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

Minutes of WG5 workshop in Milan: “Towards a new ontology of citizen science”

Hosted by: Politecnico di Milano, and Earthwatch
Milan, Italy

March 26-27, 2018

—

Attendees

Luigi Ceccaroni (chair)
Maria Antonia Brovelli
Rob Lemmens
Jaume Piera
Jakub Trojan
Frank Ostermann
Marco Minghini
Vyron Antoniou
Lucy Bastin (remotely)
Sven Schade
Chrisa Tsinaraki
Maryam Lotfian
Gloria Bordogna
Santiago De La Riva

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).

By convening COST Action's members with expertise in data and metadata representation, COST Action's Working group 5 ("Improve data standardization and interoperability") 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 further refining core requirements, ontology fields, and associated metadata with input from the larger citizen science community.

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 second period (1.5.2017-30.4.2018) 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.

Organizers

The two organizers of the meeting are: Luigi Ceccaroni (chair of the WG) and Maria Antonia Brovelli (member of the WG and professor at the Politecnico di Milano, Department of Civil and Environmental Engineering).

Politecnico Milano is a scientific-technological university which trains engineers, architects and industrial designers. It has 42822 enrolled students (a.y. 2017/2018; update: January 2018), of which: 7040 architects, 4156 designers and 31,626 engineers. It is in the top 10 Universities in the world in Civil & Structural Engineering (QS Ranking 2018: 9).

The Department of Civil and Environmental Engineering [http:// www.dica.polimi.it/en/] at **Politecnico di Milano** (PoliMI) was founded in 2013 and consists of professors and researchers working in many different research fields. Maria Antonia Brovelli is part of the Geodesy and Geomatics Section, which is deeply rooted in the history of Politecnico and is characterized as having always been a promoter of advanced research themes, both at national and international level. Following the developments in the fields of electronics and informatics, the evolution in the topics regarding earth observation and data acquisition through images has led to the establishment of a new frontier for this area, which is nowadays known as Geomatics. All the research activities share a common ground represented by the data modelling and by the use of statistic, geo-statistic and inference methods for the analysis of the data and their prediction. These fundamental aspects represent a basis for all the activities, leading to substantial contributions both in the methodological approach and in numerical applications. The main research lines of the Section include: Physical and Mathematical Geodesy; Earth observation; Surveying, Photogrammetry and Remote Sensing; Geographic Information Systems (GIS). Citizen Science is one of the expertise of the GIS area (http://geomobile.como.polimi.it/), which is leaded by Maria Antonia Brovelli (http://geomobile.como.polimi.it/website/cvs/CV-European-%20Brovelli-English-2018.pdf). The GIS team focuses its activities on many different topics closely related to geographic information field, which span from the traditional GIS applications, spatial analysis and GeoWeb services to the most recent research topics resulting from the Web 2.0 era, among these VGI (Volunteered Geographic Information), Citizen Science, geo crowdsourcing and Big GeoData.

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.

Agenda

March 26

9:30 - 11:00

1. Luigi Ceccaroni, Earthwatch: Overview of the past, present and future goals for ontology development in citizen science.
2. Sven Schade, EC JRC: Citizen Science Data and Metadata.

11:00- 11:30 coffee break

11:30-13:15

3. Rob Lemmens, ITC University of Twente: Presentation of the yEd editor and its applications.
4. Contributions to the analysis of the file "PPSR - Core project metadata mapping - multiple standards".
5. Revision and augmentation of the Citizen Science Association Data & Metadata Working Group - Report from CSA 2017 and Future Outlook (Bowser, Anne, Brenton, Peter, Stevenson, Rob, Newman, Greg, Schade, Sven, Bastin, Lucy, Parker, Alison and Oliver, Jessie) with focus on Project Data Model and Data Set Model.

13:15-14:45 lunch

14:45- 16:00

6. Connections with existing projects : [http://www.envriplus.eu/], a relevant deliverable from ENVRI - collects a number of research-infrastructure representations which have some overlaps with the citizen-science knowledge model: [http://www.envriplus.eu/wp-content/uploads/2015/08/D8.3-Interoperable-cataloging-and-harmonization-for-environmental-RI-projects-system-design.pdf]
7. Related examples : Open Information Linking for Environmental science research infrastructures (OIL-E) [http://oil-e.net/ontology/]
8. Project metadata schema - define and document a standard core.

16:00-16:30 coffee break

16:30-17.30

9. Brainstorm on what's missing from the initial list of concepts.
10. Analyze what is working well; what could still improve; select additional important fields to be included in the ontology.

11. Analyze where each of us is in their efforts with respect to priorities, motivations, drivers.
12. Session on what we agree to be core attributes and what are non-core.
13. Progress our previous discussion on standardization. Definition of processes around managing changes to vocabularies.

March 27

9:30 - 11:00

1. Refinement of core requirements of the ontology of citizen science based on existing experience (not only in biodiversity).

11:00- 11:30 coffee break

11:30-13:15

2. Refinement of the identification of core ontology classes and associated relations by drawing on previous research, existing vocabularies and instantiations.

13:15-14:45 lunch

14:45- 16:00

3. Development of a plan for further refining core requirements, ontology classes and associated relations with input from the larger citizen science community.

16:00-16:30 coffee break

16:30-17.30

4. Discussion, closing thoughts.

Notes

Day 1

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 (see SWE4CS paper [https://portal.opengeospatial.org/files/?artifact_id=70328], [http://external.opengeospatial.org/twiki_public/CitizenScienceDWG/WebHome], [https://github.com/opengeospatial/swe4citizenscience], [https://github.com/opengeospatial/swe4citizenscience/tree/master/compiledER]).

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 recently released by the CSA Data and Metadata Working Group dealing with PPSR - Public Participation in Scientific Research [https://www.wilsoncenter.org/article/citizen-science-association-data-metadata-working-group-report-csa-2017-and-future-outlook]. In particular, in this meeting focus will be placed on modelling two out of the three PPSR elements: projects and datasets. Modelling the third element (data/observations) will be left as a future task. While the PPSR-Core report does only detail a number of properties and classes, the goal of the meeting is to extend this list to any other classes and properties considered as essential, and to establish the relations between them. 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. For example, the “ProcessSteps” class (see Figure 2) acted on / collected observations to create datasets, distributions and submissions; quality can be stored at an observation level (e.g. as tests and assertions) and combined to form a package-level quality report when the data are sliced & diced in many different ways, observations can be associated with observers (agents) who have their own quality characteristics/reputations, which in turn may be inferred from the tests applied to their observations.

During the second part of the workshop, we split into two groups. One addressing the Data Set Model (DSM) in more detail, the other focussing on the Project Data Model (PDM). Both groups followed parallel modelling exercises, starting from the elements that are proposed by the international working groups, and then suggesting extensions (and modifications where considered necessary).

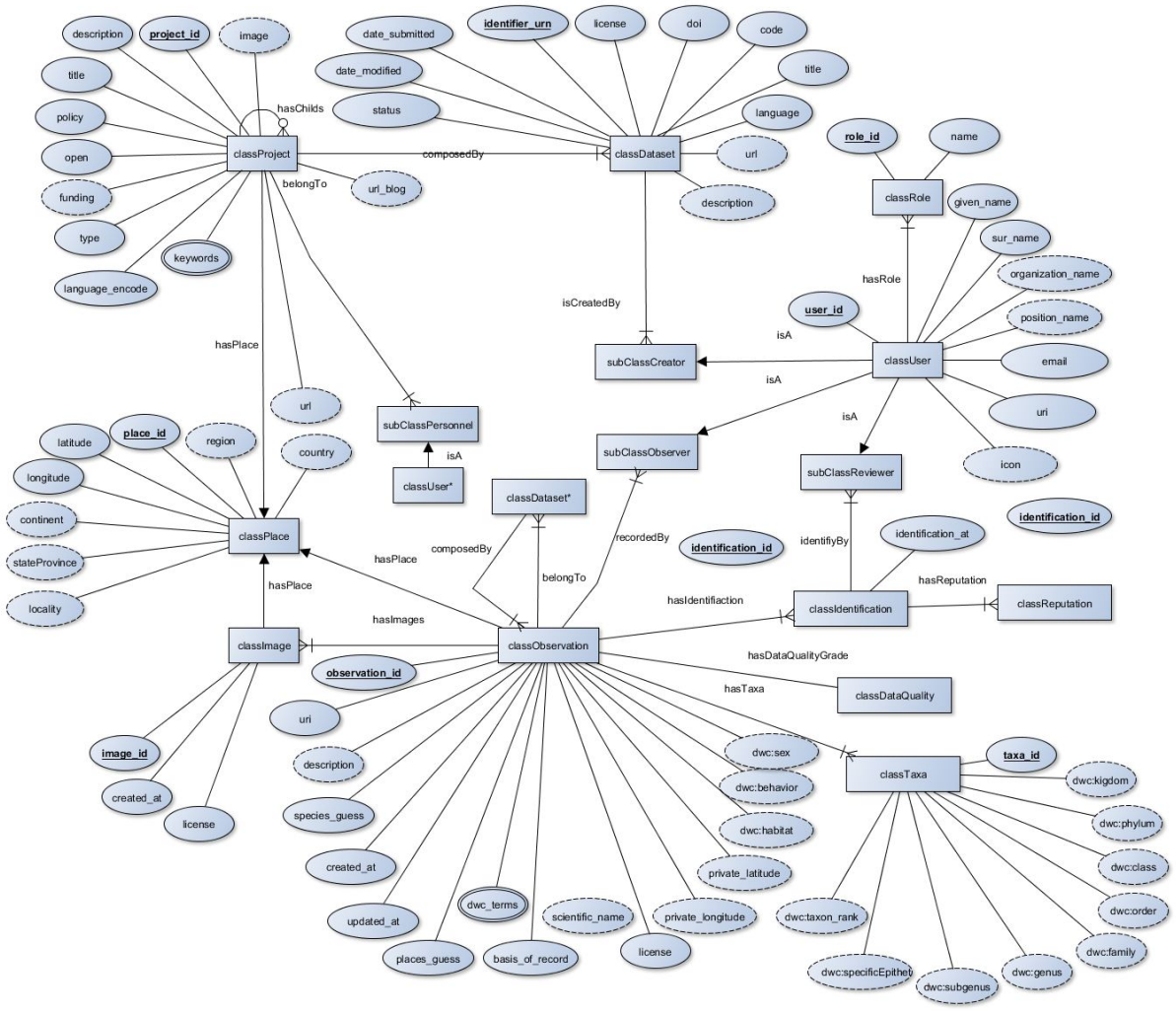

Figure 1. Natusfera's concepts

Methods of data aggregation

It is important to define different types for "data-aggregation methods" (surveys and campaigns). ISO 19115 defines this concept as "distribution" (see below).

Top-level concepts

As starting point, together with the CSA report, the top-level models developed in the Novi Sad meeting (see Figure 2) and during summer 2017 (see Figure 3) have been taken into account.

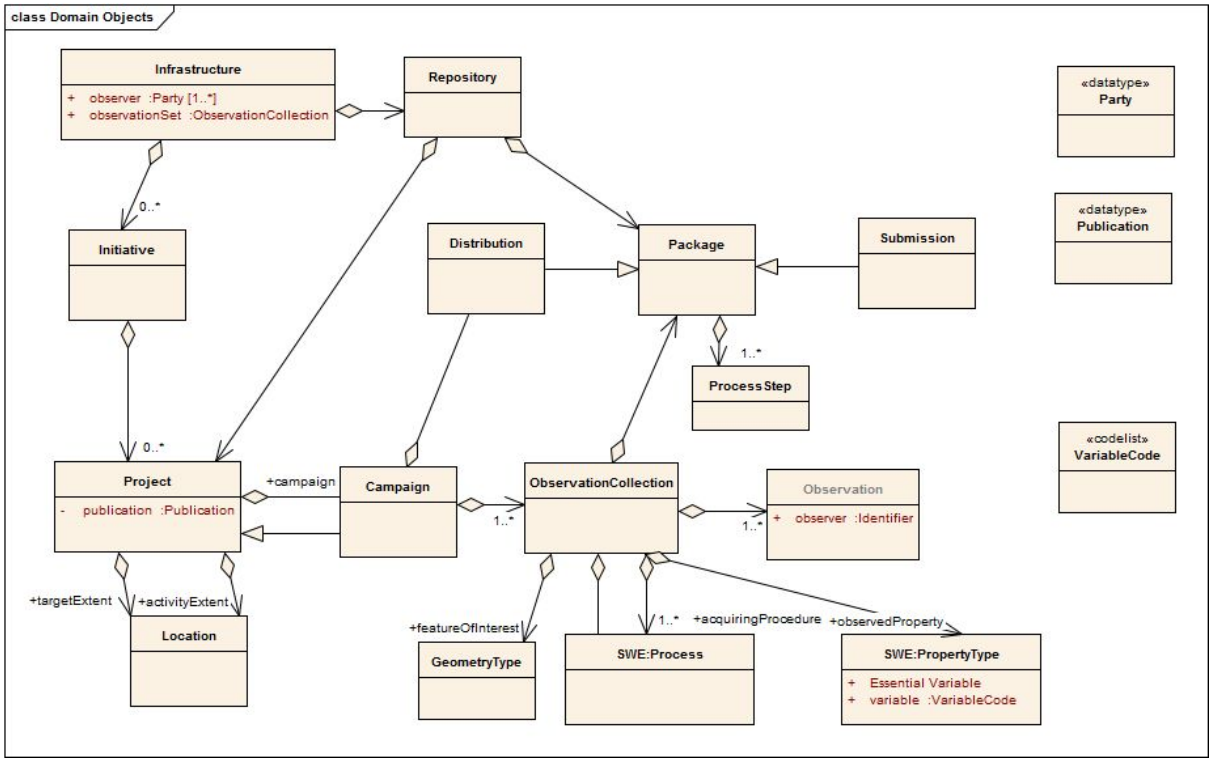

Figure 2. Novi Sad's top level model

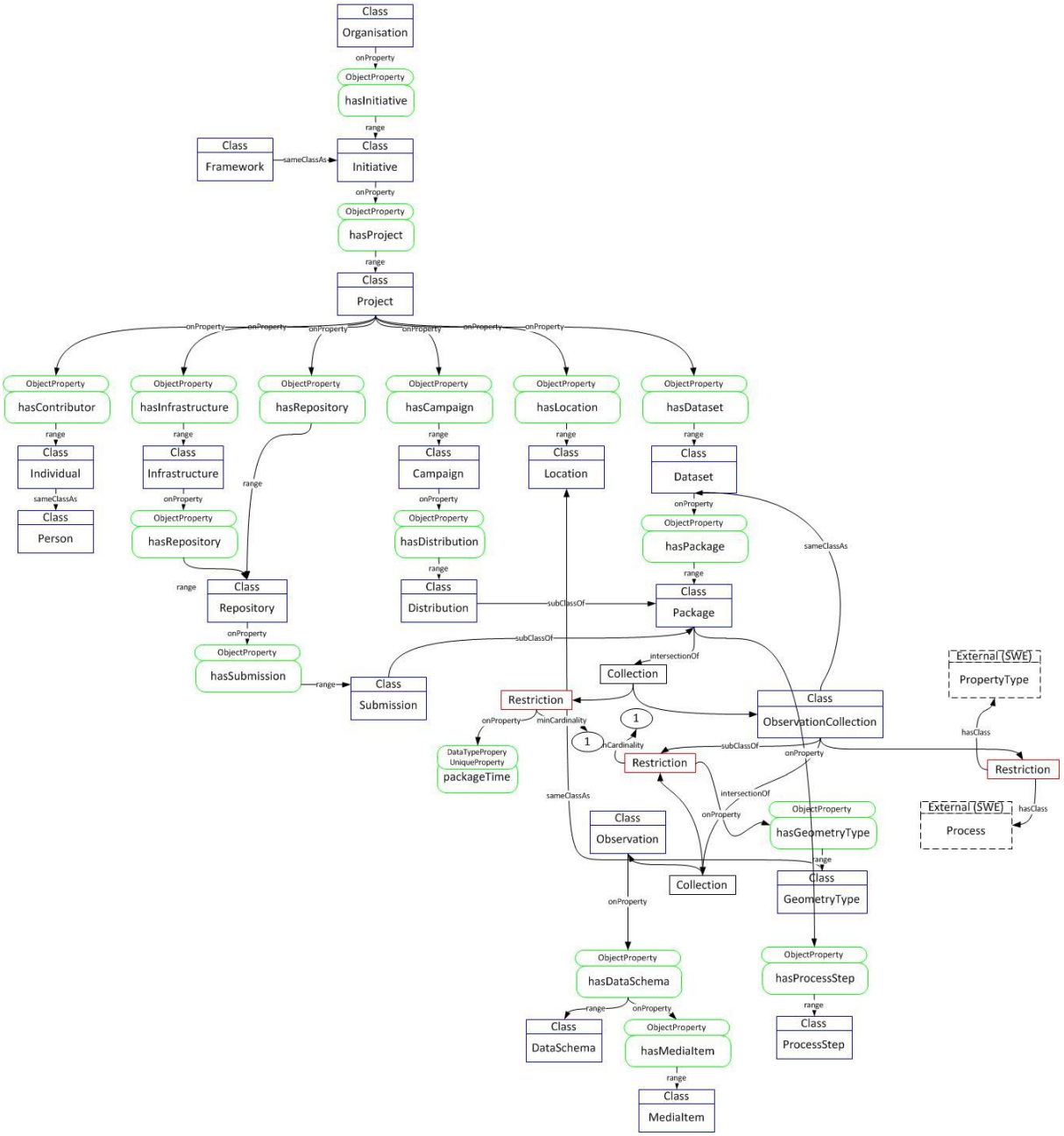

Figure 3. Post Novi Sad's top level model

Day 2

The group working on the Project Data Model (PDM) proposed a grouping of the existing attributes into a set of modules (see also Figure 4).