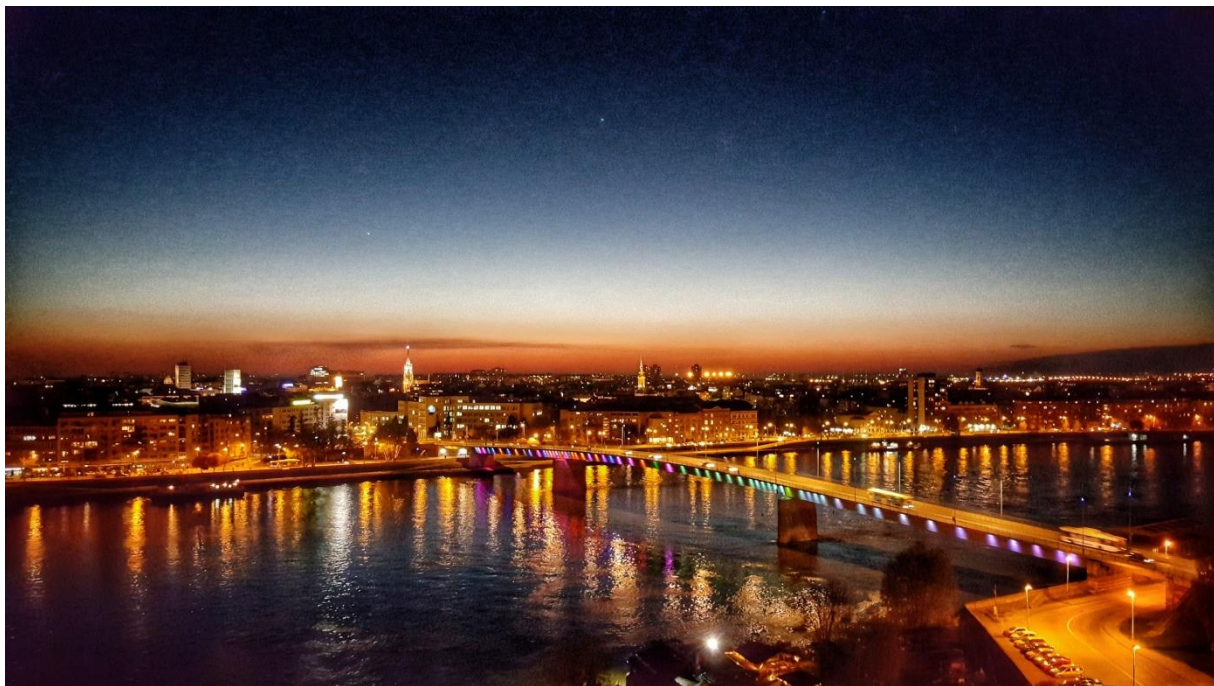

COST Action CA 15212 “Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe”

Minutes of WG5 workshop in Novi Sad: “Towards a new ontology of Citizen Science”

Hosted by:

University of Novi Sad, and 1000001 Labs

Novi Sad, Serbia



March 27-28, 2017

Attendees

Luigi Ceccaroni (chair)
Imre Lendak (co-chair)
Lucy Bastin
Rob Lemmens
Vyron Antoniou
Jaume Piera
Jakub Trojan
Mirjana Žabić
Joan Masó

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 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.

2. Continue to identify core ontology fields and associated metadata by drawing on previous research and existing vocabularies.
3. Develop a plan for further refining core requirements, ontology fields, and associated metadata with input from the larger citizen science community.

Objectives

Working group 5 “**Improve data standardization and interoperability**”’s general objectives 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 first period (1.11.2016-30.4.2017) 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 University of Novi Sad (www.uns.ac.rs) is the largest state funded higher education institution in Novi Sad, Serbia, with over 50,000 students and 5,000 employees.

One million and one labs (1000001 Labs) is a cooperative established to study artificial intelligence solutions to manage environmental data, and to advance international initiatives on citizen science data and metadata standardization and interoperability.

Agenda

March 27

1. Luigi Ceccaroni, 1000001 Labs: Lightning talk overview of the past, present and future goals for the ontology development.
2. Contributions to the analysis of the file "PPSR - Core project metadata mapping - multiple standards".
3. Decide how to distribute focus and effort across both (a) project metadata and (b) observational & analysis data.
4. Project metadata schema - define and document a standard core.
5. Brainstorm what's missing from the initial list of concepts.
6. Analyze what is working well; what could still improve; select additional important fields to be included in the ontology.
7. Analyze where each of us is in their efforts with respect to priorities, motivations, drivers.
8. Session on what we agree to be core attributes and what are non-core.
9. Progress our previous discussion on standardizing. Definition of processes around managing changes to vocabularies.
10. Custodianship, governance and maintenance of a standard.

March 28

11. Joan Masó: OGC's work related to an ontology of citizen science.
12. Scope and priorities for the ontology, including collaborating with other similar initiatives such as the TDWG Citizen Science Interest Group (<http://lists.tdwg.org/pipermail/tdwg-content/2012-May/002892.html>), CSA's International citizen science metadata Working Group and OGC.
13. Discussion of the work around attribute-level standardization and semantic enablement of attribute-level data.
14. Effective promotion and communication of the ontology to the world. How to build profile and uptake?
15. Issues and direction setting/planning session on the observation/analysis component.
16. Jaume Piera and Luigi Ceccaroni (They will participate in the CSA Conference in May in the US): Definition of a joint WG5 strategy for that big conference - e.g. whom to meet and connect with.

17. Refinement of core requirements of the ontology of citizen science based on existing experience (not only in biodiversity).
18. Refinement of the identification of core ontology classes and associated relations by drawing on previous research, existing vocabularies and instantiations.
19. Development of a plan for further refining core requirements, ontology classes and associated relations with input from the larger citizen science community.
20. 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] and [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.

Rob Lemmens presents the ontology development results of COST action IC1203 ENERGIC. He shows the ontology in WebProtege and yEd (graph editor). He also shows the Semantic Wiki page of ENERGIC, as well as the Wiki of GIS core materials used at the University of Twente.

The attendees analyze the “PPSR - Core project metadata mapping - multiple standards” document, which is the basis used for concept definition. The idea is to start from **project**-related entities. **Observation**-related entities will be defined at a future time. Initially, the focus is on ontology concepts and basic relations. Specific relations will be analyzed later. Thoughts discussed:

- Inter-relations among different types of projects are taken into considerations. A super-project (superclass of “project”) can be not only a parent, but a framework for multiple projects. A special *globally unique identifier* (GUID) will be defined for each project. This GUID might be human readable and contain explicit information about the underlying element, such as continent, country, project name (e.g. “Citclops_Europe_Spain_01”).
- There can be different datasets within a single project. (One project should be able to create multiple datasets.) These datasets have to be differentiated from sub-projects’ datasets.
- “Submission” is different from ‘project’ and ‘dataset’. It is a type of package created from an observation collection with minimum changes with the purpose of submitting it to a repository for preservation and sharing.
- Observations are important in this context as well. Each observation should have a unique identifier as well. A dataset might be regarded as a set of observations.
- The top level authority responsible for a project should be identified.

Reference links have been added to each metadata-schema column.

The following concepts (and relations among them) have been taken into account in the modeling exercise:

Project hierarchical structure

- **Infrastructures** (e.g., Natusfera [natusfera.gbif.es/], iNaturalist [<https://www.inaturalist.org/>]). They have a larger life-span than projects and different funding schema (e.g., organizations might allocate some regular budget to support them for longer periods than projects).

- **Frameworks** (or parent projects, or initiatives). Frameworks allow to group projects somehow linked (e.g., with a common purpose). For example, an organization might create a framework with its brand and then create several projects (with the same brand within the framework) with particular purposes. Usually, frameworks are linked to an infrastructure.
- **Projects**. They focus on reporting observations on particular targets (domains or locations).

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).

Day 2

Joan Masó presents the state of the situation in OGC about standardization efforts around citizen science.

A report exist on SWE4CS analyzing how SWE can be applied to citizen science (by Ingo Simonis). This report is important to lower the level of abstraction associated with the new standard. The possibility to organize a hackathon about this will be explored (by Joan Maso’ and Anne Bowser) about how to share observational data (transformation from original format to SWE4CS format). The next OGC meeting is in June in Canada; the one after that is in the UK in October; then in New Zealand in December. Regular teleconferences will be organized by Joan Masó on a monthly basis. Joan can also help in relating future implementations of the ontology in OWL and JSON-LD.

Rob and Vyron are planning also a hackathon about app development.

Presentations in the first f2f meeting on the OGC DWG.

•*Deploy 500 air quality sensors through citizens across Canada* Dr. Steve Liang, SensorUp Inc.

- *Research user feedback in spatial catalogues and realise a prototype implementation inside GeoNetwork. ELISE project.* Paul van Genuchten, GeoCat
- *A Stakeholder Analysis on Citizen Science Data Interoperability.* Anne Bowser Wilson Center
- *Land Cover and Land User Citizen Science data collected through Geo-Wiki; and the way forward within the LandSense Citizen Observatory* PERGER Christoph, IIASA
- *GroundTruth2.0 project. 6 case studies to interoperate with.* Joan Masó, UAB-CREAF

The “PPSR - Core project metadata mapping - multiple standards” document is edited, in particular adding domain objects to all properties. A UML class diagram is created with the following definitions for each class.

Definitions for the main classes (glossary):

Campaign: An effort limited in time and space to create an observation collection about variables or resources

Data: To be defined.

Dataset: To be defined.

Distribution: A type of package created with the purpose to expose a frozen view of the data as raw or aggregated:

What user want: Know all the options for accessing or getting the data

ISO 19115: orderingInstructions: general instructions, terms and services provided by the distributor

ISO 19115-1: description: brief description of a set of distribution options

Framework: A group of people with a common set of goals that facilitates the creation and maintenance of citizen-science projects. Frameworks allow to group projects somehow linked (e.g., with a common purpose). For example, an organization might create a framework with its brand and then create several projects (with the same brand within the framework) with particular purposes. Usually, frameworks are linked to an infrastructure.

GeometryType: Type of geometrical object used to describe the location of the observations

Infrastructure: A system composed by servers, hardware, sensors and the people to maintain them that can be used to host citizen-science projects (e.g., Natusfera [natusfera.gbif.es/], iNaturalist [https:// www.inaturalist.org/]). They have a larger life-span than projects and different funding schema (e.g., organizations might allocate some regular budget to support them for longer periods than projects).

Initiative: see Framework.

Observation: Individual observation. It is not described in this model.

ObservationCollection: Collection of observations that are about the same variable (property) using the same process (sensor platform) and the same geometric types

Package: A way to distribute an ObservationCollection frozen in a moment in time

Process (OM:Process): The procedure to obtain the observation. It can describe a sensor, a picture taking protocol, a human feeling.

ProcessStep: Process used to create a package

Project: A collaborative effort to report observations on particular targets (domains or locations)

Project abstract: To be defined.

Project aim: To be defined.

Project objective: To be defined.

Project outcome: To be defined.

PropertyType (SWE:PropertyType): Describes the type of variable or resource observed. It does not contain the value but could include vocabularies or value ranges.

Repository: Archive of citizen-science projects or datasets

Submission: A type of package created from the ObservationCollection with minimum changes with the purpose of submitting it to a repository for preservation and sharing

Action Items

1. Rob Lemmens will lead the preparation of a paper on the results of the workshop. We plan to have a draft version of the paper by the next WG5 meeting, which will (probably) be held in October 2017.
2. Luigi Ceccaroni will present the results of the workshop to the CSA's international Data and Metadata Working Group on March 31, 2017. (Done)

3. Luigi Ceccaroni and Jaume Piera will present the results of the workshop at the CSA Workshop on Data Quality/ WG Meeting at the CSA2017 Conference on May 17, 2017.
4. WG5 will continue to work on the file "PPSR - Core project metadata mapping - multiple standards" with the objective to create an ontology for describing citizen-science projects.
5. Clarify the distinction between Distribution and Submission. With shared characteristics defined by the Package class, Distributions and Submissions can be specialised to match the requirements of each. The key difference is that a submission is not explicitly designed for exposure, but other key distinctions will emerge when we try encoding real citizen-science projects.

Next Meeting

The next WG5 meeting will be in October 2017 (details to be confirmed).

