groundtruth2.0

Environmental knowledge discovery of human sensed data

Quality Indicators of Citizens Observatories using Sensor web standards. The Ground Truth 2.0 case

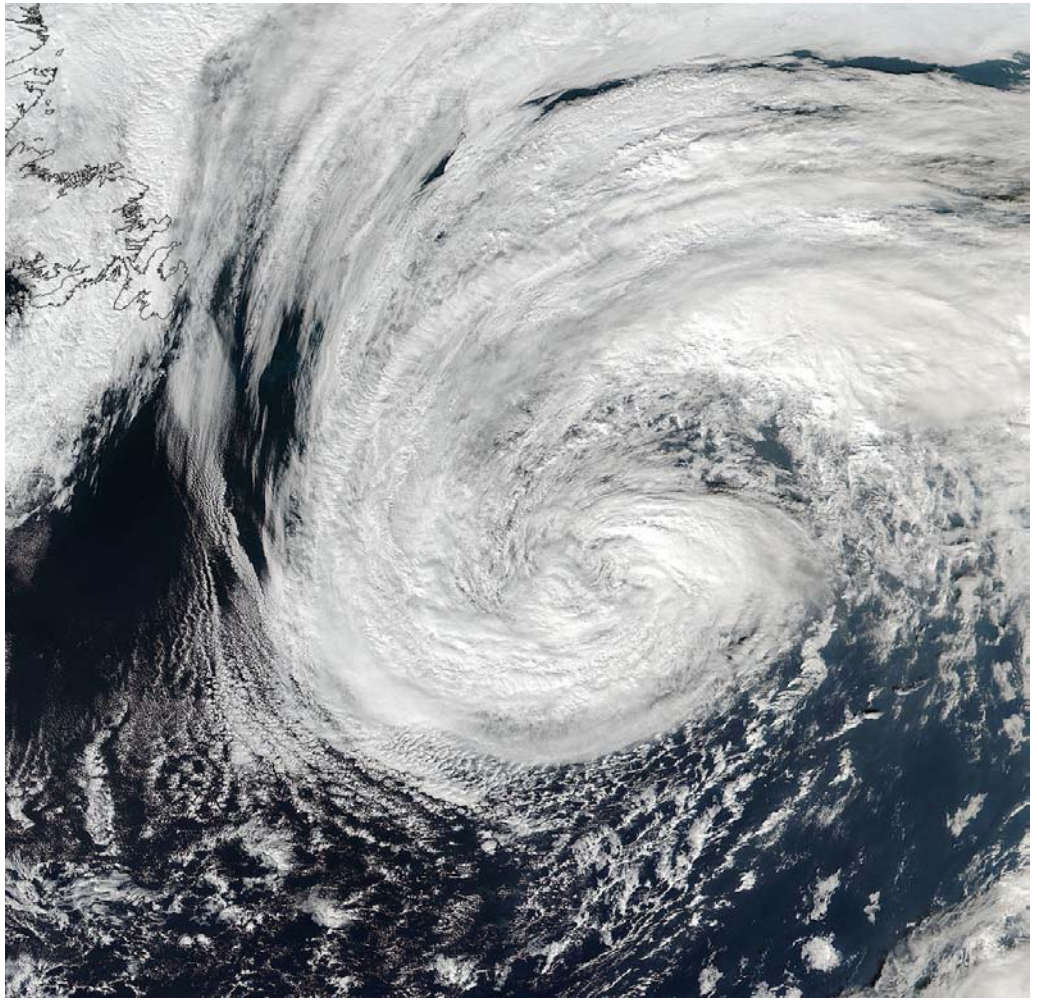
Joan Masó (CREAF)

2018-09-04

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 689744.
This presentation reflects only the author's view and the EU Agency is not responsible for any use that may be made of the information it contains.

Context: Groups for Earth Observation

GEO: an intergovernmental body that promotes open Earth Observation data availability, access and use



Source: Steve Ramage

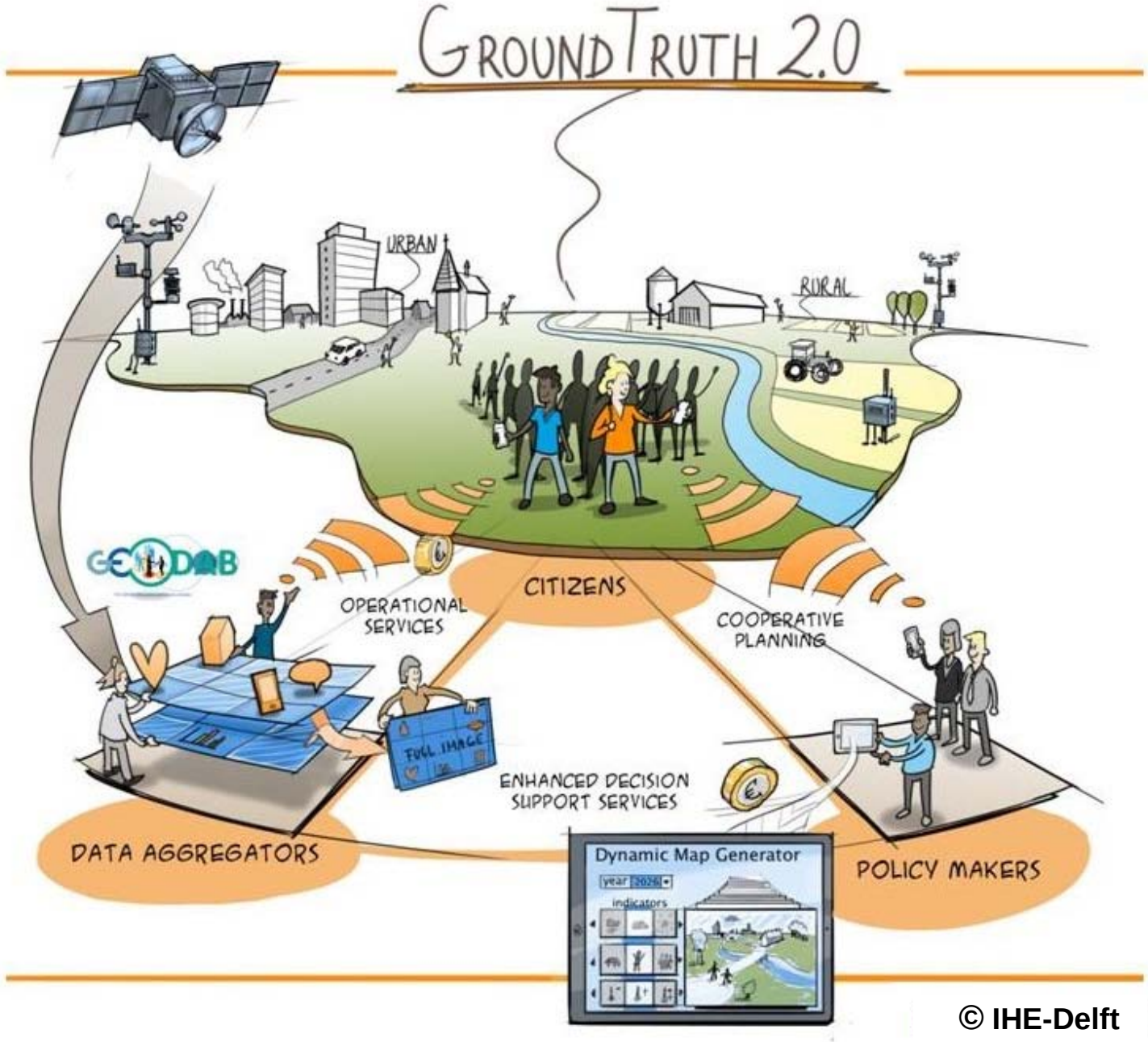
groundtruth2.0

Citizen Science and Citizen Observatories

- **Citizen Science is** the collection and analysis of data relating to the natural world by members of the general public, typically as part of a collaborative project with professional scientists.

- **Citizens' Observatories are** community-based environmental monitoring and information systems. They build on innovative and novel Earth observation applications embedded in portable or mobile personal devices. This means that citizens can help and be engaged in observing our environment.

groundtruth2.0

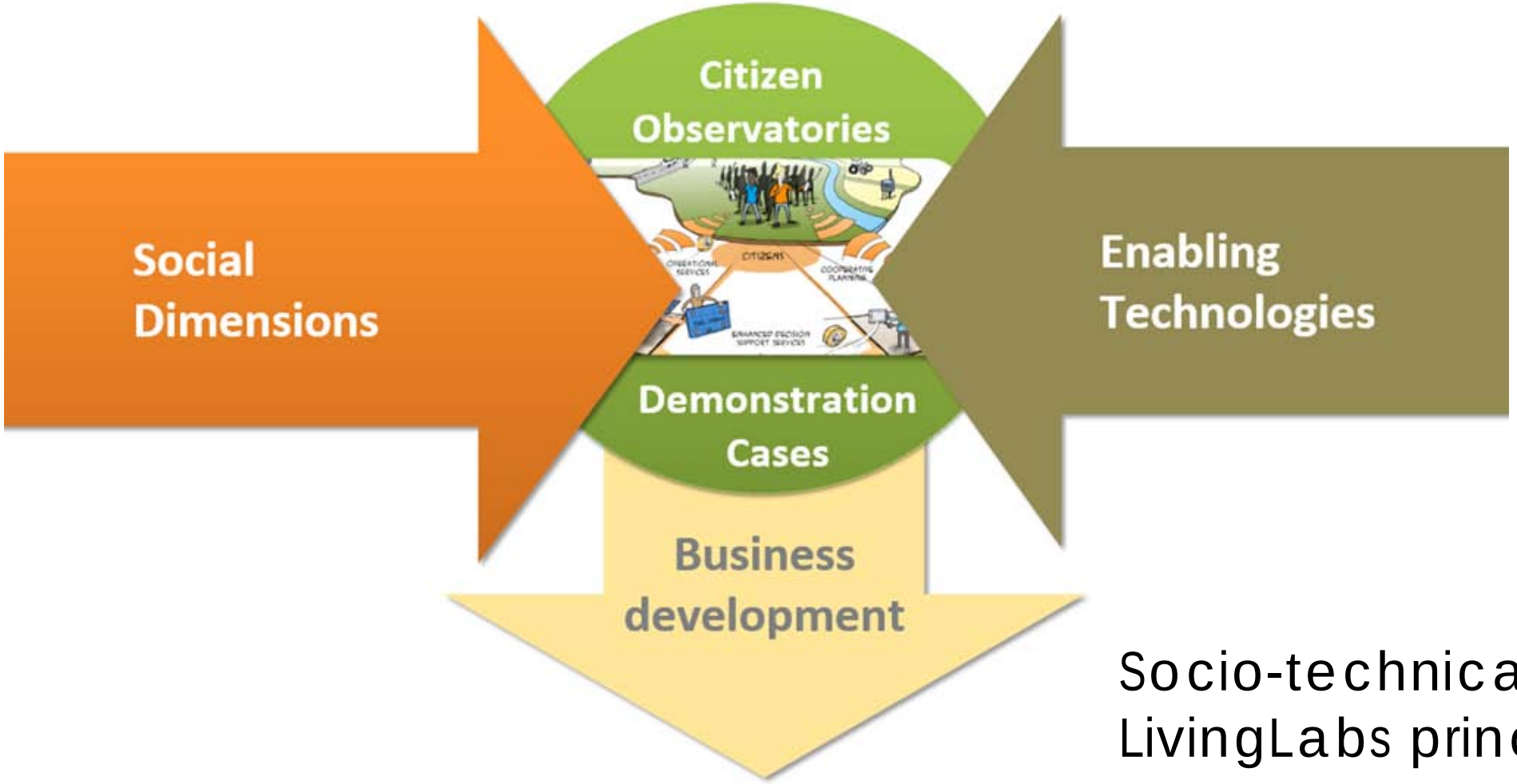


One of the EC funded Citizen Observatories that goes...
... from citizen-based data collection to knowledge sharing for **joint decision-making**, **cooperative planning** and **environmental stewardship**

Source: Uta Wehn

groundtruth2.0

groundtruth2.0 methodology



Socio-technical approach
LivingLabs principles

Source: Uta Wehn

The diversity of CS topics

- There are many CS projects.
- Example
  - Conrad CC, Hilchey KG (2011) A review of citizen science and community-based environmental monitoring: issues and opportunities
    - Identifies and analyzes 20 different CBEM projects
  - Bowser A (2016) Global Mosquito Alert: A Common Protocol and Platform for Citizen Science Vector Monitoring
    - Identifies 7 mosquito monitoring active projects.
    - Proposes the creation of the Global Mosquito alert
  - Earth challenge 2020
    - 1 billion observations made by one million people

MOSQUITO ALERT

The four COs active in H2020 for GEOSS

groundtruth2.0
Environmental knowledge discovery of human sensed data

- From citizen-based data collection to knowledge sharing for joint decision-making, cooperative planning and environmental stewardship
- Considers the social and th
- Six demo cases

scent

- Crowdsourcing platform used to collect images and text from citizens
- Calibration and validation of satellite imagery using the crowdsourcing
- h serious gaming and machine
- os river, Greece (urban) and hia (rural)

WeObserve

- Inventory of citizen observatories and sustainability
- Accelerate the uptake of Citizen Observatories
- Facilitate adoption in EO communities

GROW OBSERVATORY

- Sustainable custodianship
- Meet the demands of food production
- Calibrate/validate satellite-based soil moisture products with CS in-situ sensors
- Campaign-based approach for engaging

Landsense

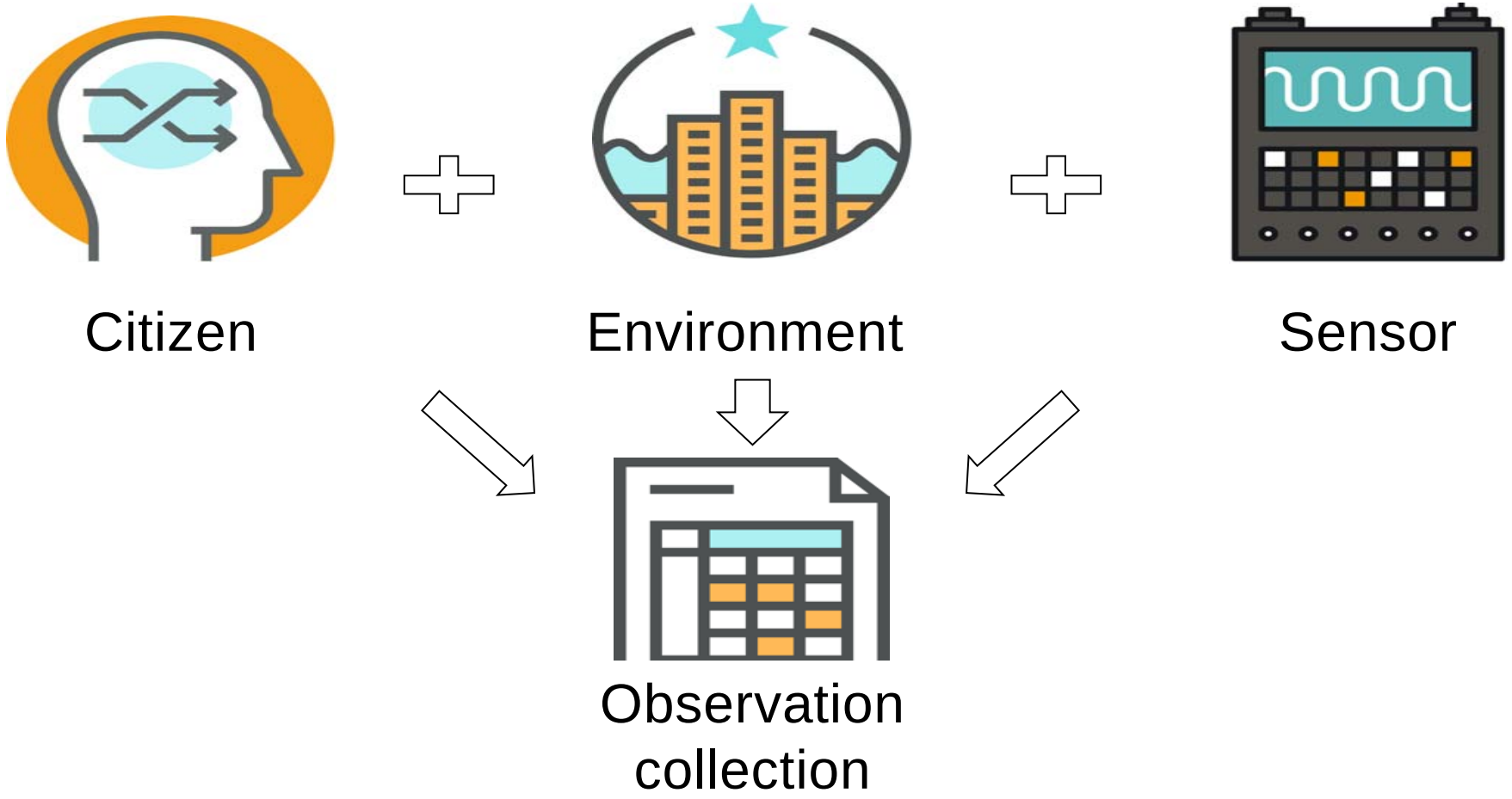
- t platform
- Emphasis on services
  - Photoquest (pictures in 4 directions), change detection in alerting about bird habitat risks, single sign-on
- Three demo cases on landscape changes, agriculture and habitat monitoring

The integration of CS

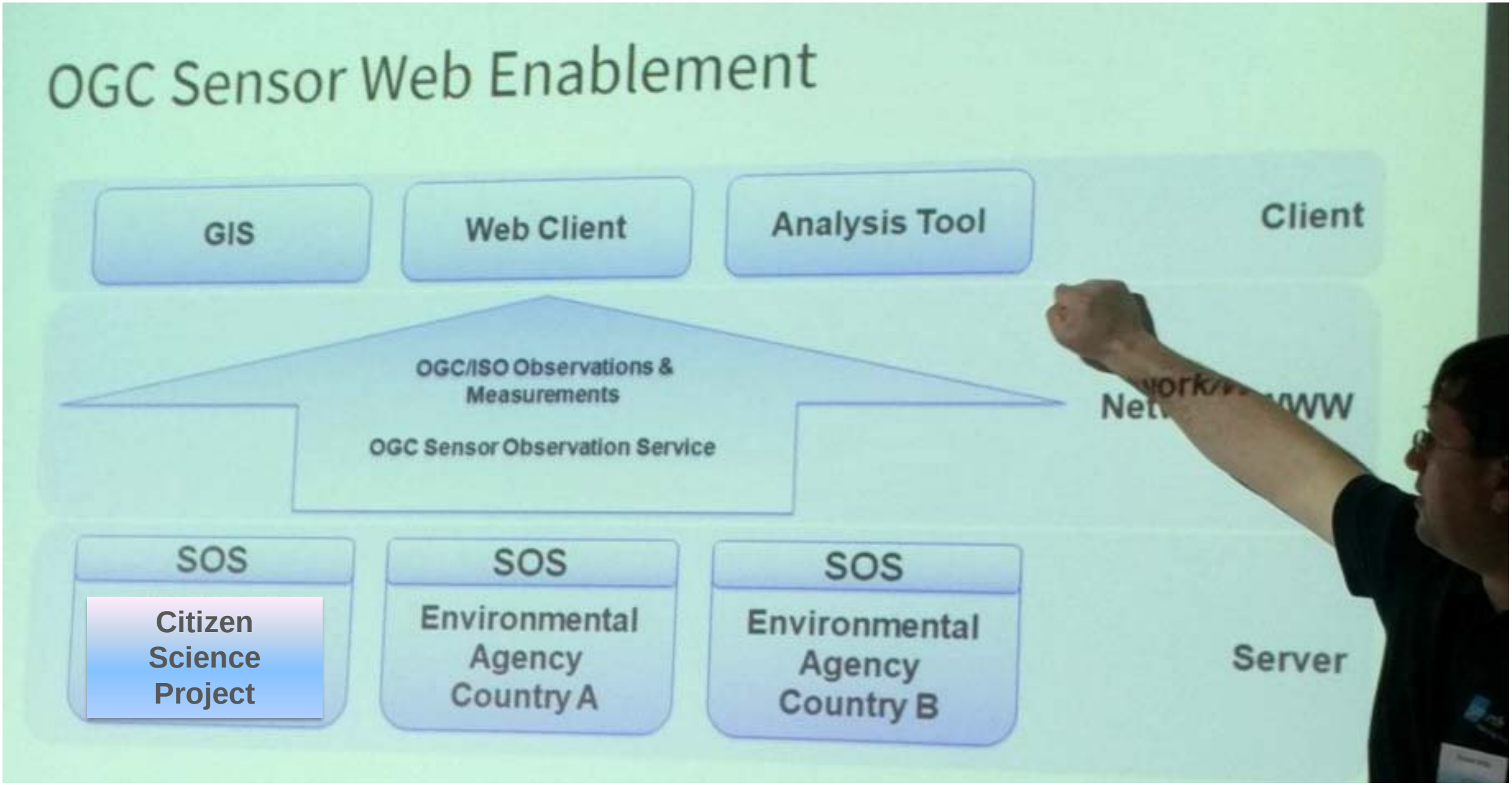
- At the project description level
  - Project documentation via Public Participation in Scientific Research metadata model (and the COST action efforts)
- At the service level
  - SWE4CS (presented in GSW 2017)
  - IoT
- At the data level
  - Conflation of datasets using semantic mediation

groundtruth2.0

Standardization



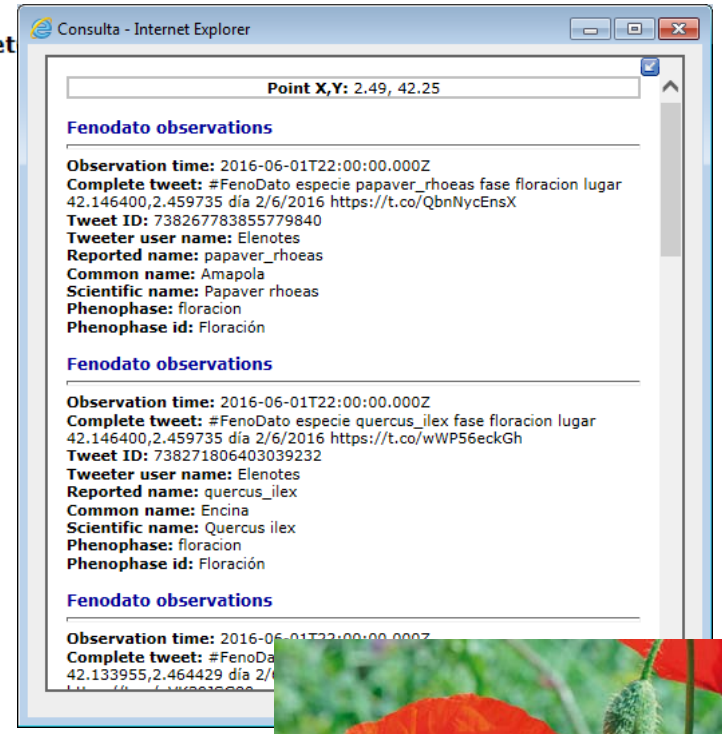
Source: Ingo Simonis


OGC Sensor Web Enablement
GIS
Web Client
Analysis Tool
Client
OGC/ISO Observations & Measurements
OGC Sensor Observation Service
SOS
Citizen Science Project
SOS
Environmental Agency Country A
SOS
Environmental Agency Country B
Server

Phenology observations in SOS

- http://localhost/cgi-bin/miramon.cgi?VERSION=2.0.0&SERVICE=SOS&REQUEST=GetObservation&featureOfInterest=http://www.opengis.uab.cat/fenodato/featureOfInterest/173

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<sos:GetObservationResponse xsi:schemaLocation="http://www.opengis.net/sos/2.0 http://schemas.opengis.net/sos/2.0/sosGet
 - <sos:observationData>
    - <om:OM_Observation gml:id="fenodato_173_5">
          <om:type xlink:href="http://www.opengis.net/def/observationType/OGC-OM/2.0/OM_TextObservation"/>
        - <om:phenomenonTime>
            - <gml:TimeInstant gml:id="phenomenonTime_fenodato_173_5">
                  <gml:timePosition>2016-06-01T22:00:00.000Z</gml:timePosition>
              </gml:TimeInstant>
          </om:phenomenonTime>
          <om:resultTime xlink:href="#phenomenonTime_fenodato_173_5"/>
          <om:procedure xlink:href="http://www.opengis.uab.cat/fenodato/procedure"/>
          <om:observedProperty xlink:href="http://www.opengis.uab.cat/fenodato/observableProperty/K_NAME"/>
          <om:featureOfInterest xlink:href="http://www.opengis.uab.cat/fenodato/featureOfInterest/173"/>
          <om:result xsi:type="xs:string" xmlns:xs="http://www.w3.org/2001/XMLSchema">Amapola</om:result>
      </om:OM_Observation>
  </sos:observationData>                                                                    ma">papaver_rhoeas</or
 - <sos:observationData>
    - <om:OM_Observation gml:id="fenodato_173_8">
          <om:type xlink:href="http://www.opengis.net/def/observationType/OGC-OM/2.0/OM_TextObservation"/>
        - <om:phenomenonTime>
            - <gml:TimeInstant gml:id="phenomenonTime_fenodato_173_8">
                  <gml:timePosition>2016-06-01T22:00:00.000Z</gml:timePosition>
              </gml:TimeInstant>
          </om:phenomenonTime>
          <om:resultTime xlink:href="#phenomenonTime_fenodato_173_8"/>
          <om:procedure xlink:href="http://www.opengis.uab.cat/fenodato/procedure"/>
          <om:observedProperty xlink:href="http://www.opengis.uab.cat/fenodato/observableProperty/P_NAME"/>
          <om:featureOfInterest xlink:href="http://www.opengis.uab.cat/fenodato/featureOfInterest/173"/>
          <om:result xsi:type="xs:string" xmlns:xs="http://www.w3.org/2001/XMLSchema">Floración</om:result>
      </om:OM_Observation>
  </sos:observationData>
```



Data quality

Includes:

- understanding of the **concepts** of data quality related to geographic data.
- defining data quality **conformance** levels in data product specifications or based on user **requirements**.
- specifying quality aspects in **application schemas**.
- **evaluating data quality.**
- **reporting data quality.**

groundtruth2.0

Data quality

Evaluation

- Apply an assessment method to a set of observations (or observation parameters) to quantify the *uncertainties* present
  - Selecting a measurement
  - From a list of values (domain)
  - Doing metrics (statistics)

Documentation

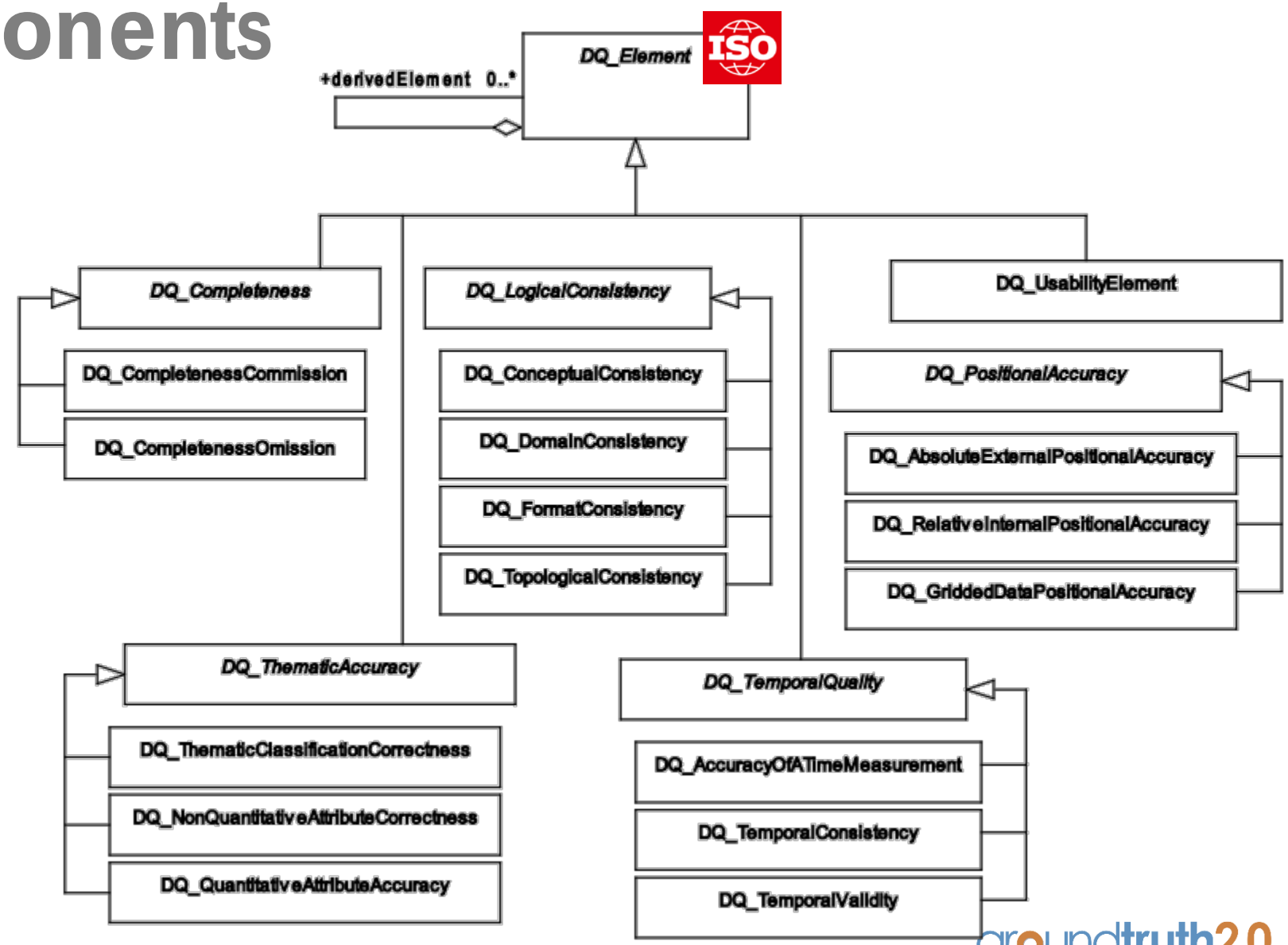
- Presenting the results in an understandable and comparable way
- In the metadata describing the set of observations

```
<gmd:DQ_QuantitativeResult>
   <gmd:value>
     <gco:Record>
        <qml:HalfLengthConfidenceInterval>
           <un:values/>11<un:values>
        </qml:HalfLengthConfidenceInterval>
     </gco:Record>
   </gmd:value>
</gmd:DQ_QuantitativeResult>
```

groundtruth2.0

Quality components

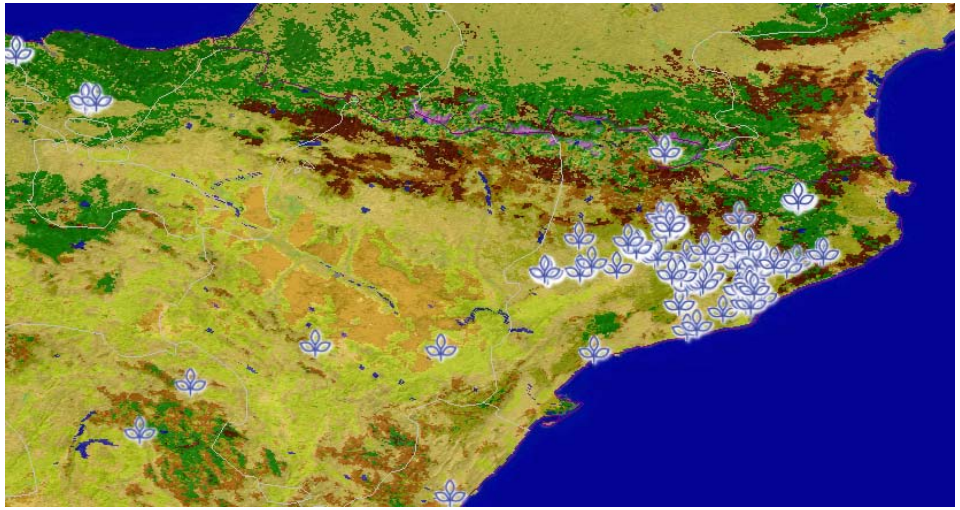
- Evaluation consists on a list of quality *elements* each one describing **a certain aspect** of the *quality* of geographic data *organized* into this **categories**:



groundtruth2.0

Data quality

- Levels
  - Observation level
    - "Positional accuracy of the observation id139 is 30m"
  - At the attribute in an observation level
    - "Species name in the observation id147 is not known"
  - Dataset (set of observations) level
    - "10% of the observations does not present a mandatory attribute."

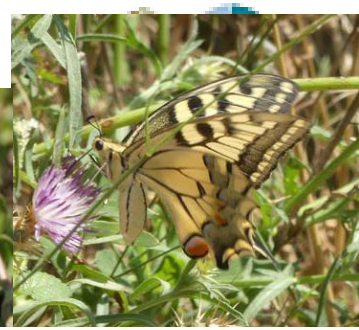


groundtruth2.0

Data quality at the *observation level*

- Each observation (or some attributes of it) has (have) an estimation of its quality by:
  - Intrinsic characteristic of the sensor (nominal or recently calibrated
    - "Temperature sensor uncertainty: ±0.2 degrees"
  - Current circumstances that limits accuracy
    - "No GPS available. Position estimated by triangulation of 3G: uncertainty:150m"
  - Estimated by repetition
    - "The uncertainty of repeated measurements of the elevation is 5m"
  - Expert reviewed
    - "Species name validated in a picture by 2 reviewers"

[Papilio machaon](#)



How to document (encode) it:

- It can be provided in several ways:
  - As an additional attribute
  - As a data quality element at the observation level
  - As a data quality value that accompanies a attribute

groundtruth2.0

Data quality as an additional attribute

```
<om:OM_Observation gml:id="ritme-natura_209_62">
  <om:procedure xlink:href="http://www.opengis.uab.cat/ritme-natura/procedure/pau_guzman"/>
  <om:featureOfInterest xlink:href="http://www.opengis.uab.cat/ritme-natura/featureOfInterest/209"/>
  <om:result xsi:type="swe:DataRecordPropertyType">
    <swe:DataRecord>
      <swe:field name="num_identification_agreements“><swe:Quantity definition="http://www.opengis.uab.cat/ritme
          <swe:value>3</swe:value>
      </swe:Quantity></swe:field>
      <swe:field name="positional_accuracy"><swe:Quantity definition="http://www.opengis.uab.cat/ritme-natura/va
          <swe:value>11</swe:value>
      </swe:Quantity></swe:field>
      <swe:field name="positioning_method“><swe:Text definition="http://www.opengis.uab.cat/ritme-natura/field/"
          <swe:value>gps</swe:value>
       </swe:Text></swe:field>
      <swe:field name="scientific_name“><swe:Text definition="http://www.opengis.uab.cat/ritme-natura/field/">
          <swe:value>Columba palumbus</swe:value>
      </swe:Text></swe:field>
      <swe:field name="FFENOFASE“><swe:Text definition="http://www.opengis.uab.cat/ritme-natura/field/">
          <swe:value>Presence detected</swe:value>
      </swe:Text></swe:field>
    </swe:DataRecord>
  </om:result>
</om:OM_Observation>
```

groundtruth2.0

Data quality element at the observation level

Observations and Measurements OGC and ISO 19156:2011(E)

```
<om:OM_Observation gml:id="ritme-natura_209_62">
  <om:procedure xlink:href="http://www.opengis.uab.cat/ritme-natura/procedure/pau_guzman"/>
  <om:featureOfInterest xlink:href="http://www.opengis.uab.cat/ritme-natura/featureOfInterest/209"/>
  <om:resultQuality>
     <DQ_AbsoluteExternalPositionalAccuracy>
        <gmd:result>
           <gmd:DQ_QuantitativeResult>
              <gmd:value>
                 <gco:Record>
                     <qml:HalfLengthConfidenceInterval level="0.683">
                        <un:values/>11<un:values>
                     </qml:HalfLengthConfidenceInterval>
                 </gco:Record>
              </gmd:value>
           </gmd:DQ_QuantitativeResult>
        </gmd:result>
     </DQ_AbsoluteExternalPositionalAccuracy>
  </om:resultQuality>
  <om:result xsi:type="swe:DataRecordPropertyType">
    <swe:DataRecord>
     <swe:field name="scientific_name“><swe:Text definition="http://www.opengis.uab.cat/ritme-natura/field/">
         <swe:value>Columba palumbus</swe:value>
```

Data quality value that accompanies an attribute

SWE Common Data Model (OGC 08-094r1)

```
<om:OM_Observation gml:id="ritme-natura_209_62">
  <om:result xsi:type="swe:DataRecordPropertyType">
    <swe:DataRecord>
      <swe:field name="elevation">
        <swe:Text definition="http://www.opengis.uab.cat/ritme-natura/field/">
          <swe:quality>
            <swe:Quality definition="http://sweet.jpl.nasa.gov/2.0/sciUncertainty.owl#Accuracy">
              <swe:label>relative Accuracy</swe:label>
              <swe:uom code="%"></swe:uom>
              <swe:value>1.5</swe:value>
            </swe:Quality>
          </swe:quality>
          <swe:value>300</swe:value>
        </swe:Text>
      </swe:field>
    </swe:DataRecord>
  </om:result>
</om:OM_Observation>
```

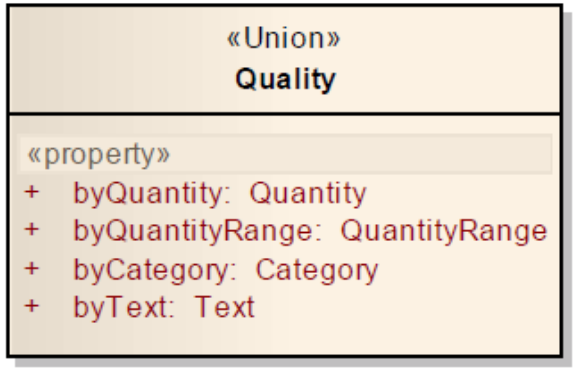
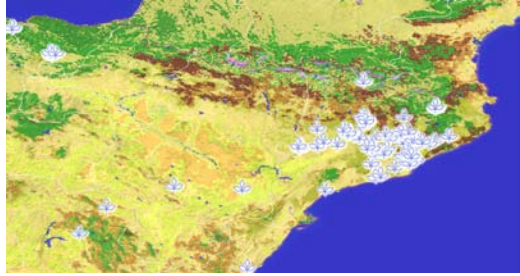


Figure 7.18 – Quality Union

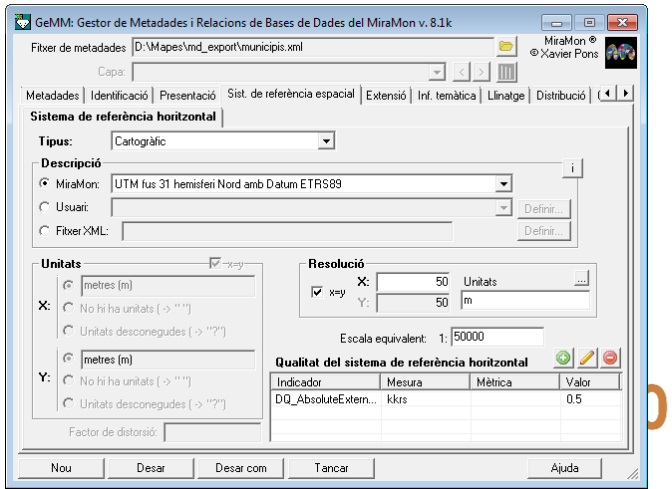
groundtruth2.0



Data quality at the *dataset level*

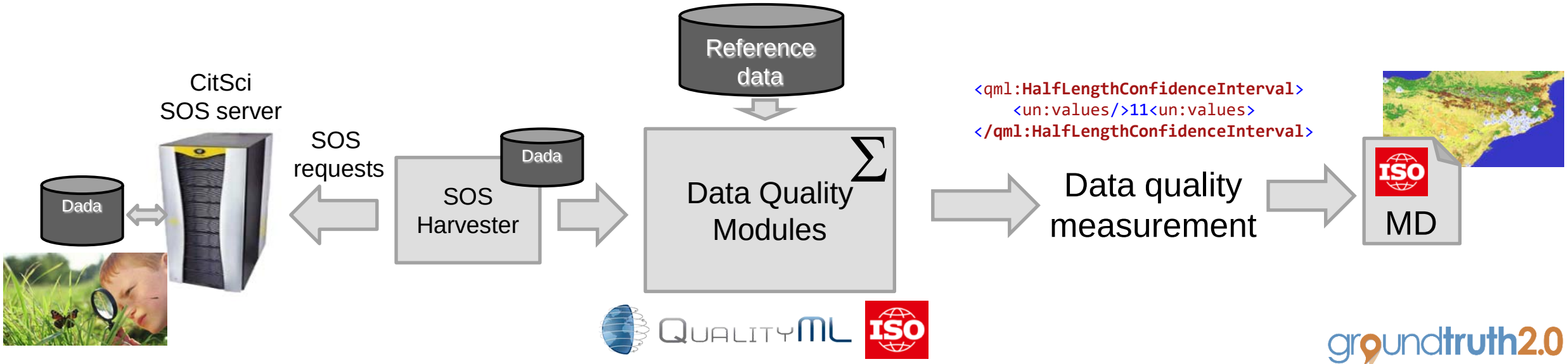
- How to evaluate it?
  - Generally is some count/average of the observation level
- How to document it?
  - It is included in the metadata of the observation collection
- How to present it?
  - Metadata viewers and editors can present this information

```
<gmd:DQ_QuantitativeResult>
   <gmd:value>
     <gco:Record>
        <qml:HalfLengthConfidenceInterval>
           <un:values/>11<un:values>
        </qml:HalfLengthConfidenceInterval>
     </gco:Record>
   </gmd:value>
</gmd:DQ_QuantitativeResult>
```



How to calculate it?

- This has been discussed several times in the literature but there are not many practical implementations.
- With SWE4CS we have an standard way to get (and harvest) the observations
- The can build modules that calculate metrics on observations



groundtruth2.0

www.QualityML.org

- It is:
  - An extension on UncertML
  - A collection of quality measurements (mainly based on ISO 19157)
  - A separation between the measure, the domain and the metrics
  - A recommendation on how to encode quality in ISO-Metadata

Quality class: Positional accuracy

Quality indicator	gmd:DQ_Element	Quality measure	Domain	Metrics
Absolute external positional accuracy	gmd:DQ_AbsoluteExternalPositionalAccuracy	**Mean absolute error** measure/MeanAbsoluteError	domain/DifferentialErrorsX	metrics/MeanAbsolute
Absolute external positional accuracy	gmd:DQ_AbsoluteExternalPositionalAccuracy	**Mean absolute error** measure/MeanAbsoluteError	domain/DifferentialErrorsY	metrics/MeanAbsolute
Absolute external positional accuracy	gmd:DQ_AbsoluteExternalPositionalAccuracy	**Mean absolute error** measure/MeanAbsoluteError	domain/DifferentialErrors2D	metrics/MeanAbsolute2D
Absolute external positional accuracy	gmd:DQ_AbsoluteExternalPositionalAccuracy	**Mean absolute error** measure/MeanAbsoluteError	domain/DifferentialErrors3D	metrics/MeanAbsolute3D
Gridded data positional accuracy	gmd:DQ_GriddedDataPositionalAccuracy	**DifferentialErrorsX**	domain/NonConformance	metrics/MeanAbsolute
Gridded data positional accuracy	gmd:DQ_GriddedDataPositionalAccuracy	**Mean absolute error** measure/MeanAbsoluteError	domain/DifferentialErrorsY	metrics/MeanAbsolute

groundtruth2.0

Example: Calculating positional accuracy of the ritme-natura service.

- CREAF SOS hosts the ritme-natura CitSci data. One of the attributes in a complex observation is the "positional_accuracy" given as a circular uncertainty value.
- QualityML database provides a matrix that can consume this type of uncertainties:
  - RMSE can use differential2D
- After getting the necessary observations we can generate an array of the result of values (the position uncertainties) an get the result RMSE

```
<swe:field name="positional_accuracy">
        <swe:Quantity definition="http://www.opengis.uab
            <swe:value>11</swe:value>
        </swe:Quantity></swe:field>
```

$$e_{xyi} = \sqrt{(e_{xi})^2 + (e_{yi})^2}$$

QUALITYML

URI:	http://www.qualityml.org/1.0/metrics/RootMeanSquareError2D
Name:	Root mean square error 2D
Alternative names:	RMSE
Definition:	Measure of the differences between 2D values predicted by a mod $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(e_{xi})^2 + (e_{yi})^2\right]}$
Parameters:	
Source:	ISO 19157
Classes:	Positional accuracy
Further information:	https://en.wikipedia.org/wiki/Root-mean-square_deviation

groundtruth2.0

Example: Completeness of the ritme-natura service.

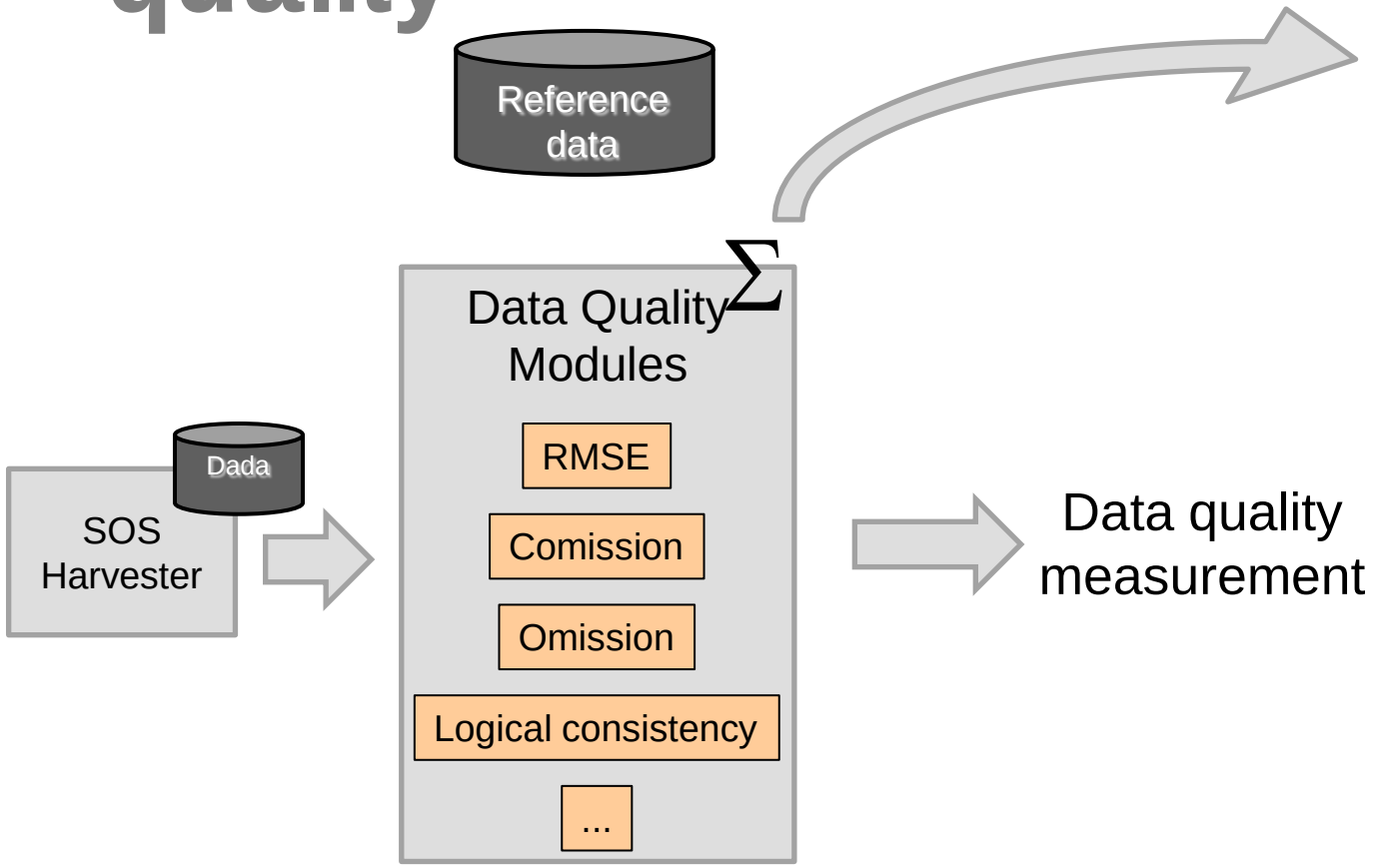
- CREAF SOS hosts the ritme-natura CitSci data. One of the attributes is the species name and the phenophase.
- We can evaluate completeness
  - Omission: are all species and phenophases represented?
  - Commission: are species or phenophases present not offered in the observation guidelines?
- After getting the necessary observations we can compare the values with a predefined list of categories and report on how many are not (or the percentages).

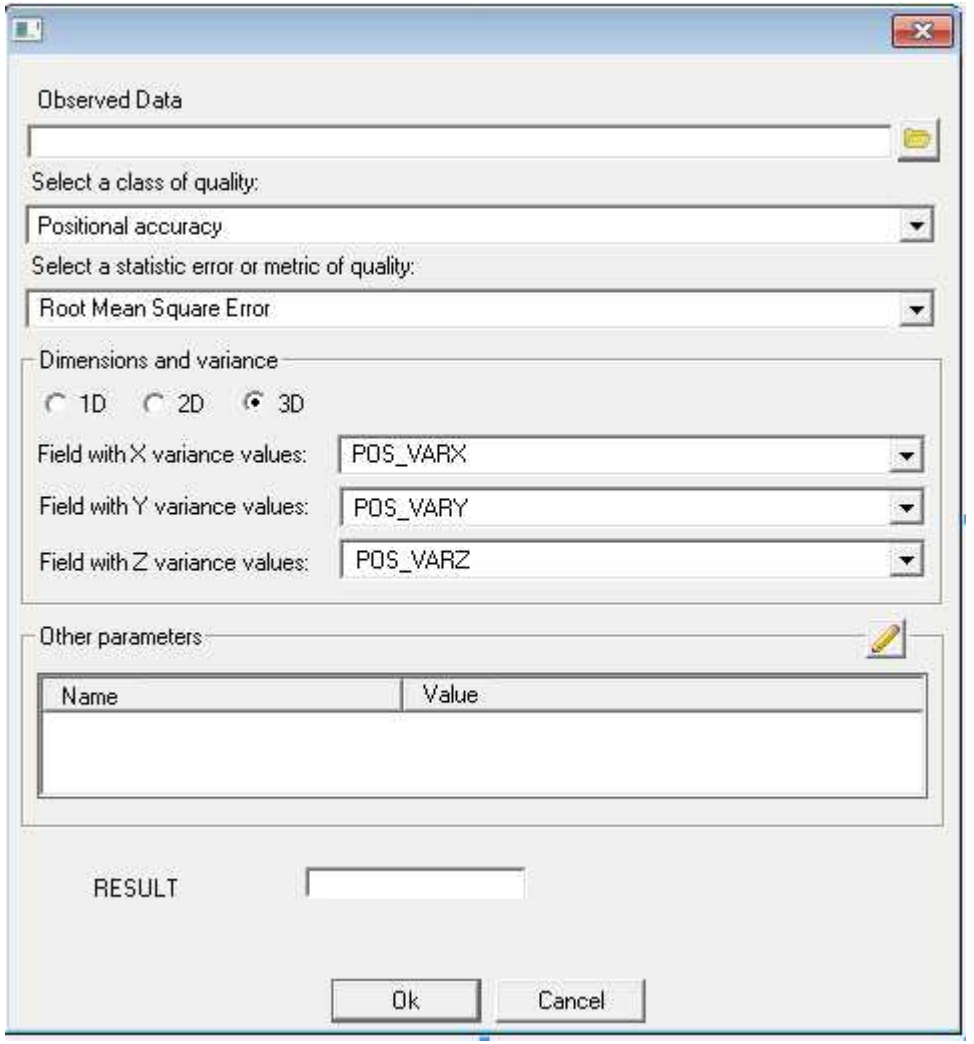
```
<swe:field name="scientific_name“>
        <swe:Text definition="http://www.opengis.uab.cat
                <swe:value>Columba palumbus</swe:value>
         </swe:Text></swe:field>
<swe:field name="FFENOFASE“>
        <swe:Text definition="http://www.opengis.uab.cat
                <swe:value>Presence detected</swe:value
         </swe:Text></swe:field>
```

QualityML

URI:	http://www.qualityml.org/1.0/metrics/RootMeanSquareError2D
Name:	Root mean square error 2D
Alternative names:	RMSE
Definition:	Measure of the differences between 2D values predicted by a mod $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(e_{xi})^2+(e_{yi})^2\right]}$
Parameters:	
Source:	ISO 19157
Classes:	Positional accuracy
Further information:	https://en.wikipedia.org/wiki/Root-mean-square_deviation

Modular architecture to calculate data quality





QualityML ISO

groundtruth2.0

How to encode it?

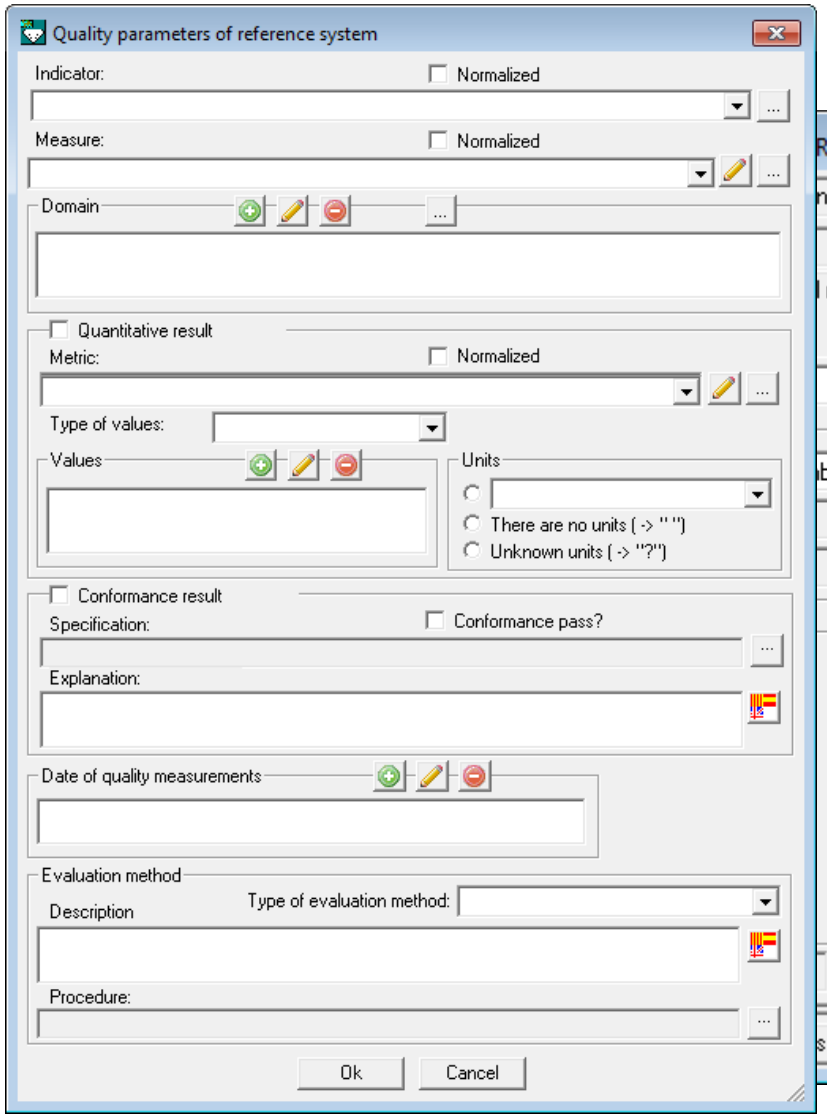
- Quality element should provide:
  - Measurement
  - Domain
  - Metrics (statistic)
  - Result
- We incorporate the data quality parameters into the Citizen Observatory metadata for the set of observations.

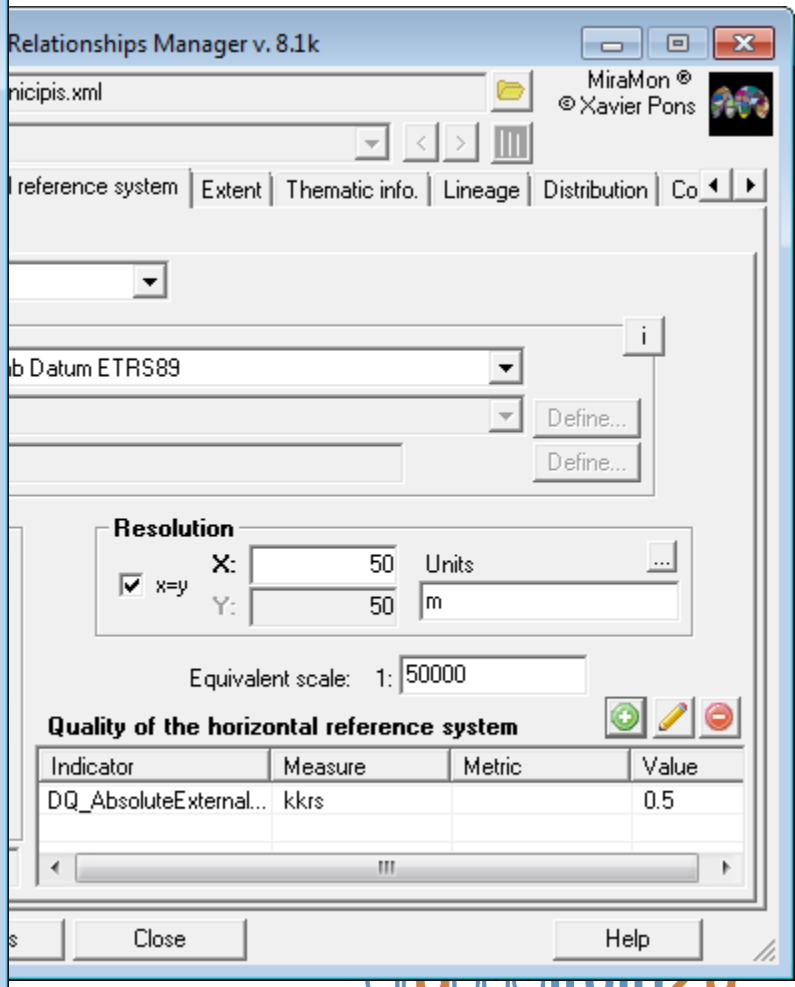
```
<gmd:MD_DataQuality>
  <gmd:report>
    <gmd:DQ_CompletenessOmission>
      <gmd:nameOfMeasure>
        <gco:CharacterString>Missing</gco:CharacterString>
      </gmd:nameOfMeasure>
      <gmd:measureIdentification>
        <gmd:MD_Identifier>
          <gmd:code>
          http://www.qualityml.org/1.0/measure/Missing
          </gmd:code>
        </gmd:MD_Identifier>
      </gmd:measureIdentification>
      <gmd:result>
        <gmd:DQ_QuantitativeResult>
          <gmd:valueType>
            <gco:RecordType xlink:href="http://www.qualityml.org/1.0/me
              Rate of missing items
            </gco:RecordType>
          </gmd:valueType>
          <gmd:valueUnit/>
          <gmd:value>
            <gco:Record>
              <qml:Items>
                <qml:rate max="100">3</qml:rate>
              </qml:Items>
            </gco:Record>
          </gmd:value>
        </gmd:DQ_QuantitativeResult>
      </gmd:result>
```

How to present it to the user?

- Metadata editors

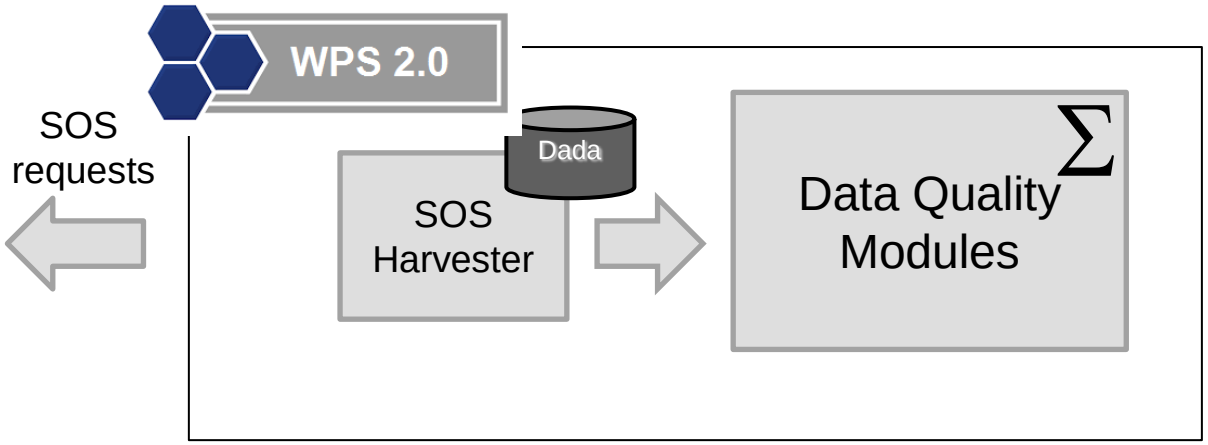
```
<gmd:DQ_QuantitativeResult>
   <gmd:value>
     <gco:Record>
        <qml:HalfLengthConfidenceInterval>
           <un:values/>11<un:values>
        </qml:HalfLengthConfidenceInterval>
     </gco:Record>
   </gmd:value>
</gmd:DQ_QuantitativeResult>
```

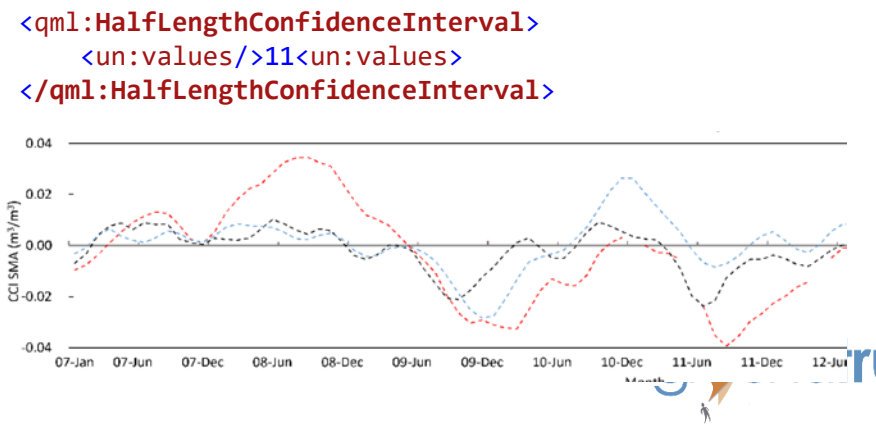




Quality as a service

- The *data quality modules* + *the SOS harvester* can be offered as a service for any Citizen Science project
- Implemented as a WPS process,
  - it can access the SOS server, retrieve the data and estimate quality indicators.
  - It can be the engine below a web that monitors the data quality of Citizen Science projects and shows the evolution of it over the time.





Citizen Science Interoperability Experiment (CitSciIE)

FRIDAY SEPTEMBER 14TH

OGC CITIZEN SCIENCE IE KICK-OFF

Kick-off in Stuttgart

- The use of OGC standards or (e.g. Sensor Web Enablement (SWE)) to support data integration among CS projects, and with other sources, esp. authoritative data (e.g by following SWE4CS);
- How to document critical metadata, including data quality aspects, and generate a data quality label.
- The integration of CS projects/campaigns in Single Sign-On system (SSO) federation;
- Project metadata standards implementation in catalogues of Citizen Science projects.

Initiators

UAB Universitat Autònoma de Barcelona

CREAF

IIASA

JRC EUROPEAN COMMISSION

esa

Wilson Center

Support

NEXTGEOSS

CoP3

OGC

We observe

Earth Challange 2020

EARTH CHALLENGE 2020 A CITIZEN SCIENCE INITIATIVE

Goal of engaging millions of global citizens in collecting one billion data points in areas including air and water quality, pollution and human health.

Citizen science volunteers around the world, working with professional scientists, will collect and share earth science data of their local communities on an unprecedented scale.

There is a sensibility for data quality in the integration of many citizen science project in a single initiative

groundtruth2.0

Thank you

Joan Masó
CREAF

groundtruth2.0 partners:

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 689744.