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

Minutes of WG5 workshop in Geneva: "On the citizen-science ontology, standards & data"

Hosted by: University of Geneva, and Earthwatch

Geneva, Switzerland



June 6, 2018

Attendees

Luigi Ceccaroni (chair) **Rob Lemmens** Jaume Piera Jakub Trojan Vyron Antoniou Lucy Bastin Sven Schade Maryam Lotfian Santiago Martínez De La Riva Friederike Klan **Gilles Falquet** Sarah Newman Jessie Oliver Jonathan Brier **Russell Scarpino** Greg Newman Anne Bowser Joan Masó Karel de Pourcq Peter Brenton

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:

- Refine core requirements of the ontology based on existing use cases and the discussion happening in the corresponding Basecamp project [<u>https://</u> <u>basecamp.com/2071195/projects/13342949</u>] (access restricted to WG5 members).
- 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-scienceassociation-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 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.

Organizers

The organizers of the meeting are: Luigi Ceccaroni (chair of the WG) and the three main citizen science associations: CSA, ECSA and ACSA, under the umbrella of the CSA's international Data and Metadata Working Group.

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

June 6

13:00 - 16:00

- 1. Luigi Ceccaroni, Earthwatch: Overview of the past, present and future goals for ontology development in citizen science
- 2. Sven Schade, EC JRC: Introduction to the Geneva Declaration on Citizen Science Data and Metadata Standards
- 3. Revision and augmentation of the Citizen Science Association Data & Metadata Working Group Report from CSA 2017 and Future Outlook (Bowser, Brenton, Stevenson, Newman, Schade, Bastin, Parker and Oliver)
- 4. Connections with existing projects
- 5. Project and dataset metadata schemas: defining and documenting a standard core
- 6. Brainstorming on what's missing from the initial list of concepts
- 7. Analysing what is working well; what could still improve; selecting additional important fields to be included in the ontology

- 8. Analysing where each of us is in their efforts with respect to priorities, motivations, drivers
- 9. Progressing previous discussion on standardisation; definition of processes around managing changes to vocabularies
- 10. Refinement of core requirements of the ontology of citizen science based on existing experience (not only in biodiversity)
- 11. Refinement of the identification of core ontology classes and associated relations by drawing on previous research, existing vocabularies and instantiations
- 12. Development of a plan for further refining core requirements, ontology classes and associated relations with input from the larger citizen science community
- 13. Drafting of the Geneva Declaration on Citizen Science Data and Metadata Standards
- 14. Discussion, closing thoughts

Notes

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.

Top-level concepts

As starting point, together with the CSA report, the top-level model developed in the Milano meeting (see Figure 1) has been taken into account, which proposed a grouping of the existing attributes into a set of modules/profiles.



Figure 1. Proposed grouping of Project Data Model (PDM)'s existing attributes into a set of modules (Milano model)

In Milano, we introduced the module structure, building around a core module with the most basic information (see Figure 2). 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 2, 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 3), 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 2. The module/profile structure introduced in Milano, building around a core module with the most basic information



Figure 3. Modules/profiles of the ontology

After introducing the state of the art and the work done previously in the COST Action, most of the meeting was dedicated to the drafting of the so called Geneva Declaration on Citizen Science Data and Metadata Standards, which is included below as integral part of the results of this COST meeting.

Geneva Declaration on Citizen Science Data and Metadata Standards

- June 6, 2018 - Geneva, Switzerland

With the growth of citizen science comes the challenge of coordinating people, projects, and data. This challenge also presents an opportunity. Through the use of data and metadata standards and other mechanisms to promote interoperability, we can support multiple research questions and use cases across numerous science domains, allowing citizen science to help address issues and problems on local, regional, national, and global scales. This declaration is about the development of a recommendation on how to represent data and metadata in citizen science.

In 2015, the *U.S. Citizen Science Association* (CSA) founded the International Data and Metadata Working Group to promote collaboration in citizen science through the development and/or improvement of international standards for citizen science data and metadata. This working group operates on citizen science at the international level, and became a coordinating and umbrella group crossing many thematic and geographically distributed organisations that provide relevant complementary work. One major body that provides input into the WG's discussions is the <u>COST Action 15212</u> "Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe" which also coordinates the inputs from the <u>European Citizen Science Association (ECSA)</u> into the overall discussions. The <u>Australian Citizen Science Association</u> (ACSA) is also a major contributor to this working group, too.

This declaration summarises the most central points that members of the **COST Action 15212's Working Group 5 ("Improve data standardization and interoperability")** agreed upon with other members of the **CSA International WG on Citizen Science Data and Metadata**. The document addresses the interoperability standards for data exchange, re-usability and compatibility. It thereby defines core building blocks of our joint work and outlines the way ahead based on the progress report that was published in December 2017. That progress report contained initial designs for the PPSR (Public Participation in Scientific Research) Common Data Model with its three main schemas. As one result of our joint discussion 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** - 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)**.

The PPSR Common Conceptual Model is designed to consist of a 'core' of key elements and concepts that are shared by all citizen science initiatives. By its nature, this core should be capable of representing generalised concepts across a range of domains. Therefore it will not completely capture the priorities and facets of all citizen science initiatives, and so it is foreseen that it may be extended using 'profiles' which include additional concepts.

Part I: Methodology

We will first agree on a small set of 'core' concepts and then offer a set of 'profiles' for more usage- and domain-specific extensions. Profiles will provide additional concepts that can be used for representing (for example) funding mechanisms, requirements for participation, necessary supporting IT infrastructure, policy support, and impact assessments and Sustainable Development Goals. We also envisage additional concepts that will help us to bridge between thematic standards of the natural sciences and social science data standards (e.g. to describe the details of particular participation mechanisms). Acknowledging the various possible uses of citizen science data and metadata, we will further develop the current version of the model following a use-case-based approach, extending the overall model in a stepwise manner, based on the priorities and requirements of the user community. Both the core model and the profiles which extend it are expected to use agreed vocabularies, and we will define and maintain the terms necessary to support the core model and the profiles we develop, while also offering a governance framework for extended vocabularies and codelists. Building bridges and providing guidance on how to use standards across communities with varying knowledge of standards and technical expertise will be as important as providing guidance on how to address a completely new domain.

We will provide a recommendation about how to represent data and metadata in citizen science, together with best practices that illustrate the use of these data and metadata. We will support emerging activities by providing an overarching framework and starting points when these activities develop their project and metadata descriptions, draft community surveys, and develop data management and sharing plans. We will be open to any feedback, especially to lessons learned and possible suggestions for improvement (e.g., new profiles).

The core model should remain stable, and we will indicate maturity levels for each profile. In this way, new features can be implemented gradually, minimizing the risk that the implementation of tools falls behind new versions of the model. Enhancements to the model and associated extensions will be made within an appropriate governance framework to ensure universal applicability, consistency of approach and appropriate version control and publication.

While discussing the next version of the model in 2018 (updating the December 2017 version), we also need to decide on the modelling language to be used. So far, the model is published as a simple list (table) of properties, their domain and range, free text definitions and allowed values. During workshops supported by COST Action 15212, we identified a clear need to supplement this list by defining relations between the concepts. A draft version of these relations, together with a number of profiles (the so called *Milano model*) is currently being documented within this COST Action.

In the medium term we will also discuss the possible provision of facilitating tools, such as metadata editors, collaboration tools, citizen science platforms and dedicated training material.

Part II: Publication of models

While continuing to improve the PPSR Common Conceptual Model (concepts, relations, definitions and ranges / allowed values), we see a need to discuss its online publication. This will require agreement on a base URL that can be used and referenced uniquely, and a common URI strategy which allows versioning. A formal decision process needs to be defined by which stable versions of the model can be released.

For the time being, the CSA International Data and Metadata Working Group is the custodian of the model. The model should be governed by a committee of the working group (Model Governance Committee, or MaGiC), which would be responsible for explicitly taking care of the model, its future versions and the communication of updates to the community. In future, as the <u>Citizen Science Global Partnership</u> evolves, this decision might be revisited, especially considering the fourth foundational goal of the CSGP: to "support the collection and exchange of open and findable, accessible, interoperable, and reusable (FAIR) data by formalizing a citizen science data standard, creating a data directory, and building an open data portal."

We commit to an open and royalty-free model, i.e. the model shall remain open to access, (full or partial) re-use and extension/augmentation (new profiles).

We generally consider a GitHub central repository as the most suitable tool currently available to collaborate on the model, including its version management. GitHub already provides a means for collaboration among developers, and supports provenance capture and tracking. As such, it is likely to reduce the time and cost to the custodian for managing and curating the model. The repository should be created and maintained by the abovementioned MaGiC committee. This committee should also define the protocol for version release and dissemination. If standardisation bodies such as ISO, W3C or the OGC wish to endorse or build upon the recommendation, then we will support their institutional processes where possible.

The associations that endorse and promote the model (above all, citizen science associations) may then provide targeted descriptions and references to the published recommendation on their respective platforms. Simplified descriptions of the model for a variety of audiences and automatic synchronisation among platforms will be considered.

Part III: Content discussion

As a first revision of the <u>progress report that was published in December 2017</u>, including a version of the Common Conceptual Model, we suggest reviewing and integrating the latest **proposal from COST Action's WG 5 addressing the Project Metadata Model (PMM) and**

the Dataset Metadata Model (DMM). This would involve agreement on attribute (properties) names, descriptions and allowed values, and on the means by which they are packaged into core information and usage-dependent profiles.

The concrete proposal of the COST Action (and any future proposal) will:

- ensure backward compatibility with the current version of the PPSR Common Conceptual Model;
- provide a full documentation and argumentation in order to justify each change proposed: for each concept the documentation will at least include: a reason/issue (something wrong or a new requirement), proposed change, argumentation for the decision;
- be subject to revision by the MaGiC committee (see Part II) before adoption as the new version of the PPSR Common Conceptual Model.

According to the suggestions in Part I, future extensions (new concepts, profiles and agreed vocabulary terms) will depend on user requirements and priorities. These include trajectories of citizen science developments related to projects whose outputs are not 'data', and projects that focus on study design, data transformation, building of physical objects, text annotations, and mathematics. It is anticipated that the Observation Data Model may be insufficient to capture all outputs of citizen science, and that an additional model, interoperable with the Project and Dataset Metadata Models, may be required in future for some products. This possibility will be actively investigated as new use cases emerge.

Part IV: Facilitating uptake

The decisions reported in this document on the approach to advance the development, publishing and maintenance of the Common Conceptual Model must be complemented with dedicated actions to allow the uptake and re-use of the results by parties that are not directly involved in the development activities of the standards for citizen science data and metadata. For this purpose, a strategic note on uptake strategy will be prepared. This strategy will be put out for discussion and will be iteratively tested and updated in order to allow for uptake of this recommendation by the wider citizen science community, and by interested standardisation bodies.

For the wider citizen science community, the uptake strategy will include examples, provide an overview of communication channels, and contain simple, targeted messages for putting the results of the standardisation work into use. We anticipate a need to use multiple communication channels to spread the word about the Common Conceptual Model and encourage its uptake. Tools, APIs and services will be required to facilitate use by a range of project participants, project managers, data users and data publishers.

We will also actively seek support and buy-in from standardisation bodies so that they can contribute, and do not propose contrary/alternative non-interoperable solutions. To achieve this we will engage with initiatives such as the *OGC Interoperability Experiment on Citizen Science* (CitScilE) 2018. In CitScilE we seek a testing ground and a proof-of-concept for

practical use of the Project Metadata Model and the Dataset Metadata Model, including through the coordinated Earth Challenge 2020 citizen science initiative. In addition, we will explore how far the Observation Data Model might be covered by existing OGC standards and how these standards might be extended or connected with other relevant models, e.g. from the social sciences. Based on the findings from this experiment, we envisage dedicated work from the standardisation track of the OGC in order to formalise the proposed citizen science recommendation within the OGC community.

Part V: Agreements

The following agreements have been reached:

- Publication of this document as the "Geneva Declaration on Citizen Science Data and Metadata Standards" on June 13, 2018
- Publication of a simple and short article describing our work and decisions to the wider citizen science and standardisation community (deadline: June 19, 2018)
- Submission, by COST Action's WG 5, of a formal proposal with its latest results to the CSA International WG on Citizen Science Data and Metadata (deadline: July 13, 2018)
- Discussing the alignment of the next version of the Project Metadata Model (PMM) and the Dataset Metadata Model (DMM) to the proposal of COST Action's WG 5 (deadline: August 2018)
- Collaboration with the approved Open Geospatial Consortium Interoperability Experiment
- Creation of the Model Governance Committee, or MaGiC, within the CSA International WG on Citizen Science Data and Metadata during the next meeting of the CSA International WG on Citizen Science Data and Metadata (proposed members: Greg Newman, Luigi Ceccaroni, 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
- Publishing an updated version of the PMM and DMM on GitHub (deadline: August 31, 2018)
- Planning for using and testing the proposed models (e.g., in context of Earth Day 2020, and the OGC CitSci Interoperability Experiment)

Part VI: Roadmap to move ahead

In addition to the above actions agreed in Geneva, the following discussion points are suggested as additional considerations for the working group:

• Procedures for responding to existing regulatory/legal frameworks related to citizen science, such as (in Europe) the implementation of the INSPIRE Directive and the provision of related best practices and tools

- Practical implementation of the model: involve ECSA, CSA, ACSA, and CSGP in the definition of an agenda for model implementation, and possibly as hosts for interoperable catalogues of citizen science projects and data (or to provide guidelines on use of existing solutions for this purpose)
- Including a dedicated section in the websites of ECSA, CSA, ACSA, the COST Action and CSGP, explaining this initiative and providing introductory information
- The need to be open to extensions/profiles related to more diverse outcomes such as policy impact, mathematical theorems, hardware, societal impacts
- The need of communication pieces to help practitioners navigate through the various standards and profiles easily (e.g., a 'choose your own adventure' approach) (see also the <u>Digital Curation Centre</u> as an inspiration)

Related efforts that might be of interest:

Metadata standards catalogue of the Digital Curation Centre (DCC) <u>http://www.dcc.ac.uk/drupal/resources/metadata-standards</u>

BioPortal - Ontology Catalogue in the biodiversity domain with search and annotation functionality

https://bioportal.bioontology.org/

A developing ontology for SDGs, one of the arenas we can hook into to describe impacts & relevance of citizen science projects

https://github.com/SDG-InterfaceOntology/sdgio

Annex: Key partners and contributors

The Citizen Science Association International Data & Metadata Working Group

In 2015, the U.S.-based Citizen Science Association (CSA) founded a Data and Metadata Working Group. The purpose of the CSA Data and Metadata Working group is to support, advance, and facilitate data interoperability among and between citizen science projects and other data repositories; and to promote collaboration in citizen science via the development and/or improvement of international standards for data and metadata. The goals of the CSA Data and Metadata Working Group are to coordinate global efforts towards data and metadata interoperability and to ensure that all relevant stakeholders are included in these discussions.

The European Citizen Science Association

The European Citizen Science Association (ECSA) is a non-profit association set up to encourage the growth of the citizen science movement in Europe in order to enhance the participation of the general public in scientific processes, mainly by initiating and supporting citizen science projects as well as performing research on citizen science. ECSA is framing citizen science as an open and inclusive approach, for example by supporting and being part

of the exploration, shaping and development of the different aspects of the citizen science movement.

The Australian Citizen Science Association

The Australian Citizen Science Association (ACSA) was formed to advance citizen science in Australia through the sharing of knowledge, collaboration, capacity building and advocacy. ACSA is a member-based community that supports, informs and develops citizen science. ACSA, including its members and partners, has a keen interest in ensuring that citizen science projects and data generated by projects are as open, accessible and re-usable as possible and is therefore committed to participating in this working group and to the objectives and outcomes from it.

The Citizen Science Global Partnership

The Citizen Science Global Partnership (CSGP) is a new partnership, which aims to become a network-of-networks that seeks to promote and advance citizen science for a sustainable world at the global level, with five initial goals: 1. become a valued partner and common point of entry for NGOs, governments, businesses, and scientific organizations seeking to understand and work with the global citizen science community; 2. network with and help support citizen science umbrella organizations, including the Citizen Science Association (CSA), the European Citizen Science Association (ECSA), the Australian Citizen Science Association (ACSA), emerging associations in other geographic regions, and related networks or communities; 3. network with and help support or manage consortia of citizen science projects conducting research around a specific topic of interest; 4. help catalogue, manage and share open and interoperable data for use in local-to-global research and decision-making; 5. understand and track the contributions of citizen science towards the sustainable development goals (SDGs).

The Open Geospatial Consortium

The Open Geospatial Consortium (OGC) is an international not for profit organization committed to making quality open standards for the global geospatial community. These standards are made through a consensus process and are freely available for anyone to use to improve sharing of the world's geospatial data.

The European COST Action

The main aim of COST Action CA15212 "Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe" (2016-2020) 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. The Action explores the potential transformative power of citizen science for smart, inclusive, and sustainable ends. The relevance of the Action derive from the recent, sudden increase of activity around citizen science, as ordinary people and researchers begin to understand the power of technological devices, which allow them to record the environment around them and share and collectively interpret data and knowledge to advance science and society.

Next meeting

The next meeting will be in Riga later in 2018 (to be confirmed), and the following one in Enschede, Netherlands, on March 21-22, 2019.

