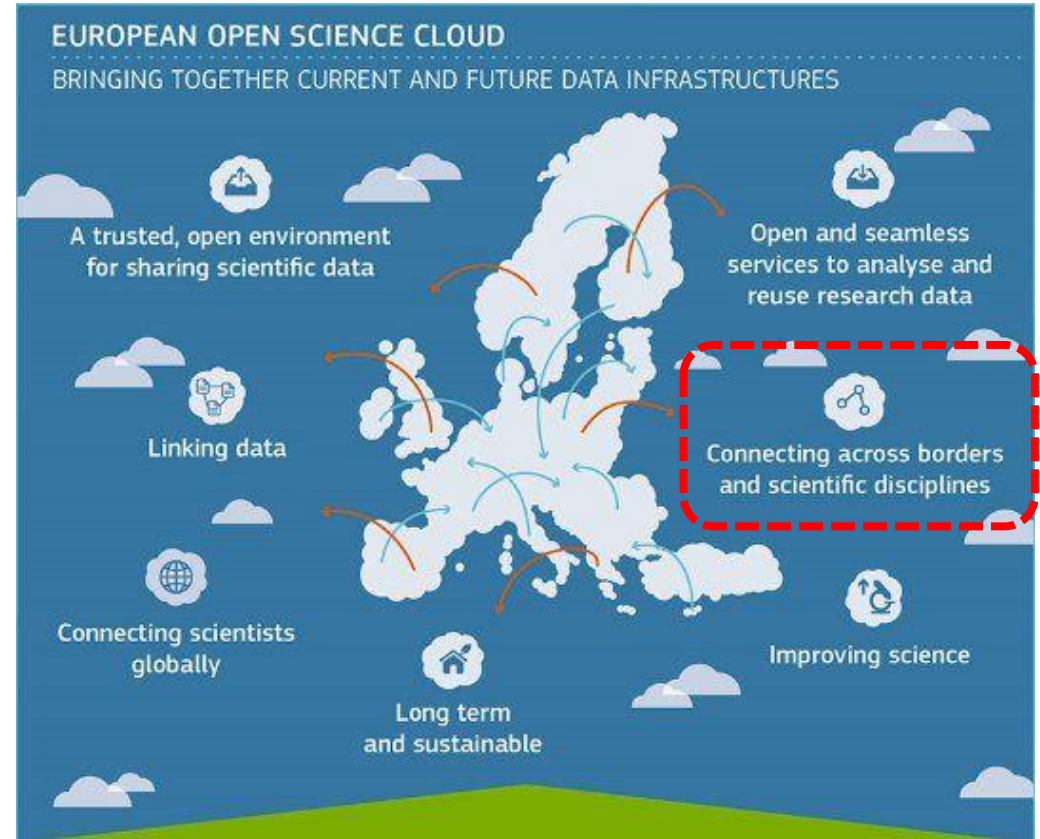
- EOSC is **considering some standards** to build the infrastructure
- But I'm not able to find **almost any reference to geospatial standards** in the EOSC development documents
- Considering that a big number of disciplines in EOSC are related to Earth science, **one would expect location to be important**

EOSC-hub

D10.4 EOSC Hub Technical Architecture and standards roadmap v2

Lead Partner:	INFN		
Version:			
Status:			

REST API's	A full REST API is used to collect metadata formatted as JSON, e.g. the referenced REST API is used to 'harvest' from Herbadrop's repository	https://helpdesk.eudat.eu/Ticket/Attachment/122586/63597/RESTAPI_HowTo_SearchUserGuide_V3.pdf
CSW / OGC	Catalogue Service for the Web (CSW) is used to collect metadata from OpenGeoSpatial Catalogues (OGC)	http://www.opengeospatial.org/standards/cat

May
- 2nd version of PID Policy
- Interoperability Framework draft version
- EOSC evaluation report

Jun
- Funding model
- Metrics & certif
- Legal Interoper
- SRIA draft versi

Jul
- 1st draft Iron Lady
- Interoperability Framework workshop

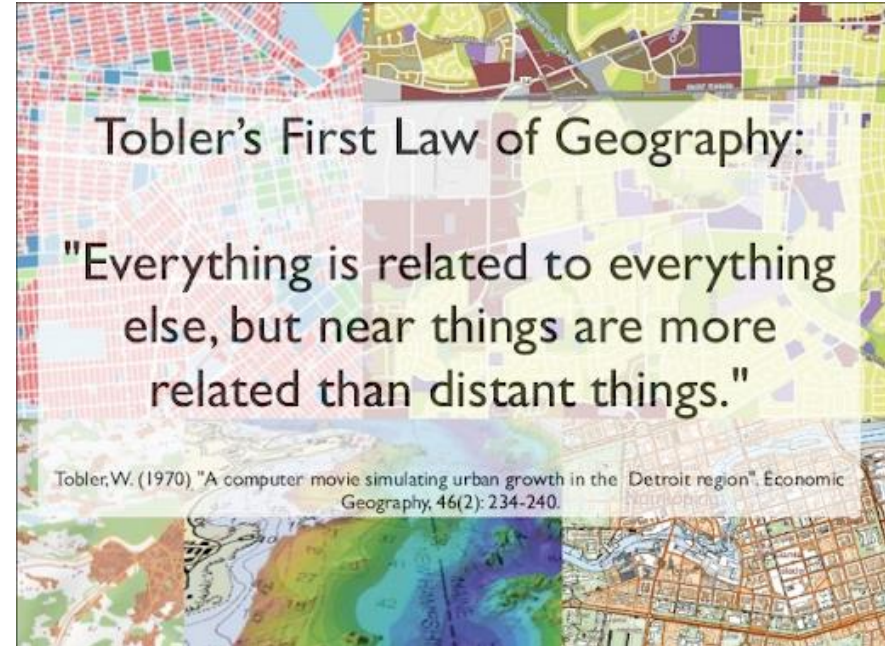https://www.eoscsecretariat.eu/eosc-working-groups

- In a position paper European Commission emphasizes the potential of the FAIR data in EOSC for ***cross-fertilization of multidisciplinary research***
  - *Relevance of the EOSC initiative and FAIR principles in the realm of Open Science and implementation phases of the EOSC* https://ec.europa.eu/research/openscience/pdf/eosc-fair_paper_schouppe-burgelman_2018.pdf

- BUT

- The **EOSC hub** defined thematic research infrastructures as **separated** Competence Centres: "*Each Competence Centre fosters the use of advanced digital capabilities and services of the EOSC hub [...] Each Competence Centre operates **independently from each other** but shares the needs in using common solutions from the EOSC hub service catalogue to setup **community-specific** services that can expand EOSC with science **discipline specific capabilities***".
  - EOSC hub D8.1 Report on progress, achievements and plans of the Competence Centres https://documents.egi.eu/document/3485

- **EOSC** provides an splendid **opportunity for** our **geospatial** sensing data, *in the form of ARD,* to be made available in an **standard** way (SOS or STA) in an infrastructure that is capable to **process** it (WPS or OGC API Processes)
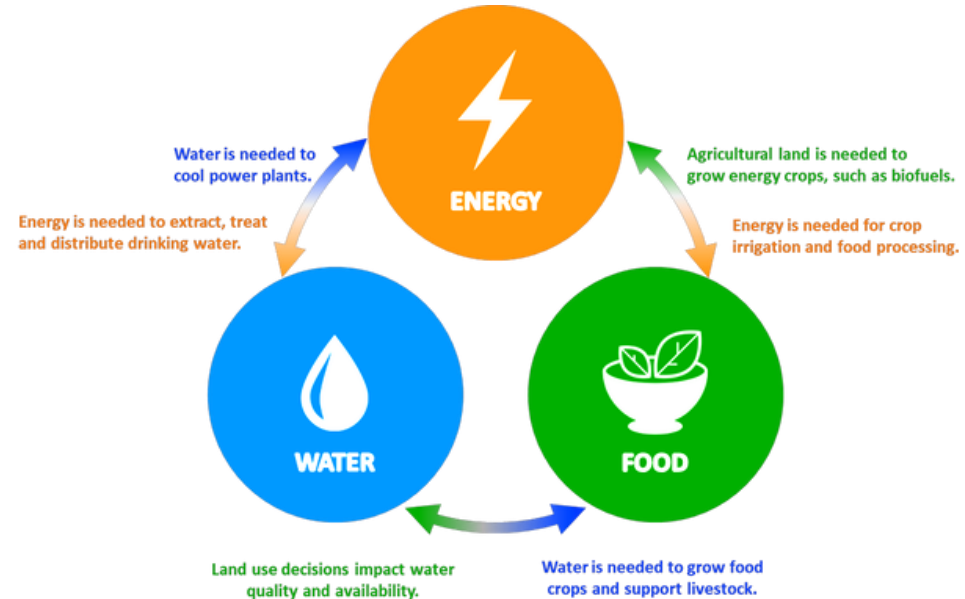
- And perform **multidisciplinary research**

- So our data (one variable) can be combined with any other data (another variable) and enable multidisciplinary research

- Why can be combine them?
  - Because they happen in the same place.

Tobler's First Law of Geography:

"Everything is related to everything else, but near things are more related than distant things."

Tobler, W. (1970) "A computer movie simulating urban growth in the Detroit region". Economic Geography, 46(2): 234-240.

- **The Food-Water-Energy nexus**
  - Combine data from **crop biomass** with **water quantity** and quality with **energy demand** and production from renewables (biofuel or hydropower)
  - We can study
    - impacts on hydropower production on food availability (irrigated crops),
    - potential land for biofuel production
    - …
- This requires **models**
  - created by **multidisciplinary teams** that **breaks** the competence center **silos**
 - run combining **open data** in research infrastructures

Water is needed to cool power plants.

Agricultural land is needed to grow energy crops, such as biofuels.

Energy is needed to extract, treat and distribute drinking water.

Energy is needed for crop irrigation and food processing.

ENERGY

WATER

FOOD

Land use decisions impact water quality and availability.

Water is needed to grow food crops and support livestock.

- Are geospatial standards well position to cover the whole scientific cycle?

Geospatial standards in the research cycle
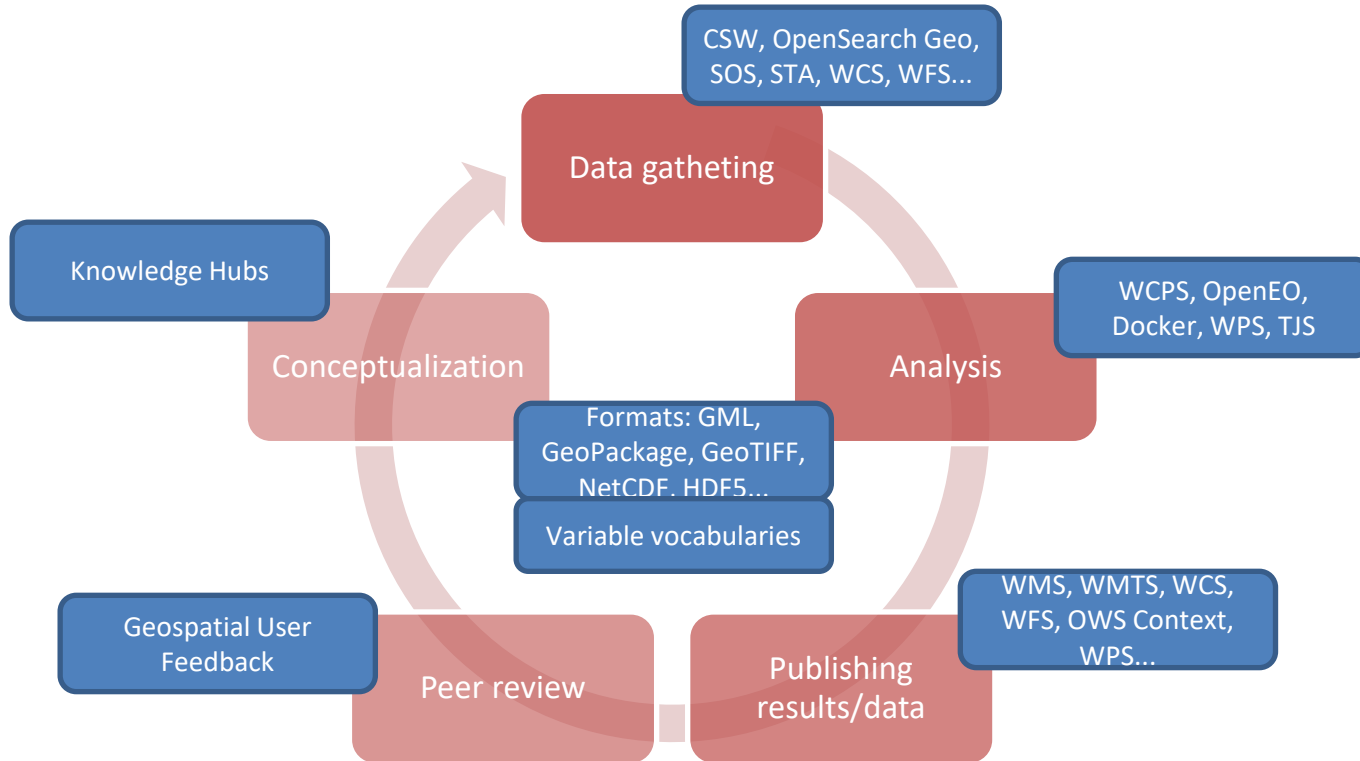
- Cos4Cloud evolving 10 services to become **EOSC services**, so that **any existing citizen science** observatory in EOSC will be able to benefit from them to improve its functionalities.

- Cos4Cloud relies on the participation of a network of 9 **citizen observatories** focused on biodiversity and the environment domain. These platforms will use and test the different services with their users.

- Experts portal to **validate observations** from multiple citizen observatories

- Data use **notification**

- **Mobile** interface for apps

- MECODA data **analysis** package

- Interactive pre-processing **camera trap**

- **Video stream** processing service

- PlantNet as a service

- **Similarity-search** based identification service

- Location-based **species prediction**

- **Training data** aggregation service

Citizen observatories and DIY initiatives

Biodiversity

- iSpot
  - Over 68,000 global nature observers identifying of 30,000 taxa, of more than 750,000 observations.

- Natusfera
  - Share biodiversity learning with the community by reporting observations on all sorts of living beings.

- Pl@ntNet
  - Collecting, sharing and reviewing plant observations based on automated identification.

- Artportalen
  - Biodiversity observations in Sweden: taxa, location, date and reporter, observation method…

Environmental

- CanAirIO
  - Colombian air quality with mobile and fixed sensors (Particle Material PM2.5) with low-cost technology and open source code.

- OdourCollect
  - Empower citizens to tackle odour pollution.

- FreshWater Watch
  - Global citizen-scientist platform, investigating the health of the world's freshwater ecosystems.

- Kduino
  - DIY application tu use Arduino with low cost sensors for to measure **water transparency**.

- iSPex
  - Aerosols and water colour based on a mobile app and a small optical add-on containing a spectrometer and a polarizer.

Massive interoperability experiment

Citizen Observatories

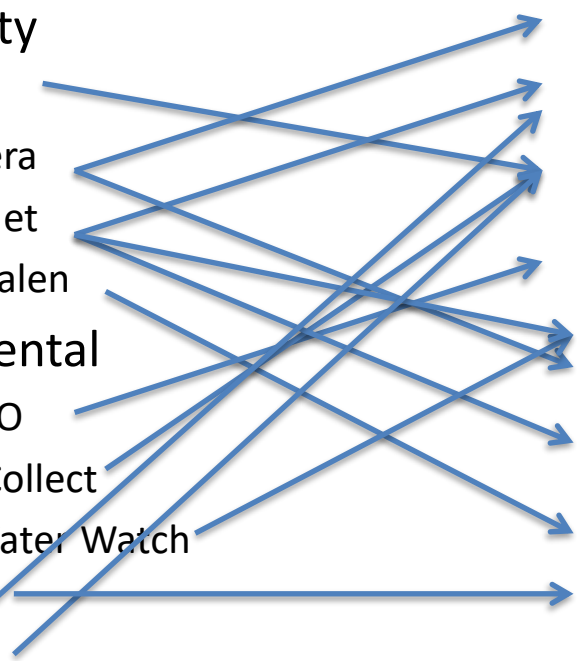- Biodiversity
  - iSpot
  - Natusfera
  - Pl@ntNet
  - Artportalen

- Environmental
  - CanAirIO
  - OdourCollect
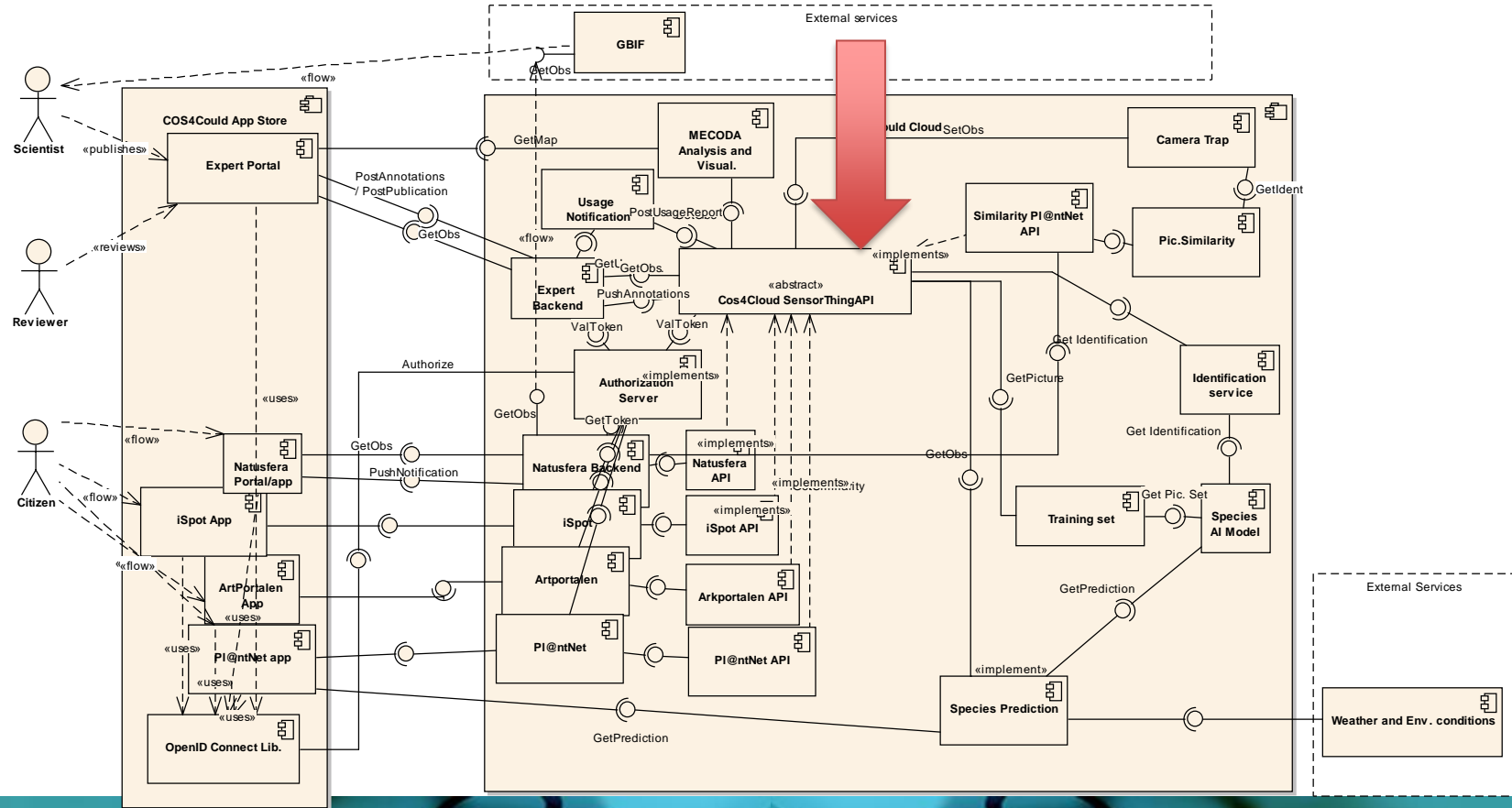 - FreshWater Watch
  - Kduino
  - iSPex

Services

- Experts portal to **validate observations** from multiple citizen observatories
- Data use **notification**
- **Mobile** interface for apps
- MECODA data **analysis** package
- Interactive pre-processing **camera trap**
- **Video stream** processing service
- PlantNet as a service
- **Similarity-search** based identification service
- Location-based **species prediction**
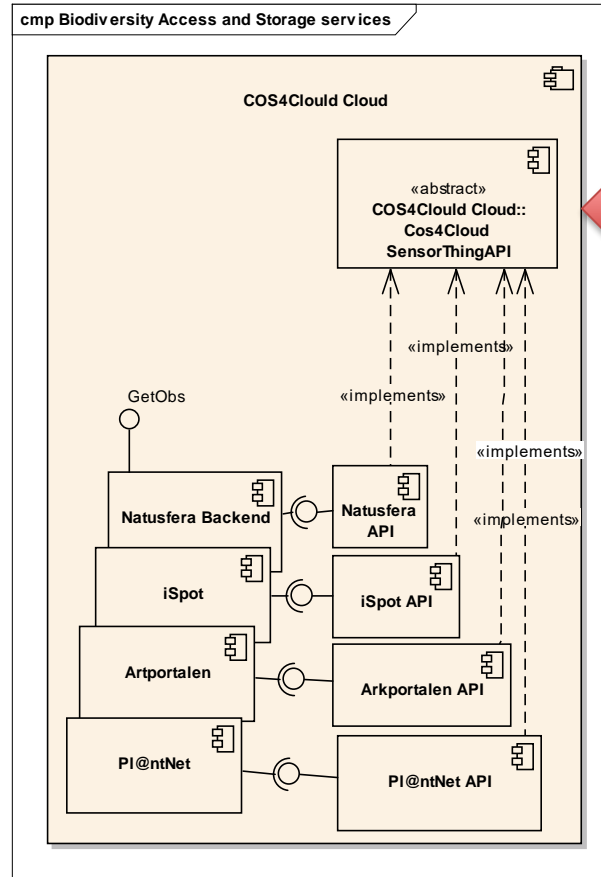- **Training data** aggregation service

Who to make it work?

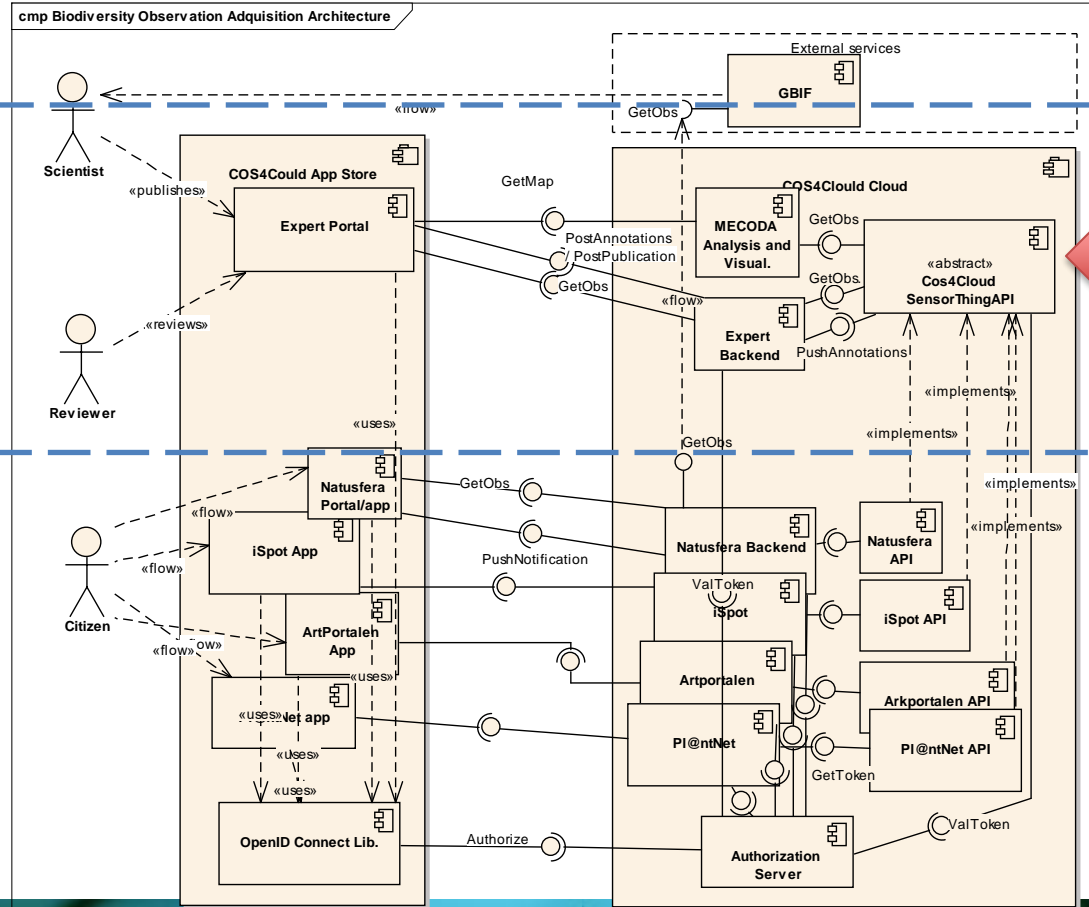# The role of Sensor Things API

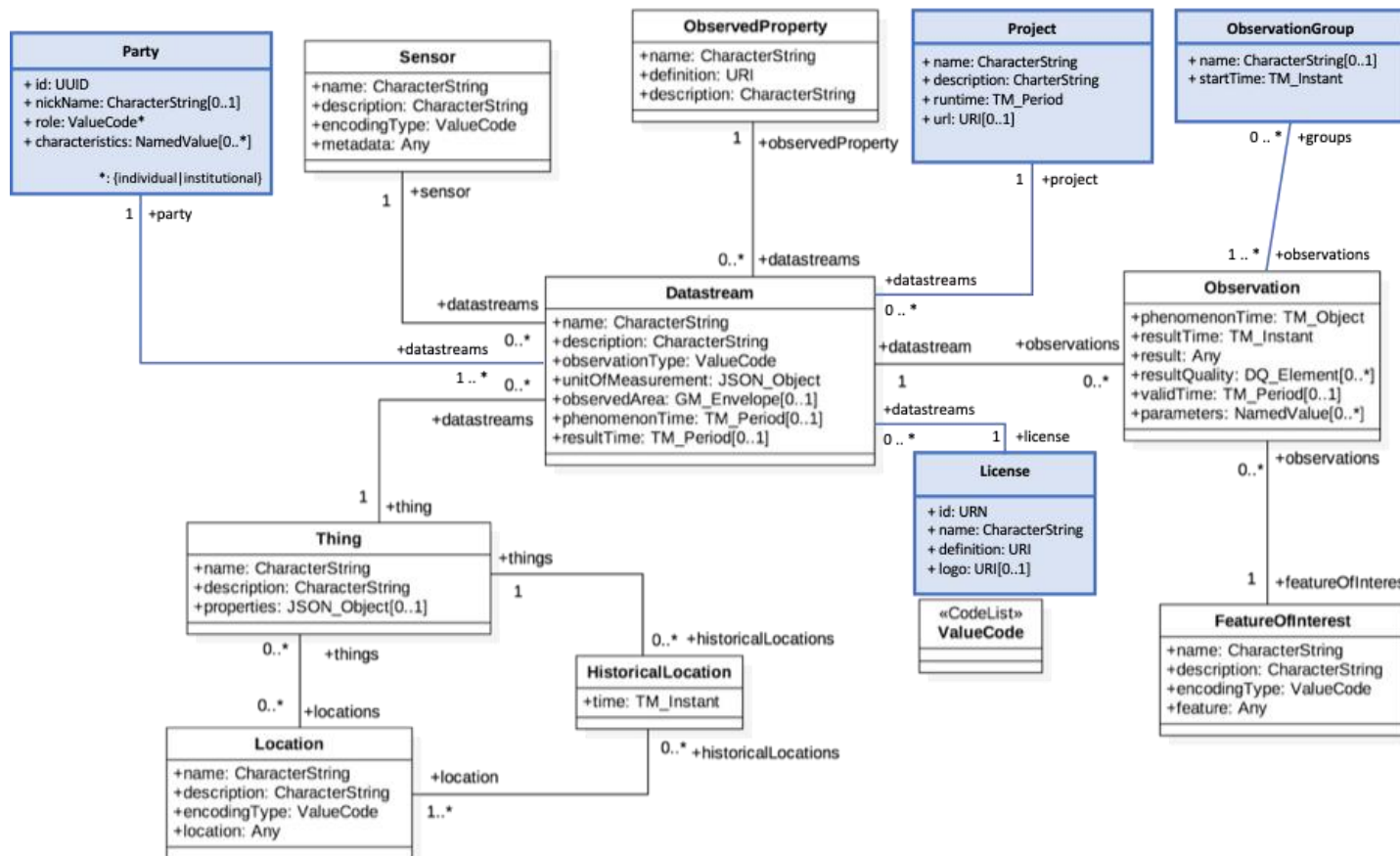# The role of Sensor Things API

Extending STA to consider the citizen

Target: services will be added to the EOSC portal

marketplace.eosc-portal.eu/services/c/sharing-discovery?service_id=&anchor=&sort=_score&q=citizen

All Services	4

CATEGORIES

Compute	0
Data management	1
Networking	0
Processing & Analysis	1
Security & Operations	1
Sharing & Discovery	**3**
Storage	0
Training & Support	0

FILTERS

Research Area ⌃

Find or choose from the list below

☐ Natural Sciences	2
☐ Biological sciences	2
☐ Computer sciences	1
☐ Other natural sciences	1
☐ Chemical sciences	0
☐ Earth sciences	0

Looking for: **citizen**

Found 3 results   Sort by: Best match

Lagunas de Sierra Nevada / Glacier Lagoons of Sierra Nevada

Science and conservation of glacier lagoons

Provided by: University of Granada – UGR
Research area: Biological sciences
Dedicated for: Providers, Research group, Research organisations, Researchers

☐ Add to comparison

Plant Classification

Plant-classification-deep-learning

Provided by: IFCA-CSIC, LifeWatch ERIC
Research area: Biological sciences, Computer sciences, Environmental engineering, Other natural sciences
Dedicated for: Research group, Research organisations, Researchers

☐ Add to comparison

EUROPEAN OPEN SCIENCE CLOUD

Conclusion

- **Analysis Ready Data** is the **fuel** for the **e-infrastructures**

- **Geospatial standards** should enable **multidisciplinary** scientific **research** by connecting **sensed** data (observations) with knowledge about variables interactions stored in **models** (processes)

- We should reevaluate our geospatial standards to better support the whole **scientific research cycle** beyond data discovery and access.

  – More on **peer review** and **conceptualization** is needed.

One complain

- This presentation assumes that sensor data can be ingested into a process (a model).
- This is done because georeference and well known **variable vocabularies**
- Sensor data has a ObservedProperty or PropertyType
- Unfortunately WPS (or OGC API Process) **does not have it** as part of the description of inputs.
  - There is a UOM but this is not enough.
  - Acceptable for generic processing tools
  - Not good for describing specialized models
- Do we need to fix that?
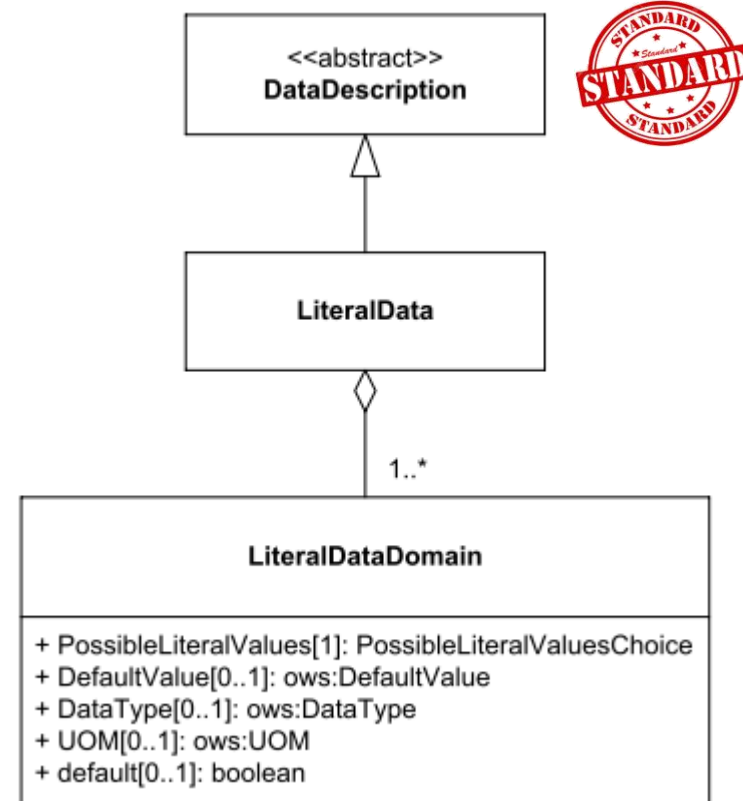
Thanks

joan.maso@uab.cat

Cos4Cloud: Co-designed
Citizen Observatories Services for the
European Open Science Cloud

https://cos4cloud-eosc.eu/