COST Action CA 15212 "Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe"

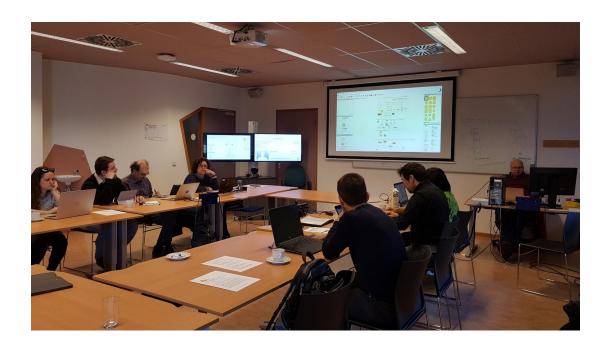# Minutes of WG5 workshop in Enschede: "On citizen-science ontology, standards and data"

Hosted by: ITC - University of Twente, and Earthwatch

Enschede, Netherlands

March 21-22, 2019

1. Attendees

Luigi Ceccaroni, Earthwatch (chair)

Rob Lemmens, ITC - Faculty of Geo-Information Science and Earth Observation - University of Twente

Gilles Falquet, University of Geneva

Vyron Antoniou, Hellenic Army Geographic Directorate

Chrisa (Chrysi) Tsinaraki, European Commission, Joint Research Center

Jakub Trojan, Tomas Bata University in Zlin, Czech Republic

Lucy Bastin, European Commission, Joint Research Center (remote)

Friederike Klan, DLR

Luis Felipe Velásquez, Earthwatch

Kalterina Shulla, RCE, Albania

Frank Ostermann, ITC, University of Twente, Netherlands

Aji Putra Perdana, ITC, University of Twente, Netherlands

Sven Shade, European Commission, Joint Research Center (remote)

Valentijn Venus, University of Twente, Netherlands

2. Acronyms

DMM:Dataset Metadata Model
PMM: Project Metadata Model

3. Background

The main aim of COST Action CA15212 "**Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe**" (2016-2020) [https://www.cs-eu.net/] is to bundle capacities across Europe to investigate and extend the impact of the scientific, educational, policy, and civic outcomes of citizen science with the stakeholders from all sectors concerned (e.g., policy makers, social innovators, citizens, cultural organizations, researchers, charities and NGOs), to gauge the potential of citizen science as enabler of social innovation and socio-ecological transition.

One goal of the COST Action is to help create an ontology (including a vocabulary) for describing citizen-science projects, observations and analyses, building upon prior research and existing standards, which any organization can model their database structure upon. This goal is also linked to the larger objectives of the international Data and Metadata Working Group of the Citizen Science Association (CSA) and the European Citizen Science Association (ECSA).

The WG aims to better define PPSR-Core and publish it as a standard for citizen science, reusing existing standards as much as possible. As a result of our joint discussion in Geneva in 2018, we (the members of COST Action 15212's WG 5 and of the CSA International Data and Metadata Working Group) agreed to change the name of this model to *PPSR - Common Conceptual Model* (CCM) - hereafter also the "model" - and to re-name its three main schemas: *Project Metadata Model* (PMM), *Dataset Metadata Model* (DMM) and *Observation Data Model* (ODM). Partners/collaborators in this effort are, among others: DataONE, CSIOR, ECSA, ACSA, CSA, and COST Action 15212's WG 5. Standards to be leverages are, among others: EML, FGDC, OGC related. The **Project Metadata Model (PMM)** will facilitate the interoperability between platforms, such as Scistarter, citsci,org, and the Wilson Center catalogue. It describes the project itself, giving the organization and project-level context for the data. We aim to retrofit existing functionality and carry out internal mappings to the locked-down published standard. The **Dataset Metadata Model (DMM)** allows for information to be consistently presented for fitness and for reusability of data at the dataset level. It describes the contextual background for the collection of data records, e.g., methods. The **Observation Data Model (ODM)** describes the actual data.

With respect to data quality we consider the following facets: accuracy, precision, completeness, consistency, validity, timeliness. The aim is to represent the "fitness for use" of citizen-science data, and to understand what is needed to improve data credibility.

4. Objectives

The general objectives of Working group 5 "**Improve data standardization and interoperability**" are:

- to explore ways for integrating data and knowledge related to citizen-science initiatives and suggest mechanisms for standardization, interoperability, and quality control;
- to improve the technical foundations to foster the impact of citizen science globally.

WG5's specific objective for the third period (1.5.2018-30.4.2019) is to contribute to develop an ontology of citizen-science projects (including a vocabulary of concepts and metadata) to support data sharing among citizen-science projects. WG5 will coordinate with activities on data and service interoperability carried out in Europe, Australia and the USA (e.g., the CSA's international Data and Metadata Working Group [http://citizenscience.org/association/about/working-groups/]), and will take into account existing standards, namely *Open Geospatial Consortium* (OGC) standards (via the OGC Domain Working Group on Citizen Science), ISO/TC 211, W3C standards (semantic sensor network/Linked Data), and existing GEO/GEOSS semantic interoperability. WG5 will investigate the best format to publish the ontology.

By convening COST Action's members with expertise in data and metadata representation, WG5 aims to:
1. refine core requirements of the ontology based on existing use cases and the discussion happening in the corresponding Basecamp project [https://basecamp.com/2071195/projects/13342949] (access restricted to WG5 members);
2. continue to identify core ontology fields and associated metadata by drawing on previous research and existing vocabularies, in particular the Report from CSA 2017 and Future Outlook by the international Citizen Science Association Data & Metadata Working Group: [https://www.wilsoncenter.org/article/citizen-science-association-data-metadata-working-group-report-csa-2017-and-future-outlook];
3. develop a plan for the uptake of the ontology by the citizen-science community at large.

5-6. Organizers

The organizers of the meeting are: Rob Lemmens (ITC - University of Twente) and Luigi Ceccaroni (Earthwatch, chair of the WG).

ITC - University of Twente [www.itc.nl]
The Faculty of Geo-Information Science and Earth Observation (ITC) of the University of Twente (UT) provides international postgraduate education, research and project services in the field of geo-information science and earth observation using remote sensing and GIS. The aim of ITC's activities is the international exchange of knowledge, focusing on capacity building and institutional development in developing countries and emerging economies. ITC has established an excellent track record in innovating and utilizing remote sensing for spatial data collection and spatio-temporal analysis in a wide variety of challenges in the development, planning and monitoring of resources. ITC is well-connected to international networks such as GEO and the Copernicus programme. Many of ITC's activities aim at the development of educational programmes for a wide range of subjects, such as climate change, population growth, and related claims for sufficient and secure food, water, energy, health, land and housing provision. In its core mission, ITC provides training and capacity building, and has built up a strong portfolio of activities in which curriculum development has played a central role. In a continuous effort to keep curricula alive and up-to-date, ITC embarks upon educational innovation, involving student-driven approaches, web-technology and fine-tuned blended learning. ITC has developed its own GIS-RS core book which has been migrated into a Wiki platform with an ontology-based approach. Dedicated training courses are offered in many countries on a regular basis in close cooperation with partner institutions. ITC has a wide network of partner institutions (universities and others) around the globe with which joint degree programmes as well as tailor-made training courses are being offered. Activities specifically focus on the use of GIS and remote sensing techniques and their applications. In many cases this is done making use of the knowledge and expertise that is available in our extensive alumni network.

Earthwatch [https://earthwatch.org.uk/] is one of the largest global backers of citizen-science--supported environmental research. For over forty years, Earthwatch has delivered a unique citizen-science model to empower individuals, students, teachers and corporate fellows to contribute to critical field-research to understand nature's response to change. Earthwatch's work supports hundreds of Ph.D. researchers across dozens of countries, conducting over 100,000 hours of research annually. Earthwatch's mission is to engage people worldwide in scientific field-research and education to promote the understanding and action necessary for a sustainable environment. To achieve its mission, Earthwatch works with multiple

partners, universities, schools, businesses and other environmental organisations, to develop robust and impactful 'citizen science' programmes that contribute to valuable environmental challenges whilst providing an engaging experience for individuals that builds commitment for environmental action.

7. Agenda

Wed 20 March - Arrivals - Informal gathering for those who want.

Thu 21 March 9:00-17:00 Meeting

09:00 - 09:10	Welcome and logistics
09:10 - 09:40	Luigi reports on CSA conference (CitSci2019)
09:40 - 10:30	Other short presentations/announcements/reports on events

1. Short presentation on Living Textbook concept map visualisation (Rob)
2. Working priorities of JRC (Chrisa)
3. Semantic annotation of Citizen Science data while they are collected (Friederike)

10:30 - 10:45	Coffee break
10:45 - 12:30	Work on action points for:

1. **International acceptance**
   a. Relation with the CSA WG
   b. Relation with the WeObserve Interop CoP & OGC CitSciIE
2. **Implementations of our model and application tooling**
3. **Data quality management practices (Peter Brenton document: [Click here](#))**

12:30 - 13:30	Lunch
13:30 - 16:40	**Ontology development and engineering.** **(Latest document on WG5 ontology: [Click here](#))** Formalising more modules of the ontology (work on the different 'layers' (project, data set and data)).

1. Round 1 (3 groups: project (lead: Luigi), data set (lead: Chrisa) and data (lead: Rob/Gilles))
2. Round 2 (3 groups: project (lead: Luigi), data set (lead: Chrisa) and data (lead: Rob/Gilles))

Coffee break around 15:30

16:40 - 17:00	Wrap-up (Luigi) and agenda for the next day (Rob)
19:00 - 21:00	Dinner (Everybody pays for themselves.)

Fri 22 March 9:00-15:00 Meeting

09:20 - 09:50	Create a document with allowed terminology.
09:50 - 10:20	Refine SDG profile
10:20-10:30	Data quality feedback
10:30-10:45	Coffee
11:00-12:00	Define a choice of use cases, e.g. from other COST WGs, or from outside the COST action, e.g. Earth Challenge 2020
12:00-12:30	International uptake
12:30-13:30	Lunch
13:30- 14:30	Follow up actions

8. Notes

8.1 Introductory presentations

Luigi Ceccaroni introduces the goals of WG5 and their setting. He explains that similar ontology development efforts exist at the international level with special contribution from the USA and Australia. In Europe ontology development is mainly covered within WG5 of COST Action 15212 and OGC. It is important to define the stakeholders who might be interested in the ontology. Different overlays or views should be specified for different stakeholders; e.g., for small projects, it is not necessary to communicate or use the whole ontology of ~200 concepts.

The purpose of this WG5 workshop is to progress in the definition of an ontology for citizen-science projects, based on interoperable standards and on the report released by the CSA Data and Metadata Working Group dealing with PPSR - Public Participation in Scientific Research. The ontology should be then tested against a number of different use cases to test its reliability. This task will be carried out after the ontology has been finalized.

So far, both in COST and in PPSR-Core, we looked at projects and datasets, and held off on observations per se because we assumed that models and ontologies would be emerging from OGC's work on an O&M profile or from other communities. We will make sure we are completely up to date with (and re-use, where appropriate) proposed observation models because so many of the higher-level entities are aggregations of observation properties.

Rob Lemmens presented the Living Textbook tool. It is a tool that allows us to create a concept map as an alternative to yEd, the current tool that we use to create the WG5 CS ontology. The Living Textbook tool contains a visual representation of a concept map and a text window with concept descriptions (see also Figures 10, 11 and 12 in this report). The Living Textbook is currently being used and further developed for educational activities at ITC, as well as in projects and research.

Friederike Klan presented a method for the semantic annotation of forms used for mobile data collection in Citizen Science projects at design time. In doing so Citizen Science data are enriched with machine-readable metadata while they are collected. Using those annotations and a set of templates, the data collected can be flexibly exposed according to different data models.

Such a method can be useful as a tool supporting the collection of standardized Citizen Science data and their interoperability. At the same time, it hides the complexity of the underlying formal data models from Citizen Science contributors and project designers, which facilitates an easier uptake.

The method has been developed in a cooperation of the DLR-Institute of Data Science and the Friedrich-Schiller-University of Jena, Germany and has been implemented as an extension to the mobile data collection framework Open Data Kit (https://opendatakit.org/). The project is available at GitHub (https://github.com/MaSteinberg/aggregate, https://github.com/MaSteinberg/build) and will be presented at the EGU General Assembly 2019 (https://meetingorganizer.copernicus.org/EGU2019/EGU2019-19123.pdf).

8.2 Ontology development

8.2.1 Top-level concepts

As starting point, together with the CSA report, the top-level model developed in the Milano meeting (see Figure 1 and 2) has been taken into account, which proposed a grouping of the existing attributes into a set of modules/profiles.
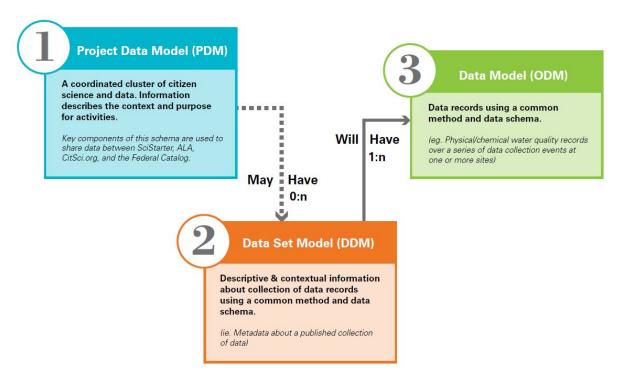
Figure 1. The PPSR-Core data model framework: A common data model with three main schemas (from
https://www.wilsoncenter.org/sites/default/files/wilson_171204_meta_data_f2.pdf).
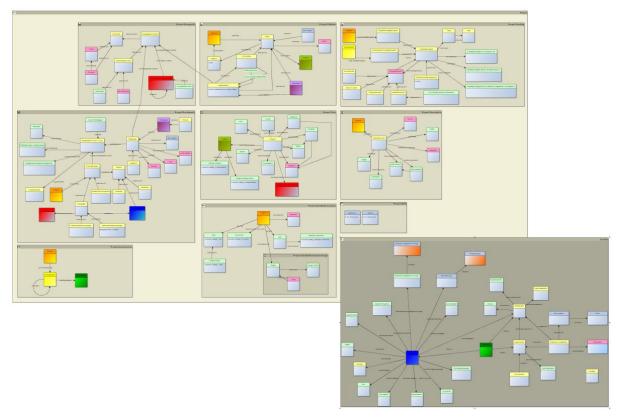

Figure 2. Proposed grouping of Project Metadata Model (PMM)'s existing attributes into a set of modules (Milano model)

8.2.2 Project Metadata Model (PMM)

In Milano, we introduced the module structure, building around a core module with the most basic information (see Figure 3). Elements with green background represent concepts/attributes taken from PPSR-Core, those with a pink background are concepts/attributes that we added. Furthermore, we included interconnecting concepts (in Figure 3, Project, Contact and GeographicAreaOfInterest) and added semantics to the relationships.

We also included a series of surrounding 'modules/profiles' (see list on the left in Figure 4), each of them covering the needs of a particular possible use case. One, for example, includes extensions related to funding, supporting technical infrastructure or information about the geography of the project. All of these extensions are considered optional and might be used depending on the purpose of particular activities.
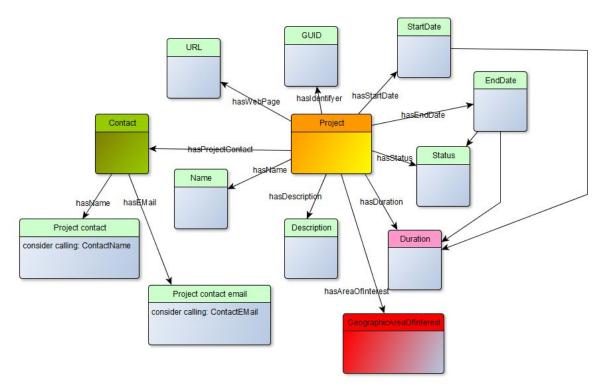

Figure 3. The module/profile structure introduced in Milano, building around a core module with the most basic information
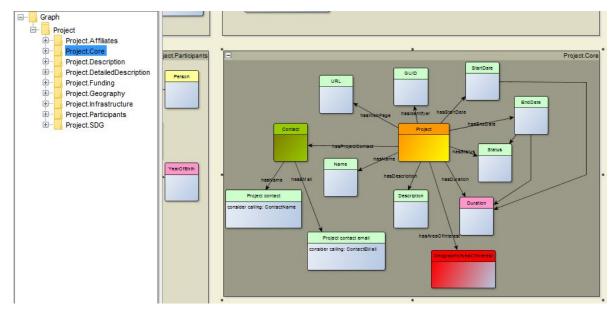
Figure 4. Modules/profiles of the ontology

Decisions made:

- We will use GitHub to host the ontology.
- Peter Brenton, of the MaGiC committee, will create the repository, as per CSA WG decision (2019).
  - From the Geneva declaration (2018): "Creation of the Model Governance Committee, or MaGiC, within the CSA International WG on Citizen Science Data and Metadata (proposed members: Greg Newman, Luigi Ceccaroni, Rob Lemmens, **Peter Brenton**, **Sven Schade**, Lucy Bastin, Anne Bowser, Robert Stevenson)"
  - "Publication of the current version of the Common Conceptual Model on GitHub using an account to be created by the MaGiC committee of the CSA International WG on Citizen Science Data and Metadata"
- We will support the population of the repository with actual content and its moderation.
- With respect to the relation with the WeObserve Interop CoP & OGC CitSciIE, we adopt the OGC standard as starting point for the ODM.
- We analysed a JRC proposal for extension based on initial requirements from a 500-project case study.
  - Proposal to add a **"hasName"** property in "Agent", which will associate "Agent" with **"Name"**. Decision: approved (see Figure 7).
  - Proposal to add a **"hasCategory"** property in "Organization", which will associate "Organization" with **"OrganizationCategory"** (see Figure 5). An initial controlled vocabulary could contain the inventory values: Governmental, Non-governmental, Academic, Private sector, Community-led, Consortium. Decision: The property should be "subClassOf". Approved with this change.

- Proposal to add a **"hasCategory"** property in "Project", which will associate "Project" with **"ProjectCategory"**. An initial controlled vocabulary could contain the inventory values: Passive sensing, Crowdsourcing, Volunteer computing, Monitoring, Occasional reporting, DIY engineering, Civic science, ~~Facilitating platform~~ (Adapted from Haklay et al. (2013) Citizen Science and Policy: A European Perspective.). Decision: rejected. This property is already there with a different name: "hasActivity" related to "ProjectActivity" (see Figure 6). The controlled vocabulary is accepted except from "Facilitating platform".
- We started to populate, in a specific file, the controlled vocabulary of the foreseen **roles** that the agents may play in a CS project with the roles that we identify in practice. Some examples: "Lead organization" (or "Project leader"), "Project member", "Project initiator". See also [https://sparontologies.github.io/scoro/current/scoro.html].

Related resources:
- [https://www.w3.org/TR/vocab-ssn] - Semantic Sensor Network Ontology (W3C Recommendation)
- [http://external.opengeospatial.org/twiki_public/CitSciIE] (OGC Interoperability Experiment)
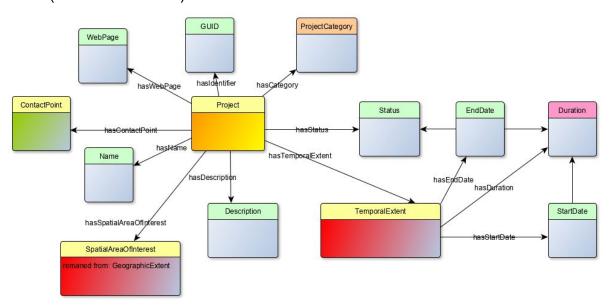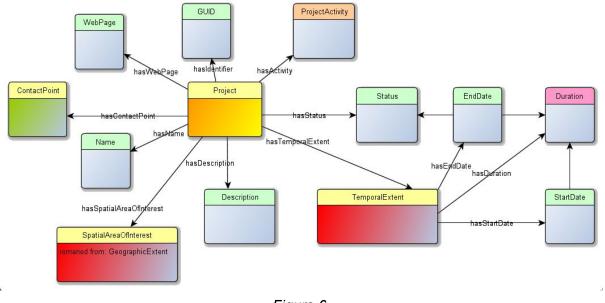- [http://external.opengeospatial.org/twiki_public/CitScieIE/SosServerClients] (list of SOS clients)
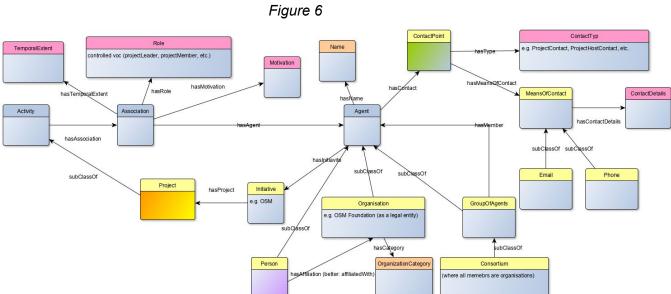

Figure 5

Figure 6

Figure 7

8.2.3 Dataset Metadata Model (DMM)

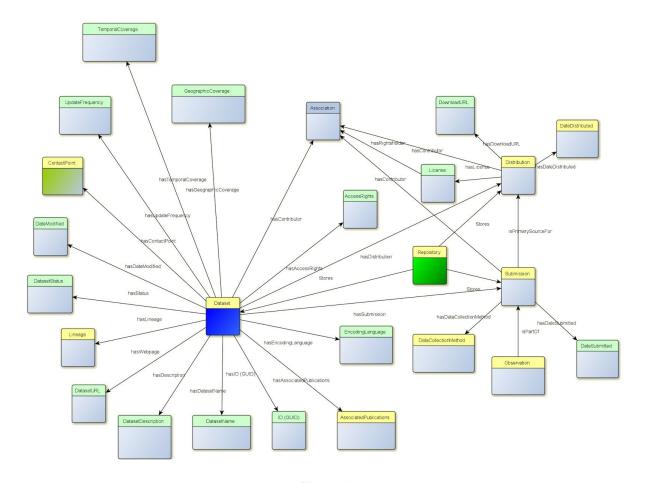The following decisions were made about the DMM (see Figure 8):

Figure 8

1. Lineage - New profile but we need to be careful to avoid overlaps.
2. DatasetGeographicCoverage - Without having a strong opinion, as it exists we shall keep it.
3. DatasetScope - As the project scope was removed, this should also be removed to avoid any confusion.
4. DataQuality - Data-quality profile might be associated with the 3 models.
5. Licence - This is required as it adds clarity to the dataset model.
6. Contributor - Ok to remove class but we must make sure that the Contributor is added to roles.

An alternative terminology has been proposed for the "Association" class (see Figure 9): the "Statement" pattern from Wikidata:

- Activity ->(p:hasAgent)-> Statement ->(q:hasAgent)-> Agent
- reference : [https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format#Truthy_statements]
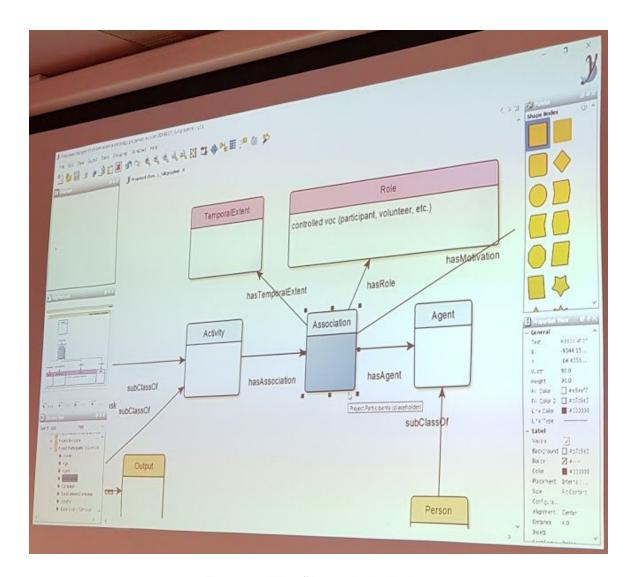
Figure 9. The "Association" class

8.2.4 Observation Data Model (ODM)

ODM was started by briefly reviewing existing standards. The following standards are considered relevant:

- https://www.w3.org/TR/vocab-ssn/#Sampling
- https://docs.google.com/document/d/1znzSW1KwDKcBWLNDfYXIAlOAZgcvc7umSnqLMG6_4M0/edit#

SSN is composed of several other models. For ODM we need to choose the relevant parts. A first top-down view has been drawn as in Figure 10.

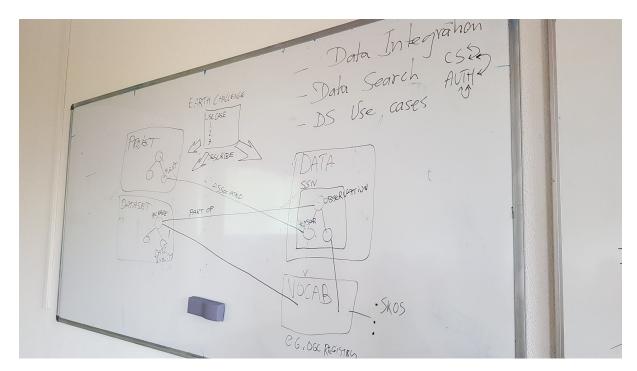The Living Textbook tool has been used to visualise the ODM (see Figures 10, 11 and 12).

Figure 10. The drawing in this picture is also represented in the Living Textbook tool [https://ltb.itc.utwente.nl/] (see Figure 11 and 12).
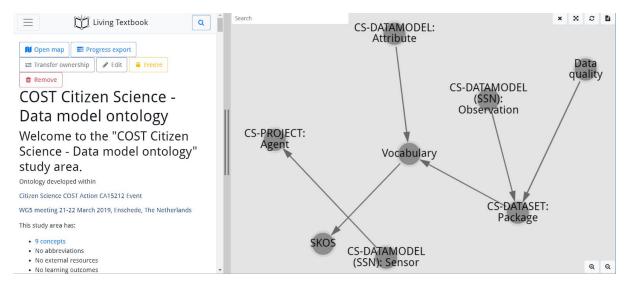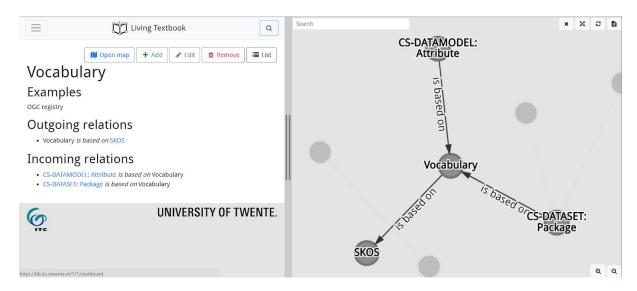

Figure 11

Figure 12

Terminology

We created an initial document with allowed terminology [https://docs.google.com/document/d/1znzSW1KwDKcBWLNDfYXIAlOAZgcvc7umSnqLMG6_4M0/edit#].

Within two weeks, the following people will have a meeting about a solution on vocabulary management (create a link between ontology and vocabularies, e.g., as construct in OWL): Gilles, Chrisa, Rob, Valentine, Stanislav and Friederike.

8.2.5 SDG profile

We created an "SDG profile" and analysed the relevance of the survey prepared by Kalterina [https://docs.google.com/forms/d/e/1FAIpQLSfVnMtYy8-PlWcRsQTIn0BNQyPEn9tz2EWPp1oHx-gXjtWWug/viewform?vc=0&c=0&w=1&usp=mail_form_link] to elements in our ontology, in order to support the identification of impacts of SC projects related to SDGs. This is also relevant to the JRC's 500-project- survey.
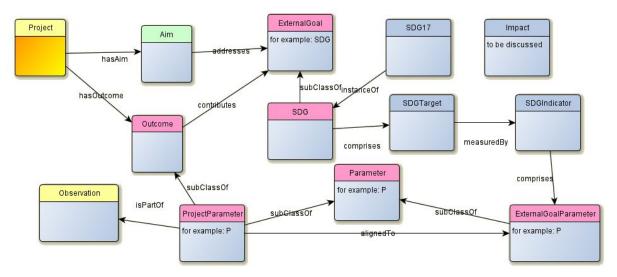

Figure 13

This module/profile should be related/linked to:

- the SDG ontology (a big one but incl. what we need, i.e. SDG, targets, indicators): [http://aims.fao.org/activity/blog/sustainable-development-goals-interface-ontology-sdgio-support-united-nations] [https://github.com/SDG-InterfaceOntology/sdgio];
- IAASA work on linking CS to SDG indicators in the context of WeObserve (contact: Dilek Fraisl);
- Global SDG and Citizen Science maximisation group [https://www.surveymonkey.com/r/LDHYRNQ] (contacts: Libby Hepburn libby@atlasoflife.org.au, Rosy Mondardini maria.mondardini@uzh.ch, Caren Cooper cbcoope3@ncsu.edu);
- CODATA group [http://www.codata.org/task-groups/citizen-science-for-the-sustainable-development-goals] (contact: Caren Cooper cbcoope3@ncsu.edu).

8.2.6 Data quality

We analysed Peter Brenton's proposal on data quality and produced the following feedback:

- Biodiversity specific concepts should be separated into a specific layer.
- All to have a closer look
- Luis Felipe to share comments on Basecamp
- Consolidation of COST feedback in Brussels (April 2)
- High/medium/low levels of quality have no meaning associated and can be considered ambiguous
- Consider terms modelled by the EBV group: Valentine to add a link

8.2.7 "Impact" profil

Earthwatch to start to work on an "Impact" profile.

8.3 Use cases

For the purpose of validating our ontology, we decided to test our model by mapping the classes on existing use cases and to define use case scenarios of the ontology.

Our high-level use cases are the following:

- Data integration
- Data search

We started with pre-selecting use cases from practice/projects. Use case scenarios can be derived from them by stating for each scenario a user, his/her activity and the reason of carrying out this activity. An example (minimal) use case (taken from a water monitoring project):

- As a water engineer (who),
- I want to receive alert on a broken water point (what),
- so that I can make a quick fix plan (why).

We will choose project cases in which we know people active in that project.

1. Try to map every class in our otology to the project content. This could be done with screenshots of our ontology and a spreadsheet with list of concepts at the side.
2. Map the content of project descriptions. This could be done with semantic annotation, e.g., GATE:

https://gate.ac.uk/

https://en.wikipedia.org/wiki/General_Architecture_for_Text_Engineering#/media/File:GATE5_main_window.png

We defined a choice of use cases, e.g. from other COST WGs or from outside the COST action, e.g. Earth Challenge 2020:

- An inventory of citizen science activities for environmental policies (http://data.jrc.ec.europa.eu/dataset/jrc-citsci-10004) - further exploitation by the JRC. Possibility to add their citizen science projects for **all participants and the public:** https://ec.europa.eu/eusurvey/runner/CSProjectInventory

- JRC Data Catalogue (http://data.jrc.ec.europa.eu/) for requirements in dataset description

- Earthwatch: MICS modelling will follow COST model (PMM)

- Earthwatch: FreshWater Watch to use the model retrospectively

- COST action on Alien Species: JRC to follow up

- CS COST Action's WGs:

  - **WG4: Luigi to get in touch with Artemis Skarlatidou to study the possibility of having a use case and to have a joint session during a WG4 meeting in April (Brussels) or November (Brno).**

  - **WG3: JRC to get in touch with Sven to study the possibility of using the work on collecting information on citizen-science portals as a use case.**

  - **WG6: Luigi to get in touch with Josep Perello' to understand if this can be the context for inter-WG collaboration on use cases.**

Research questions, Earth Challenge 2020:

http://earthchallenge2020.earthday.org/

Three main steps of Earth Challenge 2020:

1. generate questions (already done, no direct link with ontology)

2. create software (relation with ontology mostly to ensure data standards)

3. launch project using software (find suitable projects, or software to carry out a new project)

Main relevant objective: data integration and discovery.

User stories:

1.Earth Challenge 2020:

"*This will require identifying what data already exists in these areas, making sure that existing data can be documented in a harmonized way*" as per definition in the Earth Challenge 2020 website. We can offer an ontology that allows them to do that; showcase ODK with ontology template; this is a solution for the Earth Challenge 2020 mobile app integration challenge.

2.Researcher or research team:

How can I integrate data from different teams/projects or software/prototypes? What should I do to make "my" data interoperable with other Earth Challenge 2020 efforts?

If you want to do this, use our (COST) ontology!

3. Local activist or citizen scientist:

They have little knowledge in citizen science and research data so far, but knowledge in building apps, or simple interest.

Set-up or find citizen science project: identify similar projects nearby.

How can I find a project that fits my interests and is nearby?

Use natural language for query, e.g. "Which project is working on local air quality within the province of Overijssel?"

Other options:

- Zooniverse https://www.zooniverse.org/
- Interoperability experiment WeObserve: 4 citizen observatories (One is Ground Truth 2.0)
- Soil moisture project (Friederike)

8.4 International uptake

- Existing projects: tool for knowledge extraction and ability to expose the project knowledge in some high-visibility repository in an interoperable format.
- New projects: tool for representing knowledge in an implementable way in line with the ontology.
- Gilles offers a collaboration with a student of his, expert in information systems.

- Valentijn: Find the core of our ontology to support uptake. Too many fields/parameters will not appeal to end-users - Back to the drawing table!
- We use the term "core" after "Darwin Core", a common metadata standard in life sciences, which has "core" components that are uniform across implementations but has flexibility for customization for specific needs.

9. Future meetings

- Monthly teleconferences of WeObserve CoP on Interoperability (Friederike?, Sven?, Rob?)
- Quarterly teleconferences of the CSA Data and Metadata WG (Luigi, Sven)
- April 2 (Brussels): ECSA Tools and technology WG meeting, COST WG2 meeting (Luigi, Jakub, Kalterina, Sven, Jaume)
- April 10-11: COST WG4 meeting (Rob?(reach out before the meeting to WG4 to discuss the relevance of a discussion at this meeting, Kalterina)
- June 4 (Cesis): COST WG5 and WG4 meeting (Luigi, Vyron, Jakub)
- November (Brno): COST WG4 meeting (Jakub)
- March 2-3 2020 (Uherské Hradiště): COST WG5 meeting (Luigi, Jakub)

10. Overall follow-up actions

(Also check action points in the other parts in this document.)

- With respect to Peter Brenton's proposal on data quality:
  - All to have a closer look
  - Luis Felipe to share comments
  - Consolidation of COST feedback in Brussels (April 2)
  - High/medium/low levels of quality have no meaning associated and can be considered ambiguous
  - Consider terms modelled by the EBV group: Valentine to add a link
- Earthwatch to start to work on an "Impact" profile.
- Provide examples of Agent and Association on Basecamp, to clarify if it should be modified. Other modeling examples should be mentioned there.
  - Alternative terminology: Activity ->(p:hasAgent)-> Statement ->(q:hasAgent)-> Agent
    - Reference: :
      https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format#Truthy_statements

Appendix

Observations on the proposed *extensions to PPSR-Core Dataset Metadata Model on data quality* produced by Peter Brenton - by Luis Felipe PV

The comments below are a result of a personal review of the document produced by Peter Brenton, and they are aimed to feed and help enhance the process started by the aforementioned. The format of the comments is:

> ● → **Peter Brento proposed extension**¶
> Comment and justification Luis Felipe¶

- **Type of sampling method / Survey-sampling method / Description of survey-sampling method / Web site of survey-sampling method / Document of survey-sampling method**

  One believe it is particularly important to think if these suggested extensions are adding clarity or complexity to the DMM. Currently the DMM has a class named *DataCollectionMethod*, when this class was thought of, was the intention to record the information described by all the above mentioned? If that is the case, it is one's opinion that the proposed extensions are not needed. On the other hand, if the consensus is that they are needed, I would suggest these are added to the existent DMM by linking them to the *DataCollectionMethod* class using the connector *AsDescriptionOf.*

- **Data access method**

  The current DMM has an existing connection *hasDistribution* to the class *repository* and further to the class *distribution*, it is one's opinion the existent model suffices, and this proposed extension is not required

- **Legal custodian**

  The current DMM has a class called *ContactPoint*, therefore the WG might need to clarify if there is overlapping between the intended meaning of the existent class and the definition of the proposed extension. If there is an overlap in meanings, the suggestion is to use *LegalCustodian* as the class name in the DMM. Datasets tend to be related to a custodian rather than a contact point

- **Data quality assurance method**

  It is one's opinion that the description of this proposed extension has concepts related to Quality Control (QC) within data management practices i.e. *curating and managing the dataset,* thus making this extension all about the dataset method which the DMM has already described in the class *DataCollectionMethod*. Also see comment in point raised in the first comment regarding this same class. As it currently stands, the DMM suffice the description of the extension and this might not be needed.

- **Data quality assurance description**

  I agree this extension is needed. However; the proposed description refers to the data QC process in data management. Therefore, if the extension is accepted, I would like to suggest changing the name to *DataQualityControl* and for it to be related to the *DataQuality* (new profile) class using the connector *AsPartOf*

- **Data usage guide**

  It isn't clear how adding the proposed extension will enhance the information already provided by the collective group of classes defined within the DMM. As per the description in Peter's document, this extension could lead to a subjective description of the dataset which could result in an unintended bias approach to the data use. One will expect that the metadata as an entity provides enough information to the user as to inform the data usage.

Finally, the proposed extensions that have not been mentioned were identified as valuable in the enhancement of the ontology during the meeting. Please refer to the meeting's minutes in the data quality section for recommendation.