OGC Citizen Science engeneering report

Publication Date: YYYY-MM-DD

Approval Date: YYYY-MM-DD

Submission Date: YYYY-MM-DD

Reference number of this document: OGC YY-nnn

Reference URL for this document: http://www.opengis.net/doc/PER/t14-D012

Category: Public Engineering Report

Editor: Joan Masó

Title: OGC Citizen Science engeneering report

OGC Engineering Report

COPYRIGHT

WARNING

This document is not an OGC Standard. This document is an OGC Public Engineering Report created as a deliverable in an OGC Interoperability Initiative and is not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard. Further, any OGC Engineering Report should not be referenced as required or mandatory technology in procurements. However, the discussions in this document could very well lead to the definition of an OGC Standard.

LICENSE AGREEMENT

complying with any local laws in your jurisdiction which may impact your right to import, export or use the Intellectual Property, and you represent that you have complied with any regulations or registration procedures required by applicable law to make this license enforceable.

Table of Contents

Chapter 1. Summary

TBD

The Summary clause shall define without ambiguity the subject of this document and the aspect(s) covered. It shall be succinct so that it can be used as a text for bibliographic purposes. Briefly, it shall contain the key results of the work described in the ER.

The summary shall further contain a business value statement that should describe the value of this Engineering Report to improve interoperability, advance location-based technologies or realize innovations.

The summary shall contain the key findings in a concise form. A more detailed description of the findings should be in the body of the report.

This section shall be between 2-3 paragraphs and not longer than 507 words.

1.1. Requirements & Research Motivation

This experiment was designed to demonstrate how current ICT-based tools can be applied together to allow better citizen participation in CS projects and enable better reuse of the data gathered. Citizen Science is highly transdisciplinary and heterogeneous by nature, and current standardization efforts already occur in the OGC (e.g. addressing data model and sharing issues) as well as outside the OGC (primarily addressing project descriptions and data set metadata). Citizen Science projects might benefit from concrete examples and best practices required to achieve the full benefits of interoperability. OGC is in the ideal position to develop and provide such best practice guidance to the international community. Developed solutions in this IE should be applicable to most Citizen Science projects. Findings from this IE will be generalized in this manner and might set the basis for additional experimentation in the future.

The FP7 Citizen Observatory Web (COBWEB) project was the first to propose the use of SWE in CS. This work resulted in an OGC public discussion paper available on the OGC website (OGC 16-129). The discussion paper describes a data model for the standardized exchange of citizen science sampling data based on SWE standards. This discussion paper was the initial motivation for this IE.

On the other hand, the Citizen Science Association’s International Working Group on Citizen Science Data and Metadata that developed the (PPSR-Core), the European Citizen Science Association (ECSA) has a working group that recognizes the value of standardization in the CS activities (supported by a COST Action). However, these activities could benefit from some experimentation that would be able to suggest common best practices while recognizing the particularities and current approaches in different thematic domains, such as biodiversity monitoring. Citizen Science can complement authoritative in-situ observations and fill the information gaps in numerous scientific disciplines that could be essential for informed decision making. In that sense, the way Citizen Science can be integrated into The Global Earth Observation System of Systems (GEOSS) (including GEOSS-Data Core as the pool to promote and share open and free data) is still under investigation.

The 'Ecosystem of Citizen Observatories (CO) for Environmental Monitoring' — WeObserve project is a Horizon 2020 funded project focused on improving the coordination between existing COs and related regional, European and international activities. WeObserve tackles three key challenges that face COs: awareness, acceptability and sustainability. The CoP3 is about Interoperability of Citizen Science projects. WeObserve project – via its CoP activities – has represented an opportunity to promote interoperability experiment in collaboration with the OGC. Such collaboration would address questions raised in the SWE4CS discussion. In addition, it offers the possibility to directly feed the results into the OGC relevant standards and promotes their usage within GEOSS (as an important user community of OGC standards).

In anticipation of the 50th Anniversary of Earth Day in 2020, Earth Day Network, the Woodrow Wilson International Center for Scholars, and the U.S. Department of State, through the Eco-Capitals Forum, announce Earth Challenge 2020, a Citizen Science Initiative. This initiative is in collaboration with Connect4Climate – World Bank Group, Conservation X Labs, Hult Prize, National Council for Science and the Environment (NCSE), Open Geospatial Consortium (OGC), Reset, SciStarter, UN Environment and others to be announced. Earth Challenge 2020 will help engage millions of global citizens in collecting one billion data points in areas including air quality, water quality, biodiversity, pollution and human health. Earth Change 2020 data will be shared through the GEOSS Portal.

A precise descriptions of the requirements that have been addressed by the work documented in this Engineering Report; together with the research motivation that answers the fundamental question: What motivated us to address this topic in this report?

1.2. Prior-After Comparison

This section shall provide a prior-after comparison. It describes the situation/status of discussion in the OGC working groups being most relevant for the addressed topic. This part is reviewed in close detail by the appropriate SWG/DWG to ensure that the latest developments have been considered. The section will be complemented at the end of the initiative by comparing the results documented in the ER with the original situation.

1.3. Recommendations for Future Work

This OGC Interoperability Experiment ended on June 2019, but a second IE is foreseen for next year.

New possible topics for next IE to be discussed among the members include the following:

● There is a need for clarifying how to coordinate infrastructures for citizen science in Europe, and adopt standard procedures for data sharing and single sign on. Solving this issue will help in connecting CitSci to GEO. Steffen Fritz (IIASA) has proposed a side event in the next EuroGEOSS workshop to discuss this with the relevant players. This is emerging as a new activity in the WeObserve Interoperability Community of Practice that is related but not directly connected to the IEs.

- The WMO (World Meteorological Organization) is concerned about the amount of different CitSci activities that are being organized by meteorological organizations. It is looking for ways of taking advantage of this new data stream but problems of standardization of what is measured and how data is being shared arise. WMO has detected the potential of these data streams and would like to harmonize the situations to make data more useful for weather predictions in the future.

- OGC is promoting a new generation of web services based on OpenAPI. It is unclear how this could impact the use of OGC standards by CitSci projects but it is seen as an opportunity to make OGC standards more usable and compatible with IT mainstream. A hackathon to develop OGC API specifications is organized on June 20-21 in London.

The definition of the follow up IE will start in June when the first one will finalize and the ER is produced.

This section should answer the question: What does this ER mean for the Working Group and OGC in general? What aspects shall be addressed next? In any specific order? What actions are necessary?

This is a write up for why this ER should be important to the working group and OGC. This paragraph provides recommendations on how to further proceed with the achievements documented in this ER.

1.4. Document contributor contact points

All questions regarding this document should be directed to the editor or the contributors:

Contacts

Name	Organization
Joan Maso	CREAF
contributor	from org

1.5. Foreword

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

Chapter 2. References

The following normative documents are referenced in this document.

- OGC 10-004r3, Topic 20: Observations and Measurements [http://portal.opengeospatial.org/files/?artifact_id=41579]
- OGC 10-025r1, Observations and Measurements - XML Implementation [http://portal.opengeospatial.org/files/?artifact_id=41510]
- OGC 15-100r1, OGC Observations and Measurements – JSON implementation [https://portal.opengeospatial.org/files/64910]
- OGC 12-006, OGC® Sensor Observation Service Interface Standard v2.0 [https://portal.opengeospatial.org/files/?artifact_id=47599]
- OGC 08-094r1, OGC® SWE Common Data Model Encoding Standard v2.0 [http://portal.opengeospatial.org/files/?artifact_id=41157]
- 15-042r5 OGC TimeseriesML – XML Encoding of the Timeseries Profile of Observations and Measurements v1.2 [http://docs.opengeospatial.org/is/15-042r5/15-042r5.html]

Even if the following is an OGC discussion paper that is not an OGC standard and cannot be considered strictly a normative reference, it is actually the bases for several sections of this document and should be considered as and important background:

- OGC 16-129, Standardized Information Models to Optimize Exchange, Reusability and Comparability of Citizen Science Data (SWE4CS) [https://portal.opengeospatial.org/files/?artifact_id=70328]

Chapter 3. Terms and definitions

For the purposes of this report, the definitions specified in Clause 4 of the OWS Common Implementation Standard OGC 06-121r9 [https://portal.opengeospatial.org/files/?artifact_id=38867&version=2] shall apply. In addition, the following terms and definitions apply.

- Citizen Observatory (CO)

  ```
  COs are community-based environmental monitoring and information systems,
  that invite individuals to share observations, typically via mobile phone
  or the web (from: https://www.weobserve.eu/about/citizen-observatories)
  ```

- Citizen Science 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 (from: https://www.uen.org/crowdandcloud/citizen.shtml)
- Citizen Science Association

  ```
  A network that seeks to promote and advance citizen science in a region
  or around the world. Examples are the American Citizen Science
  Association (CSA), The European Citizen Science Association (ECSA) or
  even the Citizen Science Global Partnership (CSGP)
  ```

- Citizen Science Federation

  ```
  A network of Citizen Science that aims to aggregate innovative EO
  technologies, mobile devices, community-based environmental monitoring,
  data collection, interpretation and information delivery systems to
  empower communities to monitor and report on their environment. An
  example of this is the The LandSense Federation
  ```

- Community of Practice (CoP)

  ```
  Community which work to consolidate practice-based knowledge of COs
  sharing information and resources as well as developing guidelines and
  toolkits for COs (from: https://www.weobserve.eu/cops/)
  ```

3.1. Abbreviated terms

NOTE: The abbreviated terms clause gives a list of the abbreviated terms and the symbols necessary for understanding this document. All symbols should be listed in alphabetical order. Some more frequently used abbreviated terms are provided below as examples.

- CitSciIE Citizen Science Interoperability Experiment
- CO Citizen Observatory
- CoP Community of Practice
- COST European Cooperation in Science and Technology
- CS Citizen Science
- CS DWG Citizen Science Domain Working Group
- CSGP Citizen Science Global Partnership
- ECSA European Association of Citizen Science
- EO Earth Observation
- ICT Information and Communication Technologies
- IE Interoperability Experiment
- O&M Observation and Measurements
- PPSR Public Participation in Scientific Research
- SOS Sensor Observation Service
- SSO Single Sign On
- SWE Sensor Web Enablement
- SWE4CS Sensor Web Enablement for Citizen Science
- TC Technical Committee
- TIE Technology Integration Experiments
- WPS Web Processing Service

Chapter 4. Overview

This Engineering Report focuses on the findings of the first phase of the Citizen Science Interoperability Experiment (CitSciIE).

The primary focus of the OGC CitSciIE experiment was to demonstrate the interoperability of Citizen Observatories and Citizen Science (CS) projects and the way OGC standards can be applied to Citizen Science, including possible relationships to other relevant standards from the community. In particular, a subset the originally proposed topics were being addressed based on the participant organizations:

- The use of OGC standards (e.g. Sensor Web Enablement for Citizen Science SWE4CS) to support data integration among CS projects, and with other sources, esp. authoritative data;
- The use of ISO standards, and OGC publications, and community resources to document data quality aspects (e.g. UncertML, QualityML);
- The integration of CS projects/campaigns in a Single Sign-On system (SSO) federation;
- The relationships between OGC standards and data and metadata standards currently used by Citizen Science projects.

The desired outcome of this experiment was to:

1. Successfully demonstrate how OGC standards (e.g. SWE) are applicable to Citizen Science, document available supporting tools, identify the challenges of using OGC SWE standards (or Internet of Things equivalent solutions) within current Citizen Science projects, and propose a way forward. Make recommendations to the Earth Science 2020 initiative on which OGC standards should be utilized to underpin interoperable data collection and sharing.
2. Successfully demonstrate how to estimate Citizen Science data quality and make the quality indicators and conformity available in the document and in supporting tools, and link them to the OGC SWE standards (or Internet of Things equivalent solutions) within current CS projects, and propose a way forward.
3. Determine the security considerations and the available tools to support an SSO federation that helps users in participating in several projects by using a single user account.
4. Assess the possible relationships of OGC standards (e.g. SensorML) with other existing standards in the field (e.g. Public Participation in Scientific Research (PPSR) - Core, the ontology developed by the COST Action on Citizen Science, and the Citizen Science Definition Service (CS-DS) developed in the NextGEOSS project).
5. Satisfy and document the necessary requirements to integrate Citizen Science into Global Earth Observations System of Systems (GEOSS) by using OGC standards.

This IE has been promoted by the OGC Citizen Science Domain Working Group, the WeObserve and NextGEOSS H2020 projects, and The Earth Challenge 2020 project as supported by National

Geographic Society. This IE contributes not only to the interoperability and possibly standardization program of the OGC, but also to the Global Earth Observation System of Systems (GEOSS). This work is also relevant to the foundational objectives of the Citizen Science Global Partnership (CSGP). Regional and national Citizen Science Associations will equally benefit from the results of this OGC IE.

4.1. Structure of the activities

The official kick off meeting for the OGC CitSciIE experiment was held on Friday 14th September 2018 at the OGC TC meeting in Stuttgart. Activities have continued until March 2019.

During the Kick off meeting of the Experiment the following subgroups emerged:

- V: Vocabularies for organizing Citizen Science projects. There was a discussion on essential variables but also on other kind of practices that can be associated to vocabularies, i.e. on how to publish vocabularies (PublishingDefs) or on defining a list of vocabularies that could be useful to experiment with (observations, project descriptions, general glossaries of terms).
    - Working item V.1: A list of the current projects that Wilson Center knows that can align with the Earth Challenge topics (air and water quality, pollution, human health and eventually biodiversity) and extraction of a common set of variables they usually cover.
    - Working item V.2: Analysis of data models that contributors in the experiment can bring in: Air quality (HackAir), Biodiversity (Atlas of Living Australia & Natusfera), Mosquito (CREAF), Land Use (IIASA), Phenology (CREAF), Invasive Alien Species (JRC).
    - Working item V.3: Consider the COST action metadata model for inclusions as another vocabulary, this might include a set of definitions of phenomena that are being addressed by CS initiatives (based on the inventory of citizen science activities for environment policies).
- D: Data sharing using OGC standards such as O&M and SOS. A pool of services were identified for participating in an interoperability experiment, including SOS services and clients and citizen science project databases and APIs.
    - Working item D.1: A set of instructions on how a CS project can easily setup an SOS service. It could include 52North implementation and might include MiraMon SOS (with some work in the implementation). It should address the case of a small project contributing to the Earth Challenge 2020.
    - Working item D.2: Create an SOS endpoint for HackAir data with minimum resources
    - Working item D.3: Define the requirements for a data provider that could assist Wilson Center in setting up the challenge database. It should consider upload of data into the system. It seemed to go for a harvest system instead of a federated system. It could describe a possible architecture to allow the dialog between the central database and the small contributing projects. It should impose data sharing requirements (services o APIs) on the central database.

- S: Connection between Landsense federation and JRC user system.
  - Working item S.1: Interoperability test on the integration of LS-SSO and JRC-SSO
- Q: Data quality
  - Working item Q.1: Write a document on perspectives of the different quality aspects: Quality assessment (ISO 19157-QualityML), Quality improvement, Quality plan, Data Management principles (ISO 8001), Quality documentation, Quality communication
  - Working item Q.2: Perfect the quality measurement system based on WPS and SOS harvest by demonstrating the concept in practice. Also include in the SOS harvesting the possibility to have a query for assessing the quality of "views"/"selections"/"fragments" of a dataset.
    - Connection with: D.2
  - Working item Q.3: Refine the QualityML vocabulary with new entries considering the work done in Australia
    - Connection with: D.3
  - Working item Q.4: Add new entry point the QualityML for other common vocabulary formats like TTL etc.
    - Connection with: V

For each of the subgroups a chair and the main participants and contributors were identified. Responsible persons were also assigned to each of the working items.

4.2. Results detailed in next sections

These are the main activities and outcomes of the interoperability experiment detailed by activity.

Data sharing using OGC standards such as O&M and SOS

This activity has been the most active one. During the testbed the following servers have been deployed: MiraMon SOS server, Grow SOS, DLR istSOS SOS and 52north SOS. Three clients have also been produced: MiraMon SOS browser, Grow SOS data viewer and 52north Helgoland. In the last meeting at the EGU, the group was able to demonstrate interoperability by connecting the SOS clients to the SOS services and showing the data on clients, sometimes mixing data form different services and datasets in a single view. This is the most significant result of the experiment and is being extensively documented in this Engineering Report in sections 5, 6 and 7

Data quality

Two quality vocabularies have been detected: an Australian work done by Peter Brenton team (https://github.com/tdwg/bdq) and the QualityML vocabulary developed by CREAF in the GeoViQua project. The intention was to do a comparison of both approaches but we were not able to do that in the timeframe of the first interoperability experiment. Section 8 describes the current status of the activity. It is foreseen that the second IE will continue what was started here.

Definition server for organizing Citizen Science projects

The objective of this activity was to support the Earth Challenge 2020 research questions. The questions were defined during the first month of the experiment and now it is time to analyse them in terms of data needs and thematic vocabularies they are going to use. That is the reason this activity has not resulted in tangible outputs and will be reintroduced in the second interoperability experiment. Details of this development are described in section 9.

Connection between Landsense federation and JRC user system

Andreas Matheus SME (Secure Dimensions) was very active in providing demonstrations and information on how the LandSense federation works and how other projects can be included in the federation and use the single Sign on facility. Unfortunately, no other member of the CoP had the resources to apply the SSO on their services or clients and take advantage of the Andreas offering. The activity resulted in a video demonstrations that are publicly available here: https://portal.opengeospatial.org/files/?artifact_id=81550. Section 10 Details the current status of the activity

Other

Section 11 summarizes the lessons learned that can be applied to GEOSS.

In addition to these activities, another activity about quality annotating scientific documentation in a standard way was proposed by Lucy Bastin. A video was recorded that summarizes the idea: https://portal.opengeospatial.org/files/?artifact_id=82544.

Chapter 5. O&M for Cit Sci

In a *feature model* all characteristics of a feature are considered properties of the feature and are not semantically separated at the abstract level.

The Observations and Measurements(O&M) standard (OGC 10-004r3, Topic 20: Observations and Measurements) defines a data model that observations where. main concepts are separated as represented in the Figure 1:

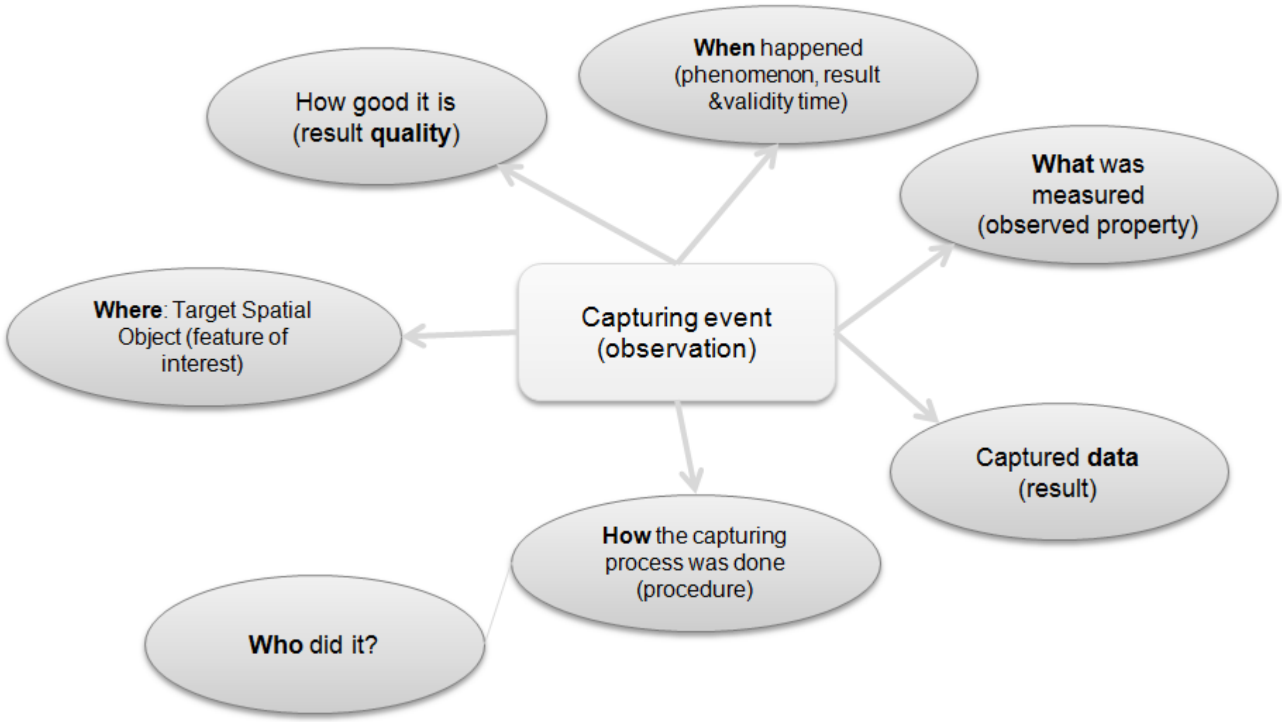


Figure 1. Observation main concepts

For each observation, the O&M allows us to document the following characteristics:

- Where the observation is located: Even if the observations was done remotely with a camera or a drone, it is commonly more relevant to know the position of the observed phenomena (the sensor position can also be recorded).
- When the observation took place and what time period it represents: Even if samples were collected and analyzed later, it is commonly more relevant to know the instant or period of the observed phenomenon.
- How the observation was done: This will describe the procedure and instrument used to capture the phenomenon.
- Who did the observation: The procedure and instrument used to capture the phenomenon was installed or used on side by someone. In citizen Science, where many observers contribute small pieces of information that together will form a dataset, it is particularly important to record at least an observer identifier.
- What was measured: This will define the property names and units of measure of the

variables observed.

- What data was collected: This will record the actual values of the properties measured.
- What is the expected quality of the observation: If an estimation of the quality of the observation was done, it is important to document it.

In the O&M data model, there aspects are clearly separated semantically as shown in Figure 2. This is the main value of the O&M model and its usage SOS (or the SensorThingsAPI that uses a very similar approach to model the data), but it is also the main handicap in applying the standard.

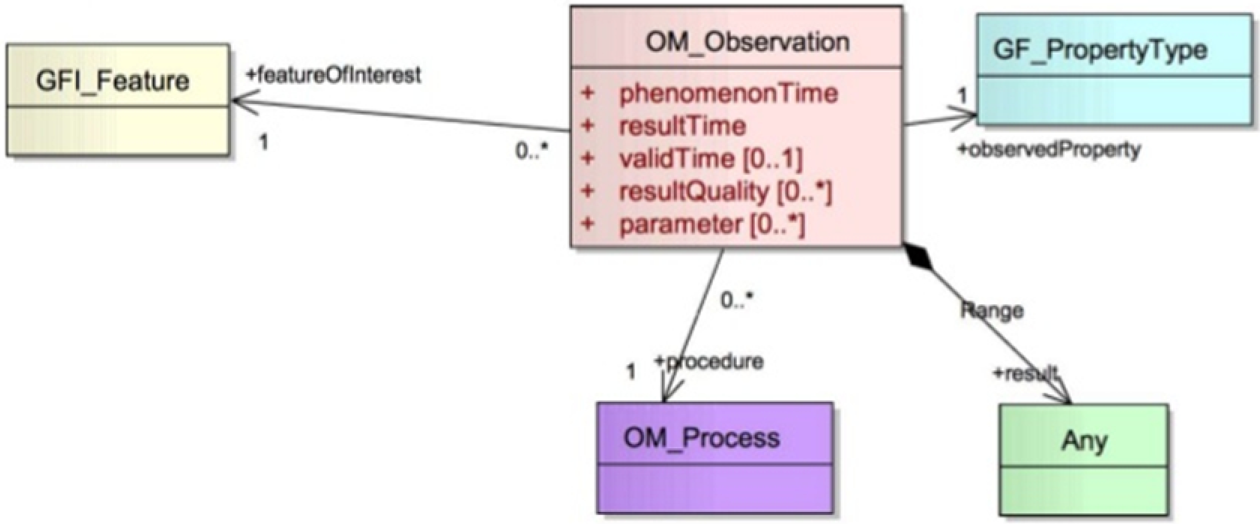


Figure 2. Common Model For Oservations and Measurements

Concept	O&M	type
Where	featureOfInterest	GFI_Feature
When	phenomenonTime, resultTime	
How	procedure	OM_Process
Who	procedure	OM_Process
What	observedProperty	GF_PropertyType
Data	result	Any
Quality	resultQuality	

Table 1. Concepts in an observation mapped to the O&M model

Even if this aspects are separated, the O&M model gives a lot of flexibility in defining the properties and this flexibility can condition interoperability when trying to combine data from different sources. The standard give us freedom to select among the different geometries provided by GML to define the featureOfInterest. The standard gives us even more freedom on the data collected that can have any imaginable structure.

That is the reason why the data model used to represent the data gathered by a Citizen Observatory need to be carefully considered before even starting the first campaign. Data models can be designed in UML for clarity but they are later encoded in XML. XML is the only official encoding that O&M referenced in the OGC website (OGC 10-025r1 Observations and Measurements - XML Implementation v2.0). Nevertheless, there is JSON alternative discussed in a OGC discussion paper (OGC 15-100r1 OGC Observations and Measurements – JSON implementation) that does not represent an official position of the OGC but can be implemented anyway. As it will be discussed latter, the interpretation of long XML files might be to slow in web browsers, and, in this cases, a JSON encodings is regarded as a good alternative. This alternative can be the proposed OGC 15-100r1 or the alternative proposed by the SensorThings API.

5.1. The GT20 examples

In the Ground Truth 2.0 project we have been using the MiraMon implementation of O&M. This implementation assumes a simplified situation that considers that each observation can be represented by a single row in a CSV or in a single record of a database table. Coordinates are represented as a single point. In this situation, we select which column names represent the *phonomenonTime*, the *procedure* (that actually is including the user name) and the *featureOfInterest* (the coordinates). The rest of the columns are considered part of the data record that needs to be provided as the *result*.

The section 8.2.1 of the OGC 08-094r1, OGC® SWE Common Data Model Encoding Standard v2.0 describes a way to encode a *DataRecord* as an array of fields that can numbers, strings, dates, etc. In our simplified assumption, this results ideal to wrap the properties of the observations that cannot be mapped to any other O&M aspect. This practice is consistent with the section 7.2.8 of the SWE4CS discussion paper.

The following is an example of how a water quality observation is represented following the O&M model and encoded in XML:

Example of XML encoding of the water quality observation

```
<om:OM_Observation gml:id="vatten-fokus_2_1">
    <om:type xlink:href="http://www.opengis.net/def/observationType/OGC-
OM/2.0/OM_ComplexObservation"/>
    <om:procedure xlink:href="http://www.opengis.uab.cat/vatten-
fokus/procedure/22655"/>
    <om:observedProperty xlink:href="http://www.opengis.uab.cat/vatten-
fokus/observedProperty"/>
    <om:featureOfInterest xlink:href="http://www.opengis.uab.cat/vatten-
fokus/featureOfInterest/2"/>
    <om:result xsi:type="swe:DataRecordPropertyType">
        <swe:DataRecord>
            <swe:field name="CREA_DATE">
                <swe:Text definition="http://www.opengis.uab.cat/vatten-
fokus/field/Creation_Date">
```

```
                    <swe:value>07/12/2018 17:23</swe:value>
                </swe:Text>
            </swe:field>
            <swe:field name="SITE_NAME">
                <swe:Text definition="http://www.opengis.uab.cat/vatten-
fokus/field/Site_name">
                    <swe:value>Dunkershall. V¤gtrumma uppst¤ms.</swe:value>
                </swe:Text>
            </swe:field>
            <swe:field name="LAND_USE">
                <swe:Text definition="http://www.opengis.uab.cat/vatten-
fokus/field/Land_use_in_the_immediate_surroundings">
                    <swe:value>Agriculture</swe:value>
                </swe:Text>
            </swe:field>
            <swe:field name="BANK_VEGE">
                <swe:Text definition="http://www.opengis.uab.cat/vatten-
fokus/field/Bank_vegetation">
                    <swe:value>Grass</swe:value>
                </swe:Text>
            </swe:field>
            <swe:field name="NITRATE">
                <swe:Quantity definition="http://www.opengis.uab.cat/vatten-
fokus/variable/NITRATE">
                    <swe:uom/>
                    <swe:value>1.50</swe:value>
                </swe:Quantity>
            </swe:field>
            <swe:field name="PHOSPHATE">
                <swe:Quantity definition="http://www.opengis.uab.cat/vatten-
fokus/variable/PHOSPHATE">
                    <swe:uom/>
                    <swe:value>0.075</swe:value>
                </swe:Quantity>
            </swe:field>
            <swe:field name="WATER_COLOR">
                <swe:Text definition="http://www.opengis.uab.cat/vatten-
fokus/field/Estimated_water_colour">
                    <swe:value>Colourless</swe:value>
                </swe:Text>
            </swe:field>
        </swe:DataRecord>
    </om:result>
</om:OM_Observation>
```

The following is an example of how two air quality observations are represented following the

O&M model and encoded in JSON:

Example of encoding of the air quality observation


```
{
    "id":"meet-mee-mechelen_1_0",
    "type" : "http://www.opengis.net/def/observationType/OGC-
OM/2.0/OM_ComplexObservation",
    "phenomenonTime" : "2017-11-19 17:20:00+01",
    "resultTime" : "2017-11-19 17:20:00+01",
    "procedure" : "http://www.opengis.uab.cat/meet-mee-mechelen/procedure/5",
    "observedProperty" : "http://www.opengis.uab.cat/meet-mee-
mechelen/observedProperty",
    "featureOfInterest" : "http://www.opengis.uab.cat/meet-mee-
mechelen/featureOfInterest/1",
    "result": {
        "type":"DataRecord",
        "field":[
            {
                "name" : "CAMPAIGN",
                "type" : "Text",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/field/CAMPAIGN",
                "value" : "Oct-Nov2017"
            },
            {
                "name" : "bc_aggr",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/bc_aggr",
                "value" : "3155"
            },
            {
                "name" : "bc_aggr_mi",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/bc_aggr_mi",
                "value" : "80"
            },
            {
                "name" : "bc_aggr_ma",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/bc_aggr_ma",
                "value" : "16413"
            },
            {
```

```
                "name" : "bc_aggr_st",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/bc_aggr_st",
                "value" : "3398"
            },
            {
                "name" : "uncertaint",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/uncertaint",
                "value" : "0.50"
            }
        ]
    }
},
{
    "id":"meet-mee-mechelen_2_1",
    "type" : "http://www.opengis.net/def/observationType/OGC-
OM/2.0/OM_ComplexObservation",
    "phenomenonTime" : "2017-11-19 17:20:06+01",
    "resultTime" : "2017-11-19 17:20:06+01",
    "procedure" : "http://www.opengis.uab.cat/meet-mee-mechelen/procedure/5",
    "observedProperty" : "http://www.opengis.uab.cat/meet-mee-
mechelen/observedProperty",
    "featureOfInterest" : "http://www.opengis.uab.cat/meet-mee-
mechelen/featureOfInterest/2",
    "result": {
        "type":"DataRecord",
        "field":[
            {
                "name" : "CAMPAIGN",
                "type" : "Text",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/field/CAMPAIGN",
                "value" : "Oct-Nov2017"
            },
            {
                "name" : "time_first",
                "type" : "Text",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/field/time_first",
                "value" : "2017-11-06 08:00:18+01"
            },
            {
                "name" : "bc_aggr",
                "type" : "Quantity",
```

```
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/bc_aggr",
                "value" : "3382"
            },
            {
                "name" : "bc_aggr_mi",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/bc_aggr_mi",
                "value" : "80"
            },
            {
                "name" : "bc_aggr_ma",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/bc_aggr_ma",
                "value" : "17256"
            },
            {
                "name" : "bc_aggr_st",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/bc_aggr_st",
                "value" : "3663"
            },
            {
                "name" : "number_of_",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/number_of_",
                "value" : "25"
            },
            {
                "name" : "number_o_1",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/number_o_1",
                "value" : "13"
            },
            {
                "name" : "mean_numbe",
                "type" : "Quantity",
                "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/mean_numbe",
                "value" : "7"
            },
            {
```

```
                  "name" : "uncertaint",
                  "type" : "Quantity",
                  "definition" :"http://www.opengis.uab.cat/meet-mee-
mechelen/variable/uncertaint",
                  "value" : "0.50"
              }
          ]
      }
}
```

These examples were produced by SOS requests to this URL: http://www.ogc3.uab.cat/cgi-bin/CitSci/MiraMon.cgi?. A client connecting to this service can be found here: http://www.ogc3.uab.cat/gt20/

5.2. HackAir examples

To illustrate the flexibility of the O&M we have included this air quality report that shows how HackAir data is presented by a 52North SOS implementation. In this case the *result* presents a single numerical value while the other information is provided as parameters. This approach is consistent with section 7.2.2.5 of the O&M standard.

Example of encoding of the water quality observation

```
<om:OM_Observation gml:id="o_499">
    <om:type xlink:href="http://www.opengis.net/def/observationType/OGC-
OM/2.0/OM_Measurement"/>
    <om:phenomenonTime>
        <gml:TimeInstant gml:id="phenomenonTime_499">
            <gml:timePosition>2019-01-01T00:00:12.000Z</gml:timePosition>
        </gml:TimeInstant>
    </om:phenomenonTime>
    <om:resultTime xlink:href="#phenomenonTime_499"/>
    <om:procedure xlink:href="sensors_arduino_1000"/>
    <om:parameter>
        <om:NamedValue>
            <om:name xlink:href="PM2.5_AirPollutantIndex"/>
            <om:value xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:type=
"xs:string">bad</om:value>
        </om:NamedValue>
    </om:parameter>
    <om:parameter>
        <om:NamedValue>
            <om:name xlink:href="http://www.opengis.net/def/param-name/OGC-
OM/2.0/samplingGeometry"/>
            <om:value xmlns:ns="http://www.opengis.net/gml/3.2" xsi:type=
```

```
"ns:GeometryPropertyType">
                <ns:Point ns:id=
"Point_sp_45C0E376C40E98E8EC0D48C05F7558C2FFD15245">
                    <ns:pos srsName=
"http://www.opengis.net/def/crs/EPSG/0/4326">52.063269625917
4.5077472925186</ns:pos>
                </ns:Point>
            </om:value>
        </om:NamedValue>
    </om:parameter>
    <om:parameter>
        <om:NamedValue>
            <om:name xlink:href="source"/>
            <om:value xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:type=
"xs:string">sensors_arduino</om:value>
        </om:NamedValue>
    </om:parameter>
    <om:parameter>
        <om:NamedValue>
            <om:name xlink:href="user"/>
            <om:value xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:type=
"xs:string">sID :1000</om:value>
        </om:NamedValue>
    </om:parameter>
    <om:observedProperty xlink:href="PM2.5_AirPollutantValue" xlink:title=
"PM2.5_AirPollutantValue"/>
    <om:featureOfInterest xlink:href="sensors_arduino_1000"/>
    <om:result xmlns:ns="http://www.opengis.net/gml/3.2" uom="µg/m3"
xsi:type="ns:MeasureType">130.67</om:result>
</om:OM_Observation>
```


A service producing this type of results can be seen here: https://nexos.demo.52north.org/52n-sos-hackair-webapp/service

5.3. GROW example

In the GROW project the SME Hydrologic has develop an SOS service that uses a O&M observation. In this case, a single number is provided as the *result* of the observation and additional parameters are transported.

```xml
<OM_Observation xmlns="http://www.opengis.net/om/2.0">
    <type gml:remoteSchema="http://www.opengis.net/def/observationType/OGC-
OM/2.0/OM_Measurement" />
    <phenomenonTime>
        <gml:TimePeriod>
            <gml:beginPosition>2018-09-03T09:01:38.000Z</gml:beginPosition>
            <gml:endPosition>2018-09-03T09:01:38.000Z</gml:endPosition>
        </gml:TimePeriod>
    </phenomenonTime>
    <resultTime>
        <gml:TimeInstant>
            <gml:timePosition>2018-09-03T09:01:38.000Z</gml:timePosition>
        </gml:TimeInstant>
    </resultTime>
    <procedure>Grow.Thingful.Sensors_je47sfac</procedure>
    <observedProperty nilReason=
"Thingful.Connectors.GROWSensors.AirTemperature" />
    <featureOfInterest nilReason="je47sfac" />
    <result>20.64</result>
</OM_Observation>
```

5.4. Future work

So far we have seen 3 servers using 2 different approaches to represent the result. That is not a problem for a web service (that only outputs data) but it is not the best situation to ensure interoperability at the client side where an integrated client will need to react to any possible encoding variation and deliver the best result.

5.4.1. How to encode the procedure.

The SWE4CS discussion paper suggest that we use an approach to encode the procedure that takes into account a recommendation extracted from section 6.18.1 of Timeseries Profile of Observations and Measurements standard (OGC 15-042r5) that suggests an encoding for both the observation *process* and the *operator* of the sensor (the *citizen* doing Citizen Science) that is based on ISO metadata. This will ensure an uniform way to report on this two important aspects of the observation.

NOTE	This approach has not been implemented during the interoperability experiment but it is considered something we can experiment with in the future. An example of this procedure is provided in the SWE3CS document and reproduced here for convenience

Example of encoding of the procedure (including process and operator) extracted from the SWE4CS document

```
<om:procedure>
  <tsml:ObservationProcess gml:id="op1">
    <!-- processType defines observation performed by human with sensor -->
    <tsml:processType
xlink:href="http://www.opengis.net/def/waterml/2.0/processType/Sensor"/>
    <!-- processReference defines sampling protocol -->
    <tsml:processReference
xlink:href="https://dyfi.cobwebproject.eu/skos/JapaneseKnotweedSamplingProtoc
ol"/>
    <!-- if a sensor is used, provide the link to the sensor definition here.
Use
SensorML if possible -->
    <tsml:parameter>
      <om:NamedValue>
        <om:name xlink:href=
"http://www.opengis.net/def/property/OGC/0/SensorType"/>
        <om:value>http://www.motorola.com/XT1068</om:value>
      </om:NamedValue>
    </tsml:parameter>
    <!-- operator defines the citizen scientist producing this observation
-->
    <tsml:operator>
      <gmd:CI_ResponsibleParty>
        <gmd:individualName>
          <gco:CharacterString>Ingo Simonis</gco:CharacterString>
        </gmd:individualName>
        <gmd:organisationName>
          <gco:CharacterString>OGC</gco:CharacterString>
        </gmd:organisationName>
        <gmd:role>
          <gmd:CI_RoleCode
    codeList=
"http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
    codeListValue="resourceProvider"/>
        </gmd:role>
      </gmd:CI_ResponsibleParty>
    </tsml:operator>
  </tsml:ObservationProcess>
</om:procedure>
```

The result is quite verbose what might affect performance when many data is transmitted.

5.4.2. Avoiding verbosity by defining a data streem

An approach based on providing a coma separated recordset that is described only once at the beginning should be more compact and efficient to parse.

The section 8.4.3 of the OGC 08-094r1, OGC® SWE Common Data Model Encoding Standard v2.0 describes a way to encode a *DataStream* only once and then send the data directly as a CSV format using HTTP of other protocol. A similar solution could be worth to be tested in the future to increase performance.

Chapter 6. SOS architectures

In this chapter we describe three architectures tested in the Interoperability Experiment that demonstrate end to end architectures as well as interoperability among servers and clients.

6.1. Architecture 1: SOS services integrated in a SOS client

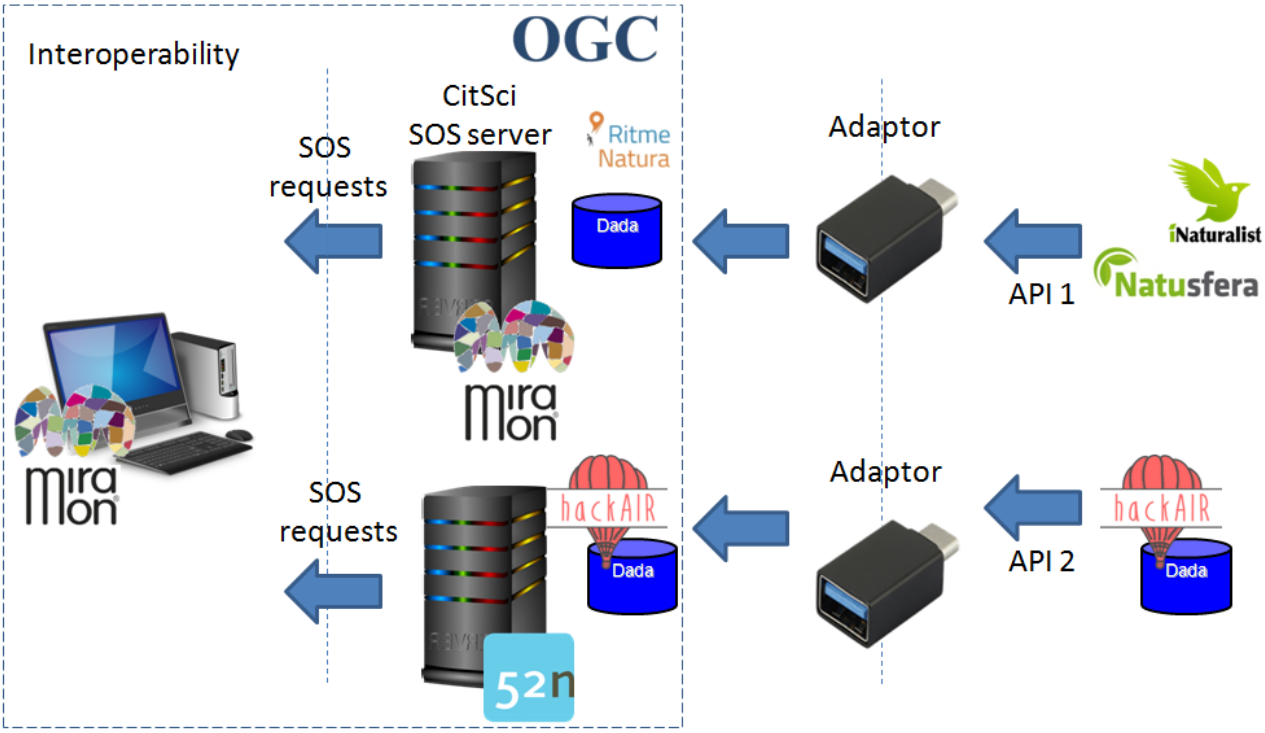


Figure 3. MiraMon SOS Architechture

In this architecture ([MiraMonSOSArchit]) the client access directly two different SOS services. It formulates a GetFeatureOfInterest to determine the positions of the individual observations and a GetObservation each time it needed to show a complete description of a single point (the user triggers this event by clicking on an icon) or if it needs to represent different icons as a function of the value of the observation. In this case, interoperability happens directly in the client. Since the SOS requests are communicated to the Internet, this client is exposing requests and response, allowing people to explore the SOS protocol with both the map browser console as well as the browser developing tools.

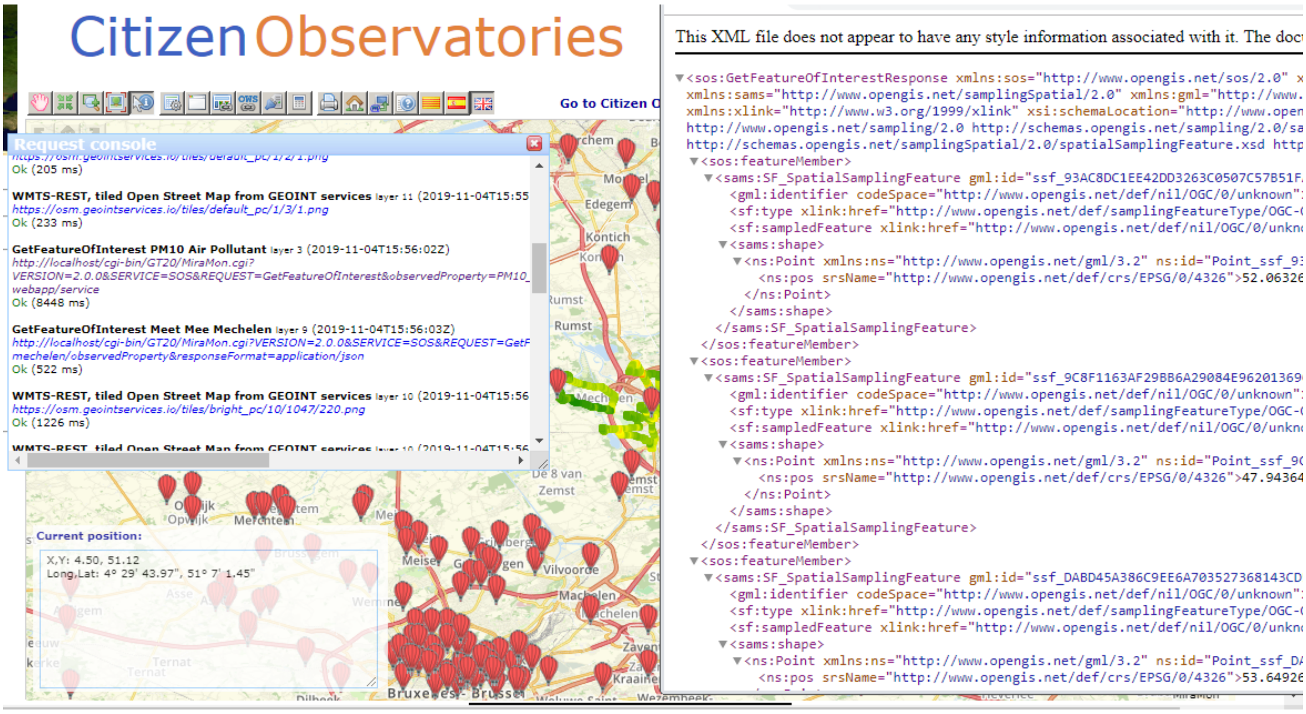


Figure 4. MiraMon Map Browser console with both server requests

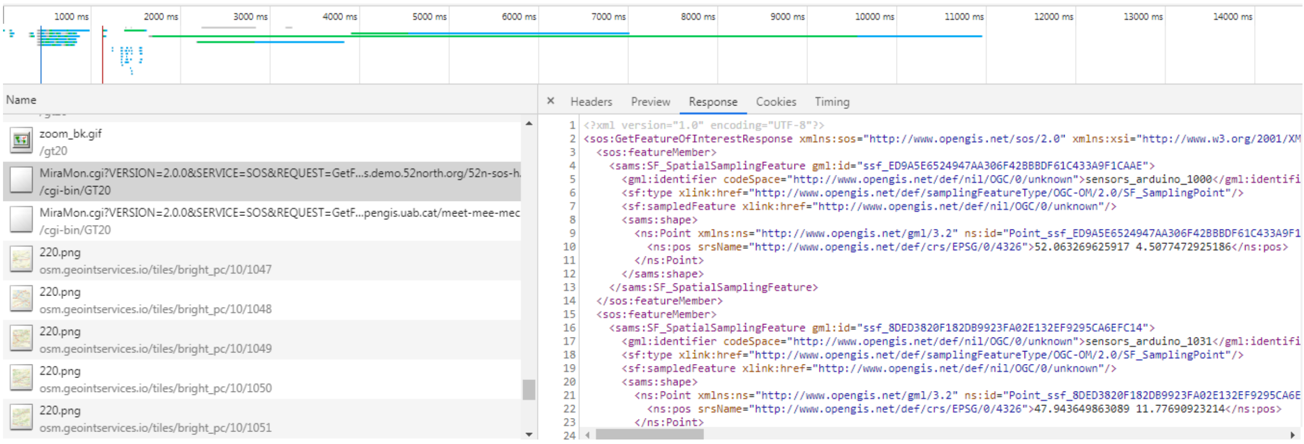


Figure 5. SOS requests as shown in the Chrome development tool

It is worth mentioning that this architecture is only possible if both services are declaring their willingness to be combined in the header of the responses. By default, programatically reading XML or JSON data coming from a Internet domain different from the client itself is not allowed except if the server states in the header that this is allowed. This is known as Cross-Origin Resource Sharing (CORS). The following headers will allow CORS with anybody.

```
Access-Control-Allow-Origin: *
Access-Control-Allow-Methods: POST, GET, OPTIONS, DELETE
```

In the case a SOS server does not allow CORS, our client is still able to force a solution by redirecting the request to our server with an extra parameter *ServerToRequest*. In this case, our server will cascade the request to the specified server and return the response back to the client as if there was only one server involved in only one domain.

6.2. Architecture 2: SOS services integrated in a combined agile service

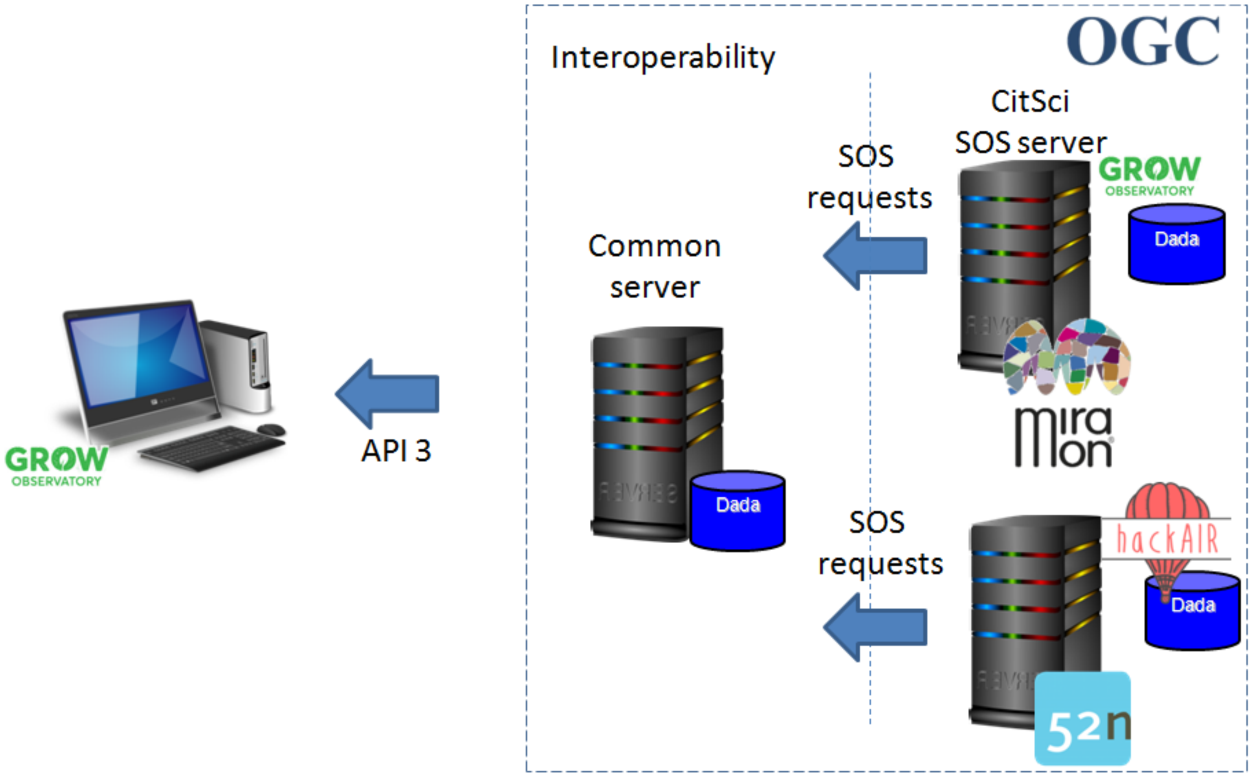


Figure 6. Grow SOS Architechture

In this architecture ([GrowSOSArchit]) a common server pulls two or more SOS requests into a central tabular datastore. This datastore records only the information from the returned SOS data that is required for the final visualisation and removes information that is redundant creating a data warehouse representing only one version of the data. In this approach the interoperability happens internally in the datastore and the SOS requests and responses are not explosed to the final client. In the diagram below ([GRowDataFlow]), data is stored in the datawarehouse and Microsoft's PowerBI does the heavy lifting for the visualisation of the combined data sources.

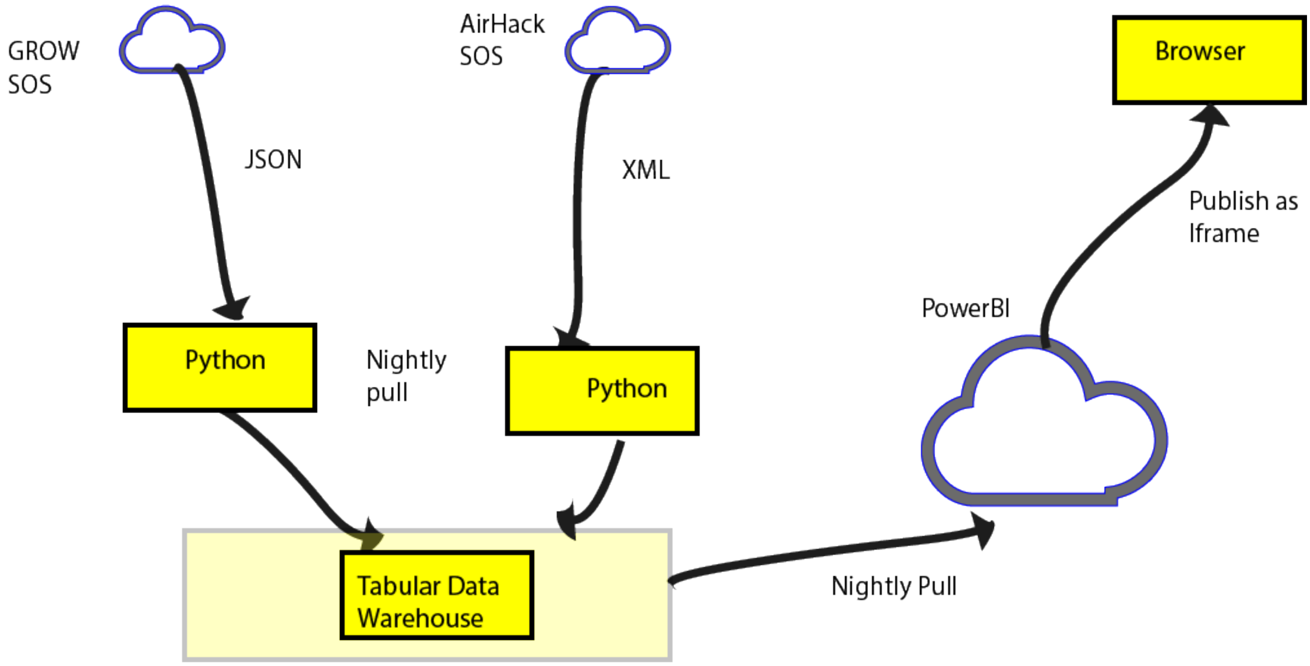


Figure 7. Grow SOS Data Flow

The common server would typically be a cloud server, but for some clients this is not necessary, in the case of PowerBI a data bridge is created between the data source and the visualisation tool before it is published to a web client. Other visualisation tools (such as Tableau) will have their own methods of connecting to the data warehouse and publishing the results to a web based client. In this architecture, data is only as up to date as the latest data pull from the SOS servers; in the case of GROW this is done nightly but this could be made more frequent or moved towards real time using a data log pipeline in a kappa architecture.

6.3. Architecture 3: SOS service for interoperability and JSON API for fast client

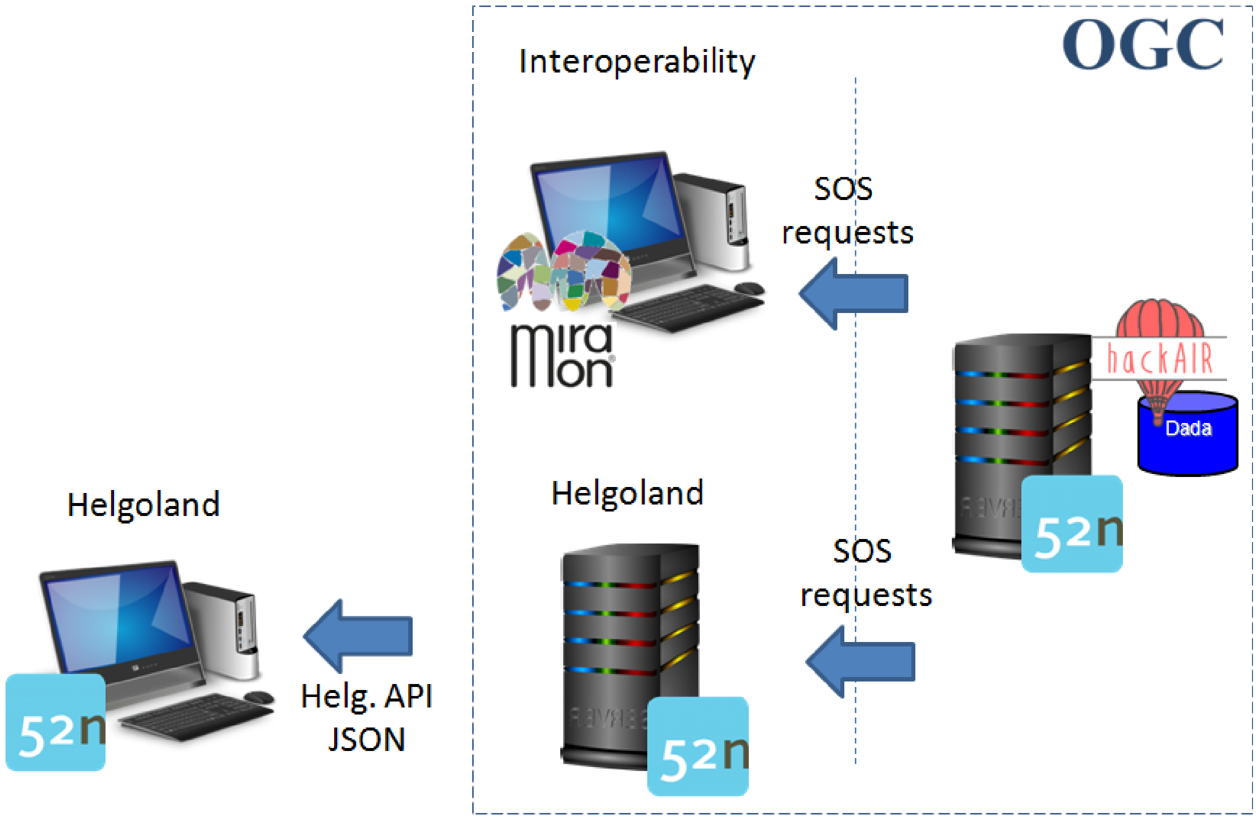


Figure 8. Helgoland SOS Architechture

This architecture was especially optimised to support the development of lightweight Sensor Web applications. This is achieved by avoiding the direct XML encoding/decoding on the client device. Instead, the interactions between the client (in this case the 52°North Helgoland Sensor Web Viewer) and the server components is achieved via a REST and JSON interface (the 52°North Sensor Web API).

This API can be directly exposed by Sensor Web servers such as the 52°North implementation. Alternatively, an available proxy component is also capable to encapsulate existing OGC SOS servers behind the lightweight interface of the 52°North Sensor Web API.

The advantage of this approach is a more lightweight communication pattern to be implemented on the client side. In addition the 52°North Sensor Web API offers further convenience methods as well as functionalities for reducing the transferred data volume (by generalising observation data) and improving the data visualisation (e.g. providing rending hints). A drawback of this approach is a less direct interaction with SOS servers, so that for integrating new SOS servers, a proxy component has to be configured/adjusted.

Chapter 7. SOS servers

In this chapter we describe four SOS servers tested in the Interoperability Experiment.

7.1. 52 North solution

The 52°North Sensor Web Server comprises several server-side modules which closely interact to provide different kinds data access functionality. In detail, this comprises the following elements:

- Data storage: The database for storing the observation data is integrated through an object-relational mapping layer based on the Hibernate framework. This allows the flexible integration of different types of database management systems (e.g. PostgreSQL, Oracle, MS SQL Server, MySQL) and data models. For this Interoperability Experiment, PostgreSQL was used.
- For the access to observation data, the server offers three dedicated modules, which use the same common Sensor Web database.
    - SOS: The SOS module offers a comprehensive implementation of the OGC Sensor Observation Service 2.0 standard (including beyond the core several extended functionalities, transactional and result handling operations). It also offers several interoperability enhancements such as a support of the INSPIRE Technical Guidance on the SOS as a Download Service.
    - SensorThings API: In addition to the SOS support, a dedicated module is available for supporting the OGC SensorThings API Part 1: Sensing (not yet evaluated as part of this Interoperability Experiment).
    - 52°North Sensor Web API: Complementary to the previous modules, also the 52°North Sensor Web API is offered. This API offers an additional but optional convenience layer for building client applications. While both the SOS and the SensorThings API standards are well suited for enabling the interoperable access to observation data, the Sensor Web REST-API allows to provide additional functionality that significantly facilitates the development of client applications. Typical examples of this additional functionality comprise: generalisation of observation data (important for developing mobile applications), provision of rendering hints (e.g. styling information for time series), conversion of data to mainstream formats such as CSV.

7.2. istSOS

Easily manage your sensor network and distribute your data in a standard way http://artemis.geogr.uni-jena.de/istsos_ie/soil?service=SOS&request=GetCapabilities&version=1.0.0

7.3. MiraMon SOS

7.4. GROW SOS

Chapter 8. SOS clients

In this chapter we describe three SOS integrated clients tested in the Interoperability Experiment.

8.1. Grow Client

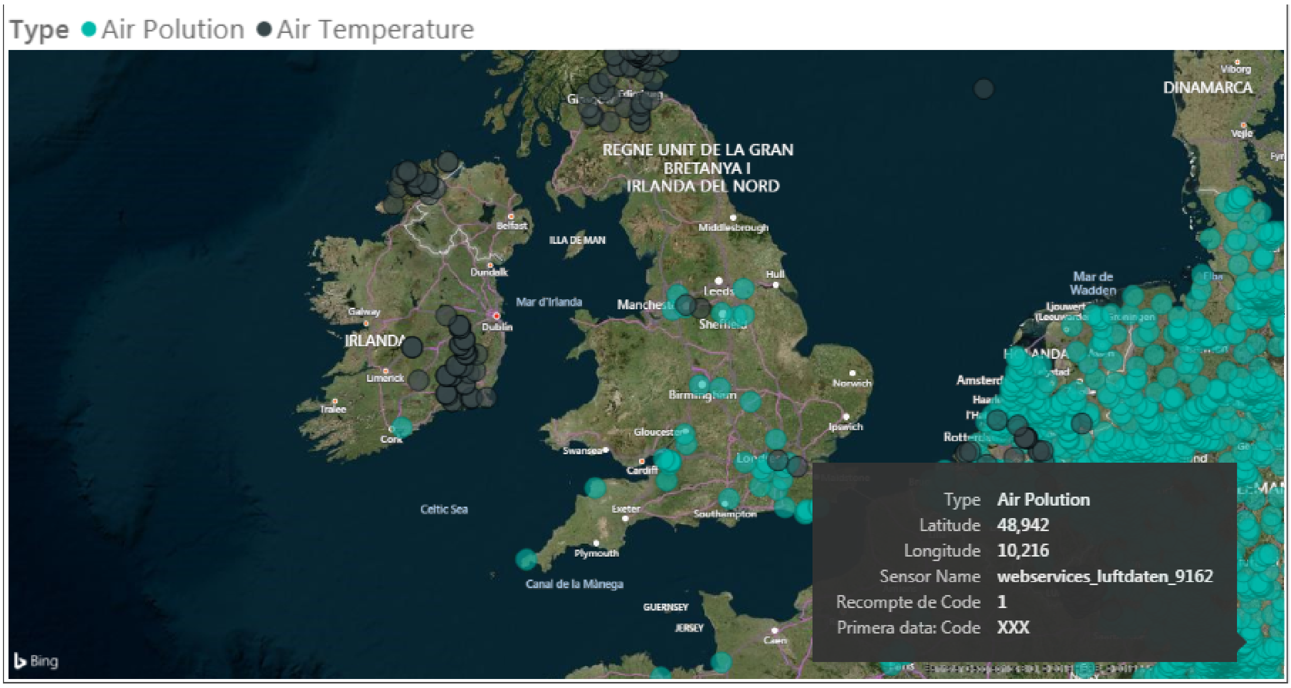


Figure 9. Andy SOS Client

8.2. MiraMon Client

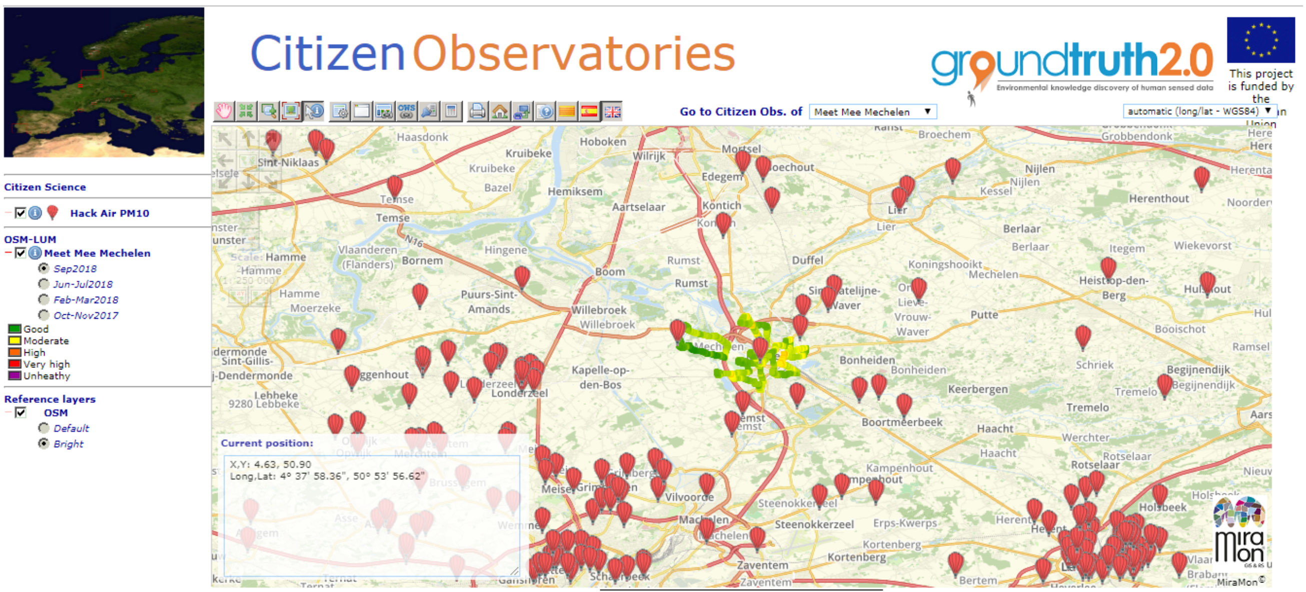


Figure 10. MiraMon SOS Client

8.3. 52°North Helgoland Sensor Web Viewer

The Helgoland Sensor Web Viewer developed by 52°North is an open source visualisation tool for different kinds of Sensor Web data. It allows to explore available observation data sets and to visualise the actual data (e.g. time series) as diagrams.

Figure 1 shows the map view of the Helgoland Sensor Web Viewer. In this case, a sample data set of the hackAIR project is explored. The map view shows the locations at which air quality sensors are located.

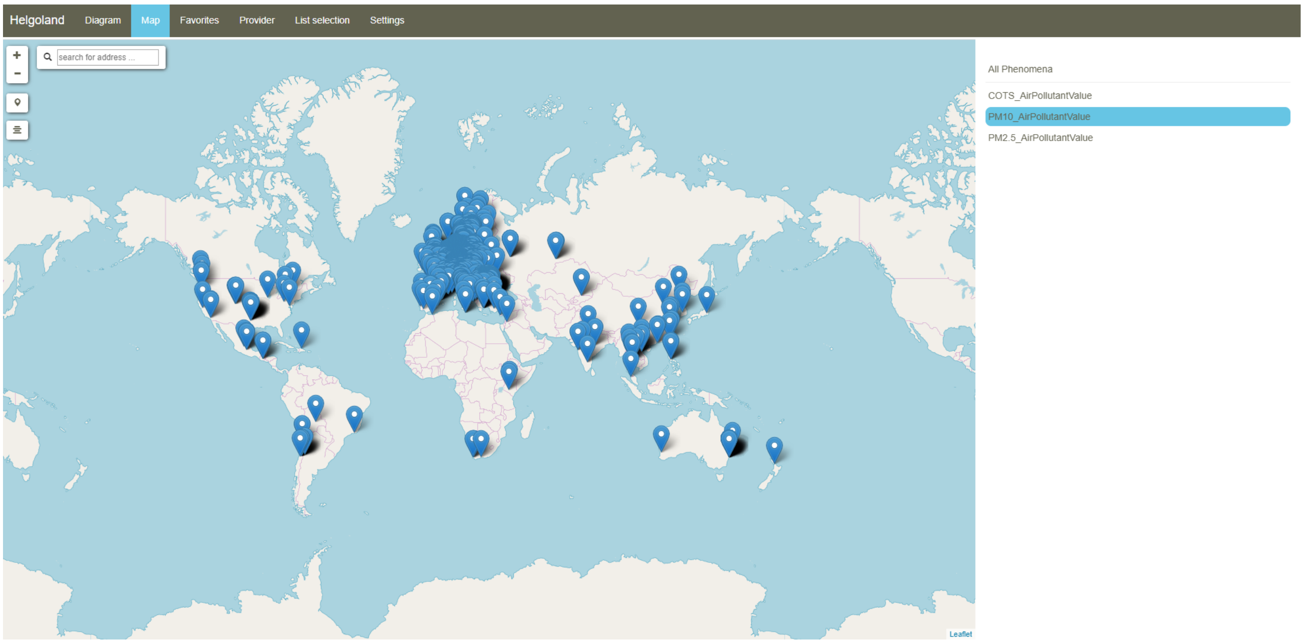


Figure 11. 52°North Helgoland Sensor Web Viewer - Map View

After selecting a specific measurement location, the data can be visualised as a diagram (see Figure 2). It is possible to combine data from multiple sensors, multiple observed properties and even from different providers into a single diagram.

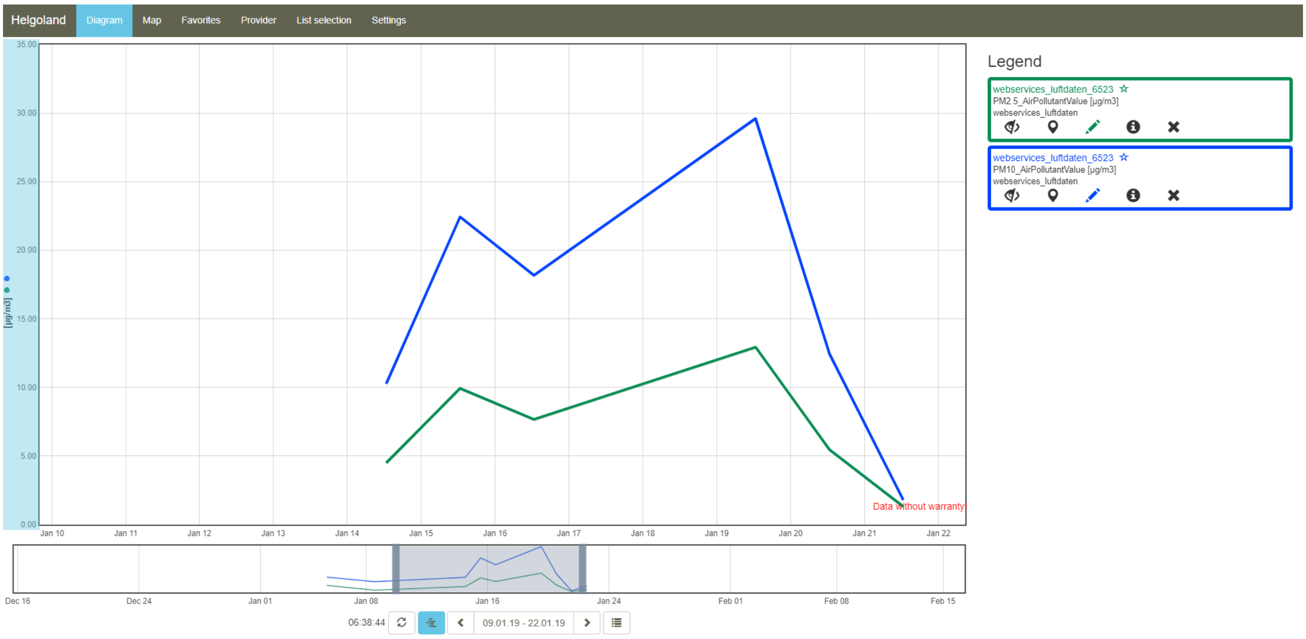


Figure 12. 52°North Helgoland Sensor Web Viewer - Diagram View

8.4. SOS Technology Integration Experiments

During the IE a set of Technology Integration Experiments (TIE) were conducted. In the client and server architecture a TIE is test that combines a server with a client and demonstrates that communication between client and server is possible and the user (operating the client) is able to see or get some data. The following table summarizes the tests conducted and the degree o success achieved.

Servers	Data	MiraMon Client	Helgoland Client	Grow Client
MiraMon	Ground Truth 2.0	Yes		
Helgoland	HackAir		Yes	
istSOS	HackAir	Yes		Yes
Grow	Grow			Yes

Table 2. Technology Integration Experiment results

Chapter 9. Data quality estimations with SOS

TBD

Chapter 10. Definition Server

TBD: Waiting for writing materials.

Chapter 11. User authentication

TBD. This are just some texts from and email discussion for inspiration.

We are aware of Gilberto's views but as Europeans we do have an issue with having in-situ European data or citizen generated data hosted on commercial platforms that do not adhere to our data protection regulation. More specifically, given the business model of these companies that is based on collecting, and integrating personal data of their users to then sell to third parties personalised databases and services for commercial or political targetting, I question the appropriateness of putting citizen science data on these platforms, thus opening the door to their use of the personal data of those who have collected the data and those who use them, without their consent. We are raising this issues within the GEO community as well because there has been no discussion so far on this point. I understand your perspective may differ, but at least you are informed about our views.

This argumentation points directly at a real problem that should be included in an GEO architecture for Citizen Science that needs to provide the necessary anonymity to the Citizens when participating in the Citizen Science activities. The H2020 Landsense federation includes the concept of a criptoname (created when citizens authenticate) that is used by a Citizen Science infrastructure to allow participation but hiding the details of the user. This architecture will be applied in the new project CO4Cloud (I'm participant. Jaume Piera Leads) that will propose a Cit Sci infrastructure on the EOSC for citizen Science.

Good point Joan. The identity management etc. is one of the important aspects. Yet, the issue how hosts, manages and own data is complementary to it. Both require a solid architectural set-up.

I agree that the need to ensure appropriate privacy of contributors is essential but there also needs to be a way of acknowledging those Citizen Science participants that may wish to be attributed. I suspect that it may need to be an opt in process.

Chapter 12. Connecting Citizen Science data sets to GEOSS

In addition to the developments outlined above, this Interoperability Experiment also examined possibilities for the Citizen Science community to make their projects discoverable and accessible via the Global Earth Observation System of Systems (GEOSS).

NOTE The worldwide effort to build GEOSS is led by the Group on Earth Observations (GEO). GEO is an intergovernmental organization working to improve the availability, access and use of Earth observations for the benefit of society. GEO works to actively improve and coordinate global Earth Observation systems and promote broad, open data sharing.

We see a major benefit in establishing this connection because GEOSS already provides an established process to clarify data policies together with established data management principles, and to provide the minimum metadata this is required to access the data sets created by Citizen Science projects. This activity thereby helps to surface and to mobilize already existing data sets - with clear acknowledgment Citizen Science contributions. We expect that this work provides concrete contributions to the increasing discussions on how Citizen Science could connect to GEOSS and, more generally, how more in-situ data and derived knowledge could become available on the global level. Within this Interoperability Experiment, we particularly explored the possible technical connection with the GEOSS Platform by facilitating OGC standards. We also identified organizational structures that would be required to be put into place in order to provide a more flexible and scalable solution to Citizen Science projects. We consider this as a major need for the future evolution of this connection because of the high number of already existing projects that could potentially be connected to GEOSS. OGC standards can play an essential role to facilitate this connection. This elaboration would also contribute to ongoing debates on increasing the availability of in-situ data in GEOSS. Notably, we see the connection with GEOSS as one highly promising way to make Citizen Science data better accessible and more widely used. Complementary efforts might be undertaken, for example, to increase the findability via mainstream search engines, or – more generally – to provide the machine readable information about projects resources that is required for automated harversting by web-crawlers. We come back to this issue in the final part of this section, when outlining possible follow up activities. With our exercise so far, we follow the following overarching principles and carry them to the Citizen Science community:

- GEO Data Management Principles [https://www.earthobservations.org/documents/dswg/201504_data_management_principles_long_final.pdf], addressing issues such as discovery, access, traceability, quality documentation and preservation.
- Not favoring any silos, i.e. necessitating open data and leveraging open solutions, which in this context means the provision of data as part of the GEOSS Data Core [https://www.earthobservations.org/geoss_dsp.shtml], but also the application of the GEO Architecture Principles, which in addition advocate flexibility, scalability, etc.
- Distributed, standards based, and flexible to support interoperability while meeting a range

of use cases and needs, which translates to the use of the GEOSS Discovery and Access Broker (GEO DAB) [https://www.geodab.net/] that includes – among other - the support to a large range of OGC standards.

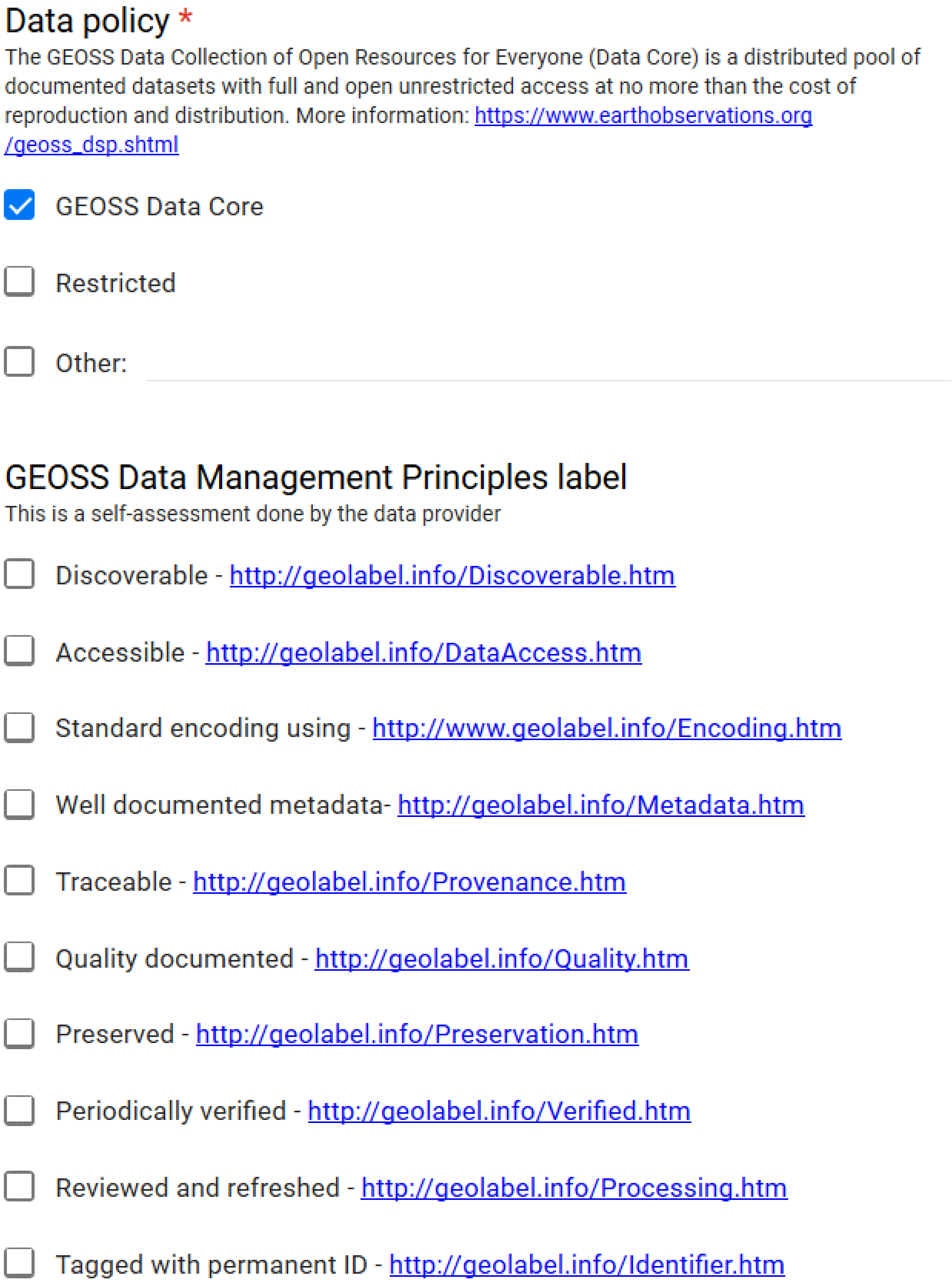
12.1. Connecting a single Citizen Science project to GEOSS

The most straight forward way to connect Citizen Science data sets to GEOSS is the registration of the project that collects these data sets at the GEOSS Platform. This process essentially requires the registration of the project in the GEOSS 'Yellow Pages' and it can be used to register also multiple data sets from one single project (see also Figure below).

Figure 13. Sketch on registering a single project at the GEOSS Platform

The entry of the Yellow Pages requires information (metadata) such as:

- Information about the lead organization (name, description, URL, geographical coverage, GEO affiliation, contact points, etc.)
- Type of provided online resource (data, vocabulary, model, algorithm, etc.) – in the context of this Interoperability Experiment we focus on data
- Data policy, including the option to declare as free and open by choosing "GEOSS Data Core" (see also the figure below)
- Indication of GEOSS data management principles that are implemented (see also the figure below)
- Relevance for the SDGs
- A service endpoint



Data policy *

The GEOSS Data Collection of Open Resources for Everyone (Data Core) is a distributed pool of documented datasets with full and open unrestricted access at no more than the cost of reproduction and distribution. More information: https://www.earthobservations.org/geoss_dsp.shtml

- [x] GEOSS Data Core
- [ ] Restricted
- [ ] Other:

GEOSS Data Management Principles label

This is a self-assessment done by the data provider

- [ ] Discoverable - http://geolabel.info/Discoverable.htm
- [ ] Accessible - http://geolabel.info/DataAccess.htm
- [ ] Standard encoding using - http://www.geolabel.info/Encoding.htm
- [ ] Well documented metadata- http://geolabel.info/Metadata.htm
- [ ] Traceable - http://geolabel.info/Provenance.htm
- [ ] Quality documented - http://geolabel.info/Quality.htm
- [ ] Preserved - http://geolabel.info/Preservation.htm
- [ ] Periodically verified - http://geolabel.info/Verified.htm
- [ ] Reviewed and refreshed - http://geolabel.info/Processing.htm
- [ ] Tagged with permanent ID - http://geolabel.info/Identifier.htm

Figure 14. Parts of the items asked for when registering in the GEOSS Yellow Pages

Once compiled, the information provided by the Citizen Science project is passed on to the GEO DAB team. Members of this team check the entries and tests, for example, if the provided endpoint actually serves the intended data, if it needed implements the required standards (from the OGC or comparable alternatives) and if it indeed follows the indicated principles and applies the indicated data policy. This testing might entail a dialogue with the registering Citizen Science project in order to make the project's offer fit to the promises in the entry for the GEOSS Yellow Pages.

SCENT Citizen Observatory [https://scent-project.eu/] successfully undertook the process to offer its citizen-generated data to GEOSS. Following the administrative registration that was described in detail above, the Interoperability Registration and Brokering workflow took place. This was the second phase and consisted of the following steps:

- Technological Information was provided to the DAB team. More specifically the SCENT web server URL, the endpoints of the SCENT OGC web services (WMS, WFS) were sent to the DAB team along with the accompanied descriptions. In addition, information was provided about WFS and WMS services versions, the supported relevant operations through the APIs (e.g. GetCapabilities, DescribeFeatureType, GetFeature, etc.), the various feature types, the SCENT data to be integrated to GEOSS (i.e. event, images, video metadata) and the associated geographic regions that they cover (i.e. Kifisos river basin, Attica Greece and Danube Delta, Romania).
- DAB team conducted a set of interoperability tests with SCENT WFS and WMS services, including discoverability, accessibility and visualisation use cases. A test report (GEO DAB/SCENT web server Brokering Test Report) was generated as an output of this process.
- Notification was received from DAB team regarding the successful conduction of the tests and that no interoperability issue was found. Taking into consideration also suggestions indicated in the test report, metadata information of the datasets provided was further enriched.
- As a final step of the process, the DAB team proceeded with the successful integration of SCENT web server into GEOSS Portal (offering its services to production environment). Thus, users can access SCENT resources via the GEOSS Portal catalogue (https://www.geoportal.org/?f:sources=wfsscentID%2CwmsSCENTID) as presented in the following figure.

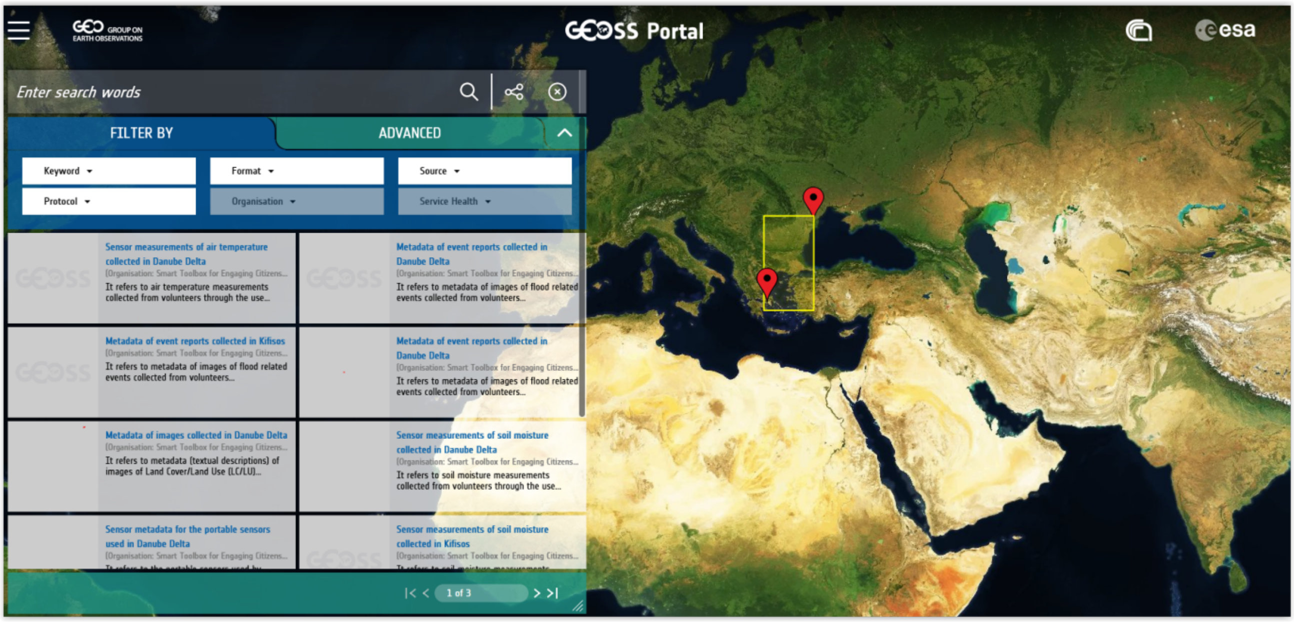


Figure 15. Scent resources available in GEOSS portal

Overall, we promote that Citizen Science projects indeed implement the free and open data policy of the GEOSS Data Core, and that all of the GEOSS Data Management Principles are implemented. Furthermore, out of the OGC Citizen Science Domain Working Group (DWG) we offer examples and guidance how this can be realized with OGC standards. However, realistically speaking, and in order to see concrete progress, it appears most feasible that projects first enter what they already support, and at least fulfill the needs to make their data sets discoverable from within the GEOSS Platform [http://geolabel.info/Discoverable.htm] using well-documented metadata [http://geolabel.info/Metadata.htm], and including information about data quality [http://geolabel.info/Quality.htm]. Following the brokering approach of GEOSS, this entails that the projects provide a minimum of required information, and ideally follow one of the multiple standards that are already supported by GEOSS. If it should not be immediately possible to also provide one of the multiple options to make data accessible [http://geolabel.info/DataAccess.htm] or to provide that data in an already recognized encoding [http://www.geolabel.info/Encoding.htm], it might still be considered to register a projects and then update the record in the Yellow Pages once the additional functionalities for harmonized data access are put in place by the project.

12.2. Why a case-by-case registration might not be the best way forward

Although the possibility to register Citizen Science e projects of any size in GEOSS exists, we do not recommend that every project should now go ahead and register itself. This is because, at least potentially, the number of relevant contributions is (at least) in the hundreds, see for example, this project inventory [http://data-staging.jrc.it/dataset/jrc-citsci-10004] and the related prototype of a project catalogue [https://ec-jrc.github.io/citsci-explorer/].

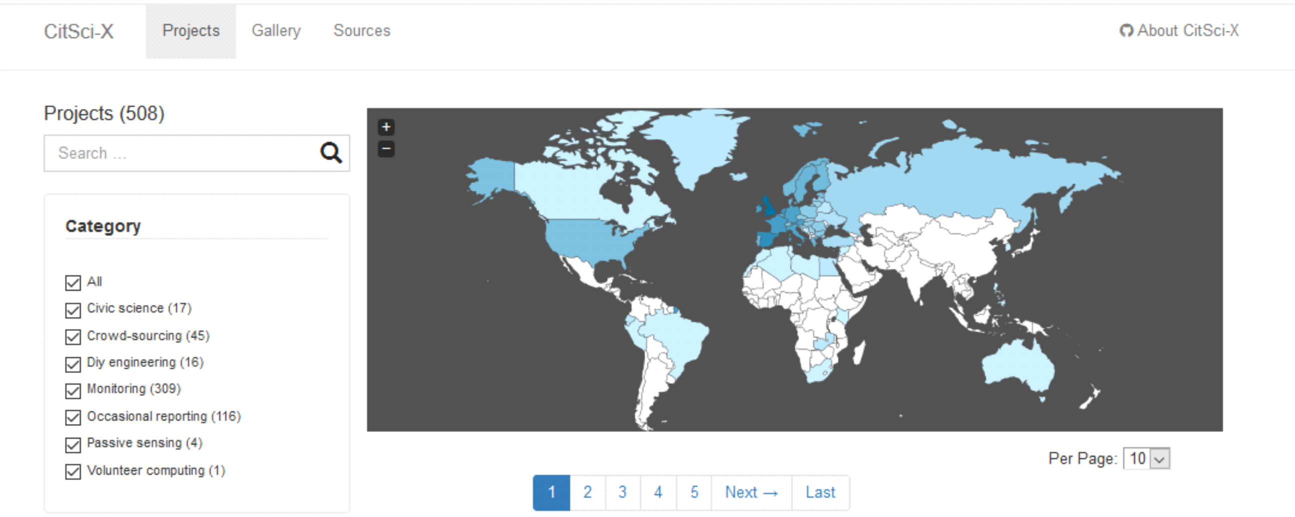


Figure 16. Example Citizen Science explorer providing access to 508 different projects (status: 13 June 2019)

It becomes obvious that a case-by-case registration per project (which each might want to register one data set or more) would create a bottle neck towards the GEO DAB and the team that is responsible to evaluate and test entries in the GEOSS Yellow Pages. As a result of our investigations within this Interoperability Experiment, we therefore suggest to elaborate on and develop an intermediate layer that provides the required organizational and technical support to the Citizen Science community so that their data sets become better discoverable, accessible, and potentially more widely used - thereby also amplifying visibility and impact of the individual projects. We consider such structures particularly important in view of larger mobilization campaigns of Citizen Science projects, as, for example, planned within the context of the Earth Challenge 2020 (EC2020). Again, also here the two/multiple-step approach - where project resources become discoverable first and commonly accessible in a second stage - might be most realistic in order to progress more quickly and to have intermediate results. In order to move ahead, we identified two main clusters of requirements. On the one hand, there appears a need to slot a technical support for Citizen Science projects in ahead the GEO DAB. This additional support should remove the potential bottle neck and help to scale up the number or Citizen Science projects and their data sets in the GEOSS Platform (and ideally in the GEOSS Data Core). Requirements for this support entail:

- Support Citizen Science projects in filling the GEOSS Yellow Pages.
- Proving examples and guidance on the use of OGC standards for implementing GEOSS requirements for data discovery, quality descriptions, data access, data encodings, etc.
- Pre-testing of yellow page entries before registration in GEOSS.
- If necessary, interaction with individual projects to correct their entries for the Yellow Pages.
- Liaise with the GEO DAB team in order to actually register the new entries.
- Possible consider an additional hub that would in fact cluster multiple Citizen Science projects and their endpoints into a single access point, which is then registered within the GEOSS Platform (see also example below).

- Establish a capacity building mechanism, capable to support and equip existing initiatives with the necessary skills to apply data management principles related to the accessibility, discoverability, re-usability and curation of their resources.

For the above mentioned EC 2020 we could, for example, imagine the following. EC2020 will collect new data and offer those via a dedicated API. At the same time, several already existing Citizen Science projects partner with EC2020 and also provide access to their data (in different forms). For the connection to GEOSS, EC2020 could provide a gateway that connects the newly collected data and the offerings of the different partners to a single discovery service and a single data access service. These two endpoints would be registered via the Yellow Pages with the GEOSS Platform and thereby make the EC2020 resources more widely visible, together with a clearly defined and well-known data policy and following most recent data management principles. The figure below depicts this setting.

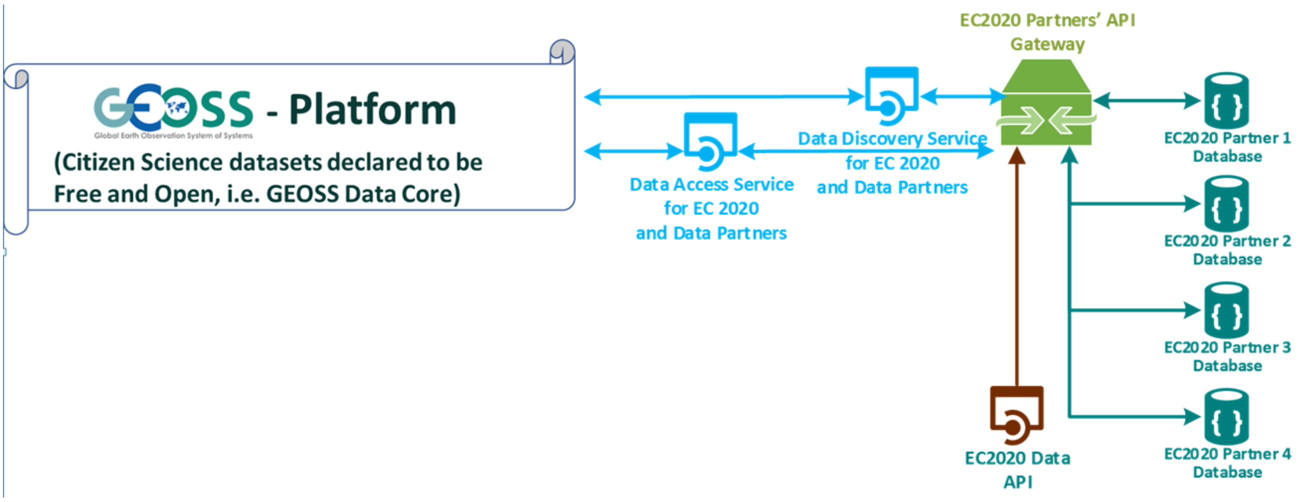


Figure 17. Sketch of a possible connection between EC2020 and the GEOSS Platform (one possible scenario), as an example of a more coordinated structure to make Citizen Science data discoverable and accessible via GEOSS

On the other hand, and because the offerings made above alone would not be enough to actually advance from the current situation, also the connection to the Citizen Science community deserves dedicated attention. We see, for example, requirements in: - Mobilizing existing data sets, i.e. reaching out to the Citizen Science community and let them know about this work and the linked offering of increased visibility and possibly impact, providing guideless and practical examples on what would need to be done from their side, offering support in establishing the connections.

- Help in preparing new data sets, i.e. be available to consult Citizen Science projects during their set-up phase, and let the community know about this offer.
- Promote FAIR data management and GEOSS as a practical way to get there.
- Provide access and training for (OGC) standard-based tool that the community can use to make the connection, and implement the desired data policy and data management principles.

For further discussions and possible realizations, it should be considered if the support outlined above could be provided in a coordinated but decentralized way. We could imagine that the above mentioned support could be set of for different geographic regions, thematic areas, or other sensible divisions (e.g. with a retaliation that is specific for EC2020, which would still need to be discussed). Such settings could also help to disseminate good practices, for example, on the use of OGC standards in this context.

12.3. Items for further investigation regarding GEOSS

The experiment helps us to identify current possibilities, and to shape parts of the way ahead. However, it also left a few questions unanswered and raised some new issues. From our experiences, we the particular needs to further investigate the following:

- Acknowledging that Citizen Science data is already included in GEOSS today, i.e. systematically flagging where Citizen Science already contributed to a knowledge resource on the GEOSS Platform (GEOSS Data Core, ideally).
- Develop detailed examples and guidance on how CS projects can implement the different GEO Data Management Principles by using the many already supported OGC standards.
- Consider promoting OGC' SensorThingsAPI to be considered by the GEO DAB, because this standard appears to be taken up by several Citizen Science projects, but so far other standards (such as WFS or WMS) need to be implemented in addition in order to allow harmonized access via the GEOSS Platform.
- Develop different scenarios to meet the identified organizational requirements.
- Consider Citizen Science not only as a data source, but also explore the possibilities and use of OCG standards when it comes to the engagement of Citizen Scientists as part of data validation.
- Also consider Citizen Science as part of the processing capacity, collective intelligence, data cubes, relationship to WPS and work on Artificial Intelligence, etc.
- While focusing on the connection to GEOSS here, we should also investigate how this work related to the provision of metadata for 'flat' online searches and the accessibility to automatic web crawlers. We might want to address both in a single go. If we will work towards intermediate organizational structures the help the Citizen Science community in using OGC standards and the GEOSS Platform for improved data policies and management, can these intermediaries – and the tools and services they provide – also automatically cover these complementary needs?

Appendix A: Abstract Test Suite

An Abstract Test Suite may be relevant to an Engineering Report.

An Abstract Test Suite is specified in Clause 9 and Annex A of ISO 19105. That Clause and Annex specify the ISO/TC 211 requirements for Abstract Test Suites. Examples of Abstract Test Suites are available in an annex of most ISO 191XX documents, one of the more useful is in ISO 19136. Note that this guidance may be more abstract than needed in an OGC® Implementation Standard.

Test identifier	/test/case/id
Test purpose:	Confirm that the IUT satisfies all applicable requirements for conformance level 1.
Test method:	Functional testing performed in an automated and/or manual manner. Verify the behaviour of the IUT for the following operations: • GetCapabilities (mandatory) • DescribeRecord (mandatory) • GetRecords (mandatory) • GetRecordById (mandatory) • GetRepositoryItem (mandatory) • GetDomain (optional)
Requirement:	OGC 07-110: cl. 2.2
Test type:	Capability

Table 3. A.1.1 Conformance level 1

Test identifier	http://www.opengis.net/spec/xxx/conf/WRS.General-ValidResponse
Test purpose:	The XML response entity is valid.
Test method:	Validate content of response entity against corresponding element declaration.
Requirement:	OGC 07-006r1: cl. 10.2.5.1, p. 118
Test type:	Capability

Table 4. A.1.2 Test case for validity of XML response entity

Test identifier	/test/case/id
Test purpose:	Confirm that the IUT satisfies all applicable requirements for conformance level 1.

Test method:	Functional testing performed in an automated and/or manual manner. Verify the behaviour of the IUT for the following operations: • GetCapabilities (mandatory) • DescribeRecord (mandatory) • GetRecords (mandatory) • GetRecordById (mandatory) • GetRepositoryItem (mandatory) • GetDomain (optional)
Requirement:	OGC 07-110: cl. 2.2
Test type:	Capability

Table 5. A.2.1 Conformance level 2

Test identifier	http://www.opengis.net/spec/xxx/conf/WRS.General-ValidResponse
Test purpose:	The XML response entity is valid.
Test method:	Validate content of response entity against corresponding element declaration.
Requirement:	OGC 07-006r1: cl. 10.2.5.1, p. 118
Test type:	Capability

Table 6. A.2.2 Test case for validity of XML response entity

Appendix B: XML Schema Documents

XML Schema Documents may be relevant to an Engineering Report.

The term “XML schema“ means all the XML schema parts having the same XML namespace, usually separated into multiple XML Schema Document files (with the file type “.xsd“. The XML schema parts in one XML namespace are usually separated into multiple XML Schema Documents to ease human understanding.

In addition to this document, this report includes several XML Schema Documents. These XML Schema Documents are bundled in a zip file with the present document.

The TBD abilities now specified in this document use TBD specified XML Schema Documents included in the zip file with this document. These XML Schema Documents combine the XML schema fragments listed in various subclauses of this document, eliminating duplications.

These XML Schema Documents roughly match the TBD UML packages described in Annex B, and are named:

```
TBD.xsd
TBD.xsd
```

These XML Schema Documents use and build on the OWS common XML Schema Documents specified [OGC 06-121r3], named:

```
ows19115subset.xsd
owsCommon.xsd
owsDataIdentification.xsd
owsExceptionReport.xsd
owsGetCapabilities.xsd
owsOperationsMetadata.xsd
owsServiceIdentification.xsd
owsServiceProvider.xsd
```

All these XML Schema Documents contain documentation of the meaning of each element and attribute, and this documentation shall be considered normative as specified in Subclause 11.6.3 of [OGC 06-121r9].

```
<ows:Operation name="GetCapabilities">
  <ows:DCP>
    <ows:HTTP>
      <ows:Post xlink:href="http://www.opengis.net/?">
        <ows:Constraint name="PostEncoding">
          <allowedValues>
            <ows:Value>SOAP</ows:Value>
          </ows:AllowedValues>
        </ows:Constraint>
      </ows:Post>
    </ows:HTTP>
  </ows:DCP>
</ows:Operation>
<ows:Operation name="GetTile">
  <ows:DCP>
    <ows:HTTP>
      <ows:Post xlink:href="http://www.opengis.net/?">
        <ows:Constraint name="PostEncoding">
          <ows:AllowedValues>
            <ows:Value>SOAP</ows:Value>
          </ows:AllowedValues>
        </ows:Constraint>
      </ows:Post>
    </ows:HTTP>
  </ows:DCP>
</ows:Operation>
```

Appendix C: UML model

A UML model may be relevant to an Engineering Report. This template thus includes this annex as the place for recording this UML model.

Instructions and guidelines on the usage of UML models are provided in OGC document OGC-121r9 [https://portal.opengeospatial.org/files/?artifact_id=38867].

Appendix D: Revision History

NOTE

Example History (Delete this note).

replace below entries as needed

Date	Editor	Release	Primary clauses modified	Descriptions
June 15, 2016	I. Simonis	.1	all	initial version
July 22, 2016	I. Simonis	.9	all	comments integrate
September 7, 2016	S. Simmons	1.0	various	preparation for publication
March 23, 2017	I. Simonis	2.0	all	template simplified
January 18, 2018	S. Serich	2.1	all	additional guidance to Editors; clean up headings in appendices

Table 7. Revision History

Appendix E: Bibliography

OGC 17-019 OGC Testbed-13 MapML ER