Citizen Science and Crowdsourcing Metadata Workshop Summary

Thursday July 9th & Friday July 10th, 2015

Prepared by: Anne Bowser, Rachel McMonagle & Elizabeth Tyson

Introduction

Citizen science is a form of open collaboration where members of the public participate in the scientific process to address real-world problems.[1] As a rich historical tradition, citizen science is gaining increased acceptance and momentum. Supporting associations are convening in the United States, in Europe, and in Australia. Within the U.S., there is unprecedented support for citizen science from communities, universities, and government.[2]

Growth brings challenges and opportunities. Researchers are beginning to study citizen science as a paradigm to understand volunteer motivation,[3] identify project outcomes,[4] and help navigate barriers such as the Paperwork Reduction Act.[5] If this research will be valuable, it needs to be discoverable. Similarly, if practitioners will be able to network to share lessons learned and data, they need to be able to find projects similar to their own.

Four databases of citizen science and crowdsourcing projects — hosted by SciStarter, the Citizen Science Association (CSA), CitSci.org, and the Woodrow Wilson International Center for Scholars (the Wilson Center) — are working on a common *project metadata schema* to support data sharing with the goal of maintaining accurate and up to date information about citizen science projects.[6] The federal government is joining this conversation with a cross-agency effort to promote citizen science and crowdsourcing as a tool to advance agency missions. Specifically, the White House Office of Science and Technology Policy (OSTP), in collaboration with the U.S. Federal Community of Practice for Citizen Science and Crowdsourcing (FCPCCS),

[1] Definition borrowed from the U.S. Federal Community of Practice for Crowdsourcing and Citizen Science (FCPCCS), http://www2.epa.gov/innovation/federal-community-practice-crowdsourcing-and-citizen-science.

[2] For example, the second Open Government National Action Plan. https://www.whitehouse.gov/sites/default/files/docs/us_national_action_plan_6p.pdf.

[3] For example, Rotman, D. Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D. & Jacobs, D. (2012), Dynamic changes in motivation in collaborative citizen-science projects. Presented at the *2012 ACM Conference on Computer Supported Collaborative Work*. Seattle, WA, 11-15 February 2015.

[4] For example, Phillips, T.B., Ferguson, M., Minarchek, M., Porticella, N. & Bonney, R. (2014). *User's Guide for Evaluating Learning Outcomes in Citizen Science*. Ithaca, NY: Cornell Lab of Ornithology.

[5] Gellman, R. (2015). *Crowdsourcing, Citizen Science, and the Law: Legal Issues Affecting Federal Agencies*. Washington, DC: Woodrow Wilson International Center for Scholars.

[6] Note that this report describes work to flesh out "project metadata," which explains how citizen science projects are described (e.g., like through the degree of volunteer participation, partnership models, and project outcomes).

DRAFT

are compiling an Open Innovation Toolkit containing resources for federal employees hoping to implement citizen science and crowdsourcing projects. Navigation through this toolkit will be facilitated in part through a system of metadata tags. In addition, the Open Innovation Toolkit will link to the Wilson Center's database of federal citizen science and crowdsourcing projects.

These groups became aware of their complementary efforts and the shared challenge of developing project metadata tags.[7] On July 9th & 10th, 2015, a Citizen Science and Crowdsourcing Metadata Workshop was hosted by the Wilson Center in coordination with the U.S. Environmental Protection Agency (EPA) and the FCPCCS. The overarching goal of this workshop was to standardize metadata tagging systems used by organizations including OSTP, the Wilson Center, the CSA, SciStarter, and CitSci.org, to support data sharing and collaboration between these organizations. Specific goals included:

1. Revisit existing metadata and tagging systems that describe citizen science projects, with the aim of selecting and defining a list of agreed upon "core"[8] metadata tags. These tags may be applicable to a broad range of projects considered citizen science and crowdsourcing, within and outside of government. *Desired Outcome: A list of agreed upon "core" fields that will be implemented by each partnering database.*

2. Determine the degree and method of data sharing between different collaborating organizations. This will include a discussion of technical implementation in the form of APIs, and a plan for ensuring that shared information remains accurate and up to date. *Desired Outcome: A draft letter of intent outlining plans for data sharing and sustainability.*

3. Support a general awareness of the accelerators and barriers to implementing citizen science and crowdsourcing projects within and outside of federal agencies, with the aim of preventing duplication and supporting future collaborations between sectors. *Desired Outcome: Enhanced knowledge and networks of workshop participants.*

[7] This was identified based on work done previously by DataOne.org and expanded upon during a workshop led by SciStarter on defining taxonomy of fields in 2014.

[8] Here, "core" designates data fields that will be shared between the four partnering databases, and also made publically available to the community at large.

Workshop Procedure and Results

Day 1 – July 9th, 2015

The first day of the workshop was dedicated to revisiting existing project metadata tagging systems to expand a list of agreed upon fields, in line with the first workshop goal. The workshop began with a statement of goals and introductions from all participants (listed in Table 1).

Attendees	Affiliation
Representing the FCPCCS	
Jay Benforado, Barbara Martinez, Rachel McMonagle, Alison Parker	U.S. Environmental Protection Agency (EPA)
Jennifer Couch, Katrina Theisz	National Institutes of Health (NIH)
Ruthanna Gordon	Booz Allen Hamilton, contract support for Intelligence Advanced Research Projects Activity (IARPA)
Heidi Hadley	Bureau of Land Management (BLM)
Stephen Lowe	U.S. Department of Agriculture (USDA)
Duncan McKinley	U.S. Forest Service (USFS)
Representing Database Collaborations	
Anne Bowser, Rohin Daswani, Dave Rejeski, Elizabeth Tyson	The Wilson Center
Darlene Cavalier	SciStarter
Greg Newman	CitSci.org
Jennifer Shirk	The CSA
Bringing Key Citizen Science Metadata Experience	
Andrea Wiggins	University of Maryland, College Park

Table 1. Workshop participants

This was followed by brief presentations from the four project databases and the Innovation Toolkit development team. Each presenter gave a brief explanation of their project and, if applicable, a technical demonstration.

DRAFT

Each organization expressed different, but complementary goals:

- SciStarter's mission is to support citizen science volunteers, by offering a centralized portal for participation in citizen science.
- The Citizen Science Association's project database is designed to support networking between citizen science and practitioners.
- CitSci.org hosts information about projects making use of the citsci.org platform – a platform that allows project leaders to create and host their own citizen science projects and customize their own data collection, management, visualization, and analyses.
- The Wilson Center collects information about citizen science projects supported by federal agencies.
- The OSTP Innovation Toolkit's goal is to provide Federal agencies with best practices, training, policies, and guidance on authorities related to open innovation, including approaches such as incentive prizes, crowdsourcing, and citizen science.[9]

Prior to the workshop a list of 75 project metadata tags was compiled by combining the fields from SciStarter's database, with tags designed for the OSTP Innovation Toolkit with PPSR-Core v8.[10] The workshop began with a sorting session, where each tag was collaboratively labeled "easy," "complex," or "triage."

Easy tags were defined as simple and non-contentious (e.g., "projectURL"). Complex tags, such as "projectDescription," were considered tags that merited discussion—in this case, because different databases may collect descriptions of different lengths, containing different information. Project metadata tags were triaged if they were relevant only to a small set of projects, or if they were too complicated to finalize in a short workshop (e.g., disambiguating between characterizations of the geographic scope of activity).

Following this triage exercise, participants formed groups of 3 or 4. Easy terms were divided between the groups, and defined through a three-step process:

1. Each group wrote a tentative definition for each term.
2. Groups exchanged terms for a round of peer review, where the initial definition was often modified.

[9] Second Open Government National Action Plan, 2013

[10] PPSR-Core is a list of standardized project metadata tags that was first supported by DataONE. The PPSR-Core list of tags went through eight iterations prior to this workshop. Note that many metadata tags from each source incorporated existing research, for example typologies of citizen science including: Wiggins, A. & Crowston, K. (2012). Goals and tasks: Two typologies of citizen science projects. Presented at the *Forty-fifth Hawai'I International Conference on System Science.* Wailea, HI, 4-7 January, 2012.

3. Following peer review, all terms were shared with the larger group for a three-minute discussion. If consensus was reached in three minutes, the term was considered finalized; if not, it was set aside for later work.

Following a lunch break, this process was repeated with the complex terms. Note that during the duration of the workshop, important issues were documented on a public whiteboard for later discussion.

Day Two – July 10th, 2015

The second day of the workshop was dedicated to fleshing out issues around data sharing, in line with the second workshop goal. At the beginning of the second day, each of the five stakeholders was given five minutes to define their data sharing agreements or needs. There was broad consensus around the following points:

- Prior to the workshop, SciStarter, the CSA database, CitSci.org, and the Wilson Center reached a landmark agreement to share seven "core" fields.[11] Data sharing should be expanded beyond these seven fields.
- Sharing project data will support the sustainability of each database by ensuring that each is kept accurate and up to date.
- Yet, it may not be appropriate to share all data. For example, information collected that relates to current research projects may not be shared prior to publication.
- Additionally, there is a need to distinguish between sharing project data and publishing project data.[12]
- APIs must be developed to facilitate project data sharing.
- Stakeholders outside the room, including collaborators, lawyers, and other federal employees, will need to be consulted before data sharing agreements are finalized.
- There is an opportunity for private sector workshop attendees to provide feedback on the OSTP Innovation Toolkit.
- Fleshing out project metadata will facilitate the research and practice related to citizen science. In addition to a Citizen Science Association designed to support practice, there may be a need for a National Center, and/or regional centers to support excellence in academic research.[13]

[11] List of fields may be found in: The Citizen Science Association. (2015, June). Efforts underway to share project metadata. http://citizenscienceassociation.org/2015/06/11/efforts-underway-to-share-project-metadata/.

[12] For example, not all projects are active and seeking volunteers; thus, while they may exist in the SciStarter database it would not be appropriate for SciStarter to publish these projects.

[13] In line with Bonney, R. Shirk, J., Phillips, T., Wiggins, A., Ballard, H., Miller-Rushing, A. & Parrish, J. (2014). Next steps for citizen science. *Science, 28,* 6175, 1436-1437.

- To prevent disciplinary and geographic silos from forming, any discussion of citizen science project metadata should reach an international audience. This will be accomplished through the formation of a CSA working group (described later).

Participants split into groups to discuss the needs articulated in this discussion, and issues brought forth on the first day. Discussions were structured with the goal of outlining action items for moving forward in each area. The results of these discussions are described below.

Outcomes and Next Steps

Participants reached several conclusions, some paired with an action point for moving forward:

1. **There is a need to demonstrate proof of concept project metadata sharing** to illustrate the value of this work, and the willingness of different databases to collaborate.

2. **Existing databases need to determine their terms of use for sharing project metadata**. Collaborators agree that a "core" set of fields will be made publically available, but have yet to agree on the terms for sharing these core fields, or exactly what these core fields will be. In addition to these core fields, databases may share additional information with select collaborators. All data sharing should be transparent.

   Action item: The value of project metadata sharing will be illustrated through use cases written by Ruthanna Gordon. A demonstration of project metadata sharing will be led by Greg Newman and the CitSci.org team, who will expand upon an existing API currently used to exchange data between CitSci.org and SciStarter to implement APIs to share seven basic metadata fields between SciStarter, the CSA, CitSci.org, and the Wilson Center. These APIs and documentation (including a terms of use agreement for core fields) will be published on the CSA website and SciStarter. Each database will also write an individual "terms of use" statement describing additional data sharing policies.

3. Participants reached consensus on 30 fields[14], presented in the Appendix. However, **existing project metadata schemes, including PPSR-Core v9, require more work**. Although important progress was made, many fields are still elusive. This development will require a broad examination of existing work on metadata standards in related areas, such as the Informal Science Education metadata[15] published by the Center for the Advancement of Informal Science Education (CAISE). It will also require referencing current research on citizen science, for example through examining the articulation of

[14] These fields were based in part on the SciStarter database
[15] See http://informalcommons.informalscience.org/wiki/index.php/IC_Metadata_v_1.0.

project outcomes included in the Developing, Validating, and Implementing Situated Evaluation (DEVISE) project scales.[16]

4. In addition, **the unstructured list of current metadata tags needs to be expanded into a formal ontology**, offering definitions of project metadata terms and illustrating the relationships between different project metadata terms. A formal ontology will collect and make explicit assumptions about citizen science, supporting research, allowing the tagging of resources, and facilitating data sharing.[17] As part of this effort, appropriate standards for metadata in different fields of science (e.g., the Biological Data Profile; the Ecological Markup Language) need to be identified and linked to a formal ontology of concepts related to citizen science.

   Action item: Prior to the workshop, a number of participants published their intent to form a CSA data and metadata working group. The intent and rationale for this working group was presented at the recent meeting of the Australian Citizen Science Association (ACSA). [18] As a result, Anne Bowser recruited two members of European Citizen Science Association (ECSA) to act as European liaisons for this work, and two members of ACSA to act Australian liaisons. The U.S. CSA will sign a Memorandum of Understanding (MOU) with ECSA and ACSA to formalize this collaboration, with input and agreement from SciStarter, CitSci.org and the Wilson Center. Working group members from each association and partner databases such as SciStarter.com will collaborate on a funding proposal to support future work.

5. Workshop participants determined that a number of the agreed upon project metadata fields may not apply to crowdsourcing projects. Thus, **project metadata for the field of crowdsourcing needs more discussion.**

6. As a precursor to this discussion, **precise definitions of citizen science and crowdsourcing need to be formulated** (Figure 2 on page 9 depicts three of these concepts that are currently represented in the literature). This will also require

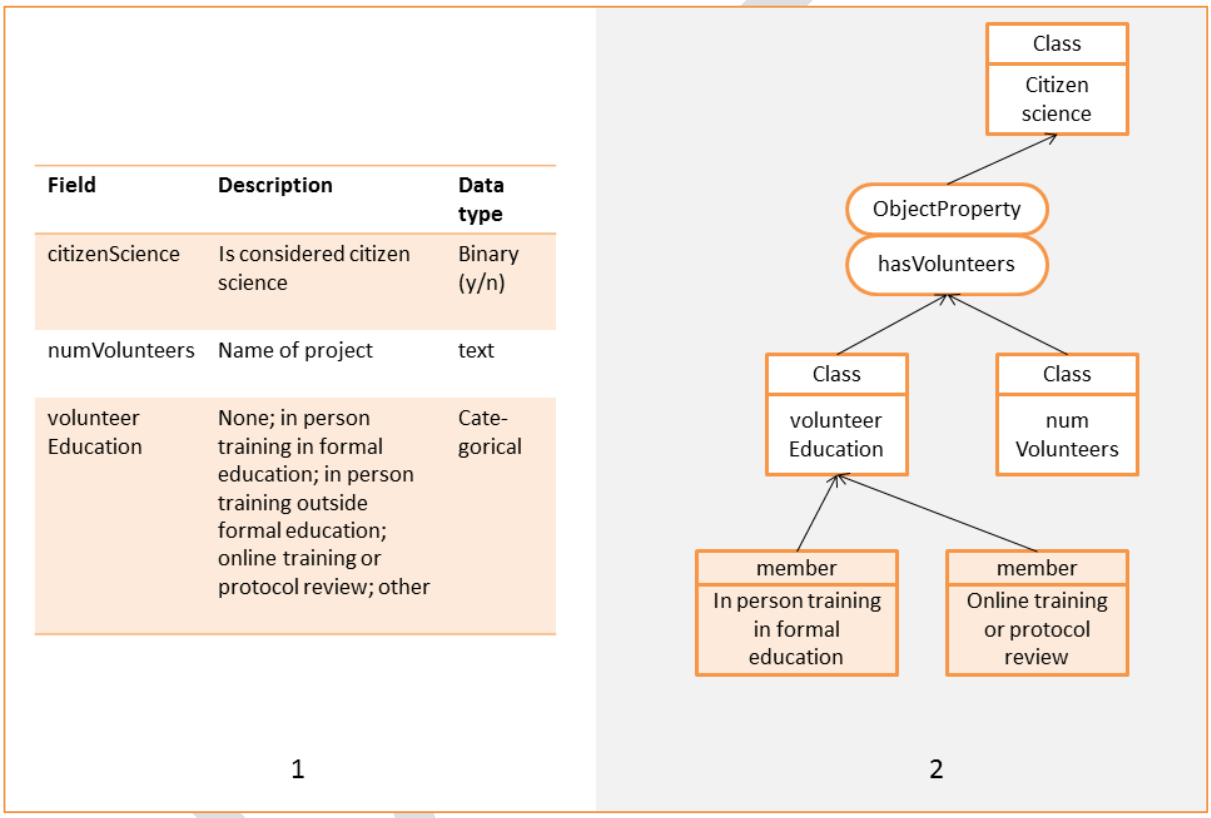
[16] Phillips et al., 2014.

[17] Noy, N. McGuinness, D. (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford, CA: Stanford University.

[18] Newman, G., Shirk, J., Cavalier, D. & Bowser, A. (2015). Introducing PPSR-Core: Standardizing metadata to support a global community. Presented at the *Australian Citizen Science Conference*. Canberra, Australia, 23-24 July 2015.

describing how these concepts map to related terms like participatory sensing, volunteered geographic information, and public participation in scientific research.[19]

Action item: This work could begin with a second workshop, where participants bring case studies of citizen science and crowdsourcing to discuss how these terms relate to one another, and how crowdsourcing can be unique. After definitions are fleshed out, work on constructing crowdsourcing project metadata may begin.



Field	Description	Data type
citizenScience	Is considered citizen science	Binary (y/n)
numVolunteers	Name of project	text
volunteer Education	None; in person training in formal education; in person training outside formal education; online training or protocol review; other	Cate-gorical

Figure 1. The difference between unstructured metadata (1) and a simplified formal ontology (2). An ontology illustrates the relationships between key terms, and provides opportunities for adding new terms to the matrix.

[19] Researchers conducted semantic analysis to examine the links between publications using terms such as "citizen science," "crowdsourcing," and "volunteered geographic information." They found a closely linked matrix in the 1990s and early 2000s, compared to greater topical distances in later years. Comber, A., Schade, S., See, L., Mooney, P. & Foody, G. (2014). Semantic analysis of citizen sensing, crowdsourcing and VGI. Presented at the *International Conference on Geographic Information Science*. Castellon, Spain, 3-6 June 2014.

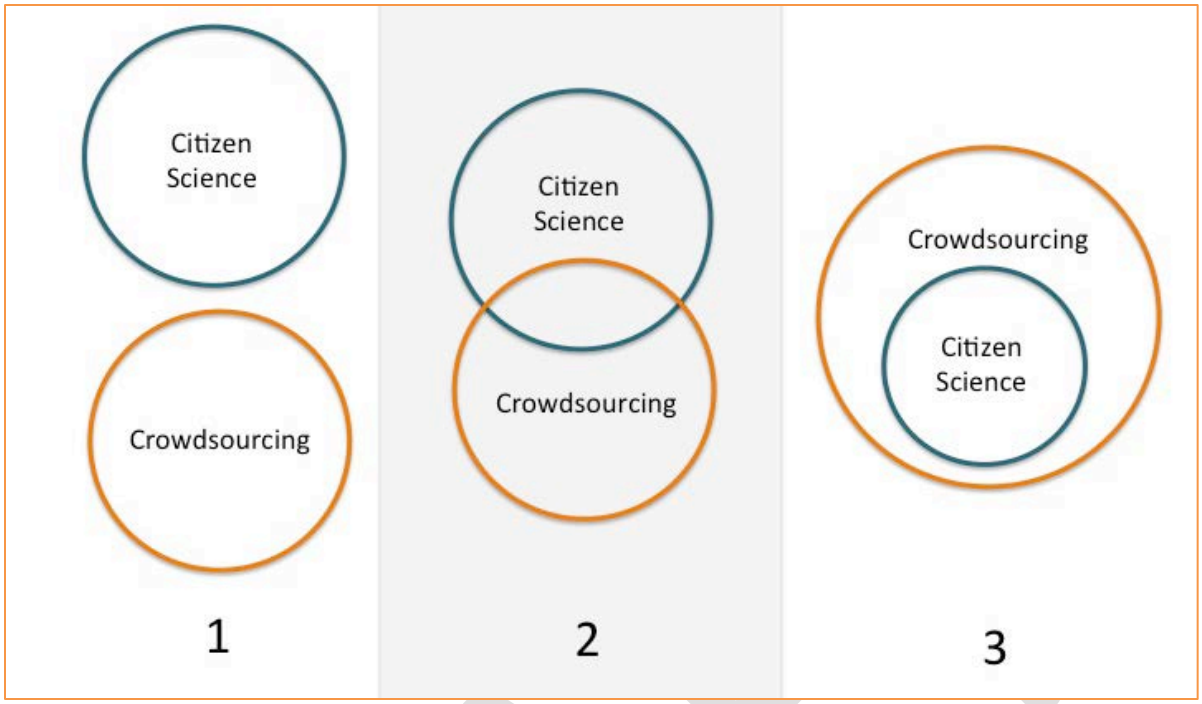


Figure 2. Common portrayals of the relationship between crowdsourcing and citizen science. In the first model, citizen science and crowdsourcing are portrayed as separate concepts with an ambiguous or undefined relationship[20]. In the second model, there is some overlap between crowdsourcing and citizen science; for example, crowdsourcing may be portrayed as "one of the possible approaches to supporting citizen science."[21] In the third model, one paradigm—typically citizen science—is subsumed in the other.[22]

Conclusion

On July 9th & 10th, 2015, a Citizen Science & Crowdsourcing Metadata Workshop was hosted by the Wilson Center in coordination with the U.S. EPA and the FCPCCS. The overarching goal of this workshop was to standardize metadata tagging systems used by organizations collecting information about citizen science and crowdsourcing projects, including OSTP, the Wilson Center, the CSA, SciStarter, and CitSci.org, to support data sharing and collaboration. The results of this workshop included 30 fields describing citizen science project metadata (Appendix). Seven of these fields will be applied in a demonstration of data sharing between four databases. However, the current list of project metadata tags requires expansion, especially around concepts related to crowdsourcing.

[20] This is often implicit. For example, the FCPCCS presents citizen science and crowdsourcing as separate terms on a list of different types of open innovation.

[21] Craglia, M. & Shanley, L. (2015). Data democracy- Increased supply of geospatial information and expanded participatory processes in the production of knowledge. *International Journal of Digital Earth,* DOI:10.1080/17538947.2015.1008214.

[22] For example, Zhao, Y. & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future directions. *Inf Syst Front, 16*, 417-434.

Workshop participants also believe that the current list should be organized into a formal ontology. A standardized ontology will allow databases of citizen science and crowdsourcing projects to share information about these projects. Such information will enable volunteers to find projects to contribute to from anywhere in the world, and practitioners to network with similar researchers. It will reduce the information burden placed on users, by ensuring that a record updated in a single database will be shared with all collaborating databases.

Formalizing a citizen science ontology will also make domain knowledge about citizen science and crowdsourcing centralized and explicit. This will support research on citizen science by articulating key characteristics of these paradigms, such as intended audiences and project outcomes. Similarly, such an ontology will allow for keyword tagging—for use in journal articles, shared bibliographies, etc.—to help others find resources and ultimately reduce duplication.

Finally, an ontology of project metadata can be linked to other existing standardized metadata, and therefore used to describe citizen-generated data sets. This will enable researchers to find and reuse citizen science data. Given that one open question asks "whether there is enough information about the data for professional scientists to know that they can trust and use them," metadata will also be important for describing protocols such as quality assurance and validation strategies that illustrate careful research design.[23] Thus, good project metadata describing research design may one day enable formal recognition or endorsement of data quality.

This work will be continued by an international working group hosted by the CSA, in collaboration with ECSA and ACSA.[24] The goal of this working group is to develop a standardized global ontology for citizen science, describing project metadata that may be linked to data sets. This work will help citizen science continue to evolve into a standardized tool utilized across disciplines to promote scientific research, support science education, and advance public policy.

DRAFT

[23] Williamson, K., Kennan, M.A., Johanson, G. & Weckert, J. (2015). Data sharing for the advancement of science: Overcoming barriers for citizen science. *JASIST,* DOI: 10.1002/asi.23564

[24] With input from SciStarter, CitSci.org and the Wilson Center

Appendix: Project Metadata Fields

Field name	Description	Data type	Controlled vocabulary
Basic information and provenance			
projectName	Name of project	text	-
projectGUID	Globally unique identifier	text	-
projectData Provider	Name of data provider/ source of first registry	text	-
projectDate LastUpdated	Date the project information was last updated in any database	text	ISO 8601 Date Time
Project personnel			
database Contact	Person that should be contacted to update the database record	Person object or text	-
publicContact	Person that interested public or researchers should contact	Person object or text	-
publicContact Affiliation	Public contact primary affiliation	text	-
publicContact email	Public contact email address	text	-
publicContact phone	Public contact phone number	text	-
publicContact address	Public contact mailing address	text	-
Web presence and social media			
projectURL	URL for project homepage	text	-
Project Facebook	URL for project Facebook page or group	text	-
projectTwitter	Project Twitter handle	text	-
projectBlog	URL for project blog	text	-
preferred SocialMedia	Preferred social media outlet (may or may not be listed above)	text	-

DRAFT

Field name	Description	Data type	Controlled vocabulary
Project description			
project Description	Description of project related to goals, objectives, purpose, vision, etc.	text	-
projectImage	Image	Image	-
projectStatus	Current status of the project activity	text	Active; active but seasonal; complete; hiatus; pending
fieldOfScience	General field of science	text	Top level categories defined by National Academes
Project metrics			
num Participants	Total number of registered participants	text	-
active Participant definition	Definition used by the project for an “active participant”	text	-
numActive participants	Total number of active participants based on project’s definition of activity	text	-
Participation considerations			
participation Tasks	Participation tasks	text	Observation; identification; classification or tagging; annotation; transcription; problem solving; data entry; finding entities; measurement; specimen/sample collection; sample analysis; site selection and/or description; geolocation; learning; photography; audio or video recording; data analysis; DIY hacking/ making; other
participantAge	Age of participants (if known)	text	-
participant Audience _SciStarter	Target participants (defined by user research with SciStarter.com)	text	General public; families; elementary school children; middle school children; teens; college; graduate students

DRAFT

Field name	Description	Data type	Controlled vocabulary
participant Education	Type of education required for participation	text	None; education through online training or protocol review; in-person training in formal education; in-person training outside of formal education; other
projectGear	Essential equipment for participation (tools, gear, field equipment, apps, Internet access, etc.)	text	-
educator Materials	Link(s) to project's teacher guides and/or note availability in print form	text	-
Supporters and partners			
Project Institution Name	Institution or organization that is hosting the project	text	-
Research and data management			
projectDataAPI	url or API for current data	text/url	-

DRAFT