Development of Standards for Metadata Documentation in Citizen Science Projects

Lizet Doriela Mantari Mincami1,*, Hilario Romero Giron1, Edith Mariela Quispe Sanabria1, Luis Alberto Poma Lago1, Jose Francisco Via y Rada Vittes1, Jessenia Vasquez Artica1, Linda Flor Villa Ricapa1

1Universidad Peruana Los Andes (UPLA), Perú, *Corresponding author: d.lmantari@upla.edu.pe

Abstract

Introduction: Citizen science has generated large volumes of data contributed by citizens in the last decade. However, the lack of standardization in metadata threatens the interoperability and reuse of information.

Objective: The objective was to develop a proposal for standards to document metadata in citizen science projects in order to improve interoperability and data reuse.

Methods: A literature review was conducted that characterized the challenges in metadata documentation. Likewise, it analyzed previous experiences with standards such as Darwin Core and Dublin Core.

Results: The review showed a high heterogeneity in the documentation, making interoperability difficult. The analyzes showed that standards facilitate the flow of information when they cover basic needs.

Conclusions: It was concluded that standardizing metadata is essential to harness the potential of citizen science. The initial proposal, consisting of flexible norms focused on critical aspects, sought to establish bases for a collaborative debate considering the changing needs of this community.

Keywords: Citizen science; Metadata; Standards; Interoperability; Data reuse; Collaboration.

Received on 8 January 2024, accepted on 15 April 2024, published on 24 April 2024

Copyright © 2024 Mincami et al., licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ettsis.5704

1. Introduction

In the last decade, Citizen Science has emerged as a transformative paradigm that redefines the traditional dynamics between the scientific community and society at large (1) (2) (3). This innovative approach not only democratizes access to scientific research, but also empowers non-specialized individuals to actively and valuably contribute to major scientific projects (4) (5) (6). This collaborative revolution has generated a substantial flow of data, fueled by the enthusiastic participation of a diverse network of citizen contributors ranging from passionate amateurs to amateur experts (7).

Despite significant achievements, Citizen Science faces a crucial challenge related to heterogeneity in the documentation of metadata in its projects (8). The diversity in capturing and describing this metadata has resulted in a sometimes complex and disorganized information landscape (9), as "the variety in metadata coding generated a 'tower of babel' of data that is difficult to integrate" (10). This patchwork of information, while reflecting the richness and breadth of citizen science projects, has raised the imperative need to establish clear standards for metadata documentation, as it limits the combination of results across initiatives (11) (12). Also, it is crucial to understand these tendencies within multiple social networks and digital spaces (13) (14) (15).

This brief communication seeks not only to highlight the relevance of developing these standards, but also to contextualize the critical importance of this step in the continued advancement of Citizen Science. The diversity and breadth of information generated by citizen projects are valuable assets that, without adequate standards, and given that "the adoption of common standards will allow harnessing the scientific potential of projects developed globally in a decentralized manner", risk losing their usefulness and coherence (16) (17) (16).
In the course of this paper, the intrinsic motivations driving the need for robust standards are explored (19). A brief analysis of past experiences related to the definition of standards is also conducted, highlighting valuable lessons that can inform the way forward. In addition, an initial proposal is presented that is intended not only as a starting point, but also as a catalyst for discussion and collaboration in this critical area of Citizen Science (20) (21).

Understanding the importance of standards for metadata documentation proves to be a cornerstone for the continued progress of Citizen Science. It will not only strengthen the integrity and utility of citizen data, but also catalyze more effective collaboration between citizen participants and professional scientists. In this journey, Citizen Science is positioned not only as a means to maximize the potential of scientific research, but as a powerful tool for active participation and collective knowledge building (22) (23) (24).

2. Methods

This study corresponds to a literature review based on editorial features (25) (26), also known as a literature review or background review, which is a systematic process of collecting, evaluating and synthesizing existing research on a specific topic (27) (28) (29). It consists of identifying, analyzing and synthesizing relevant previous research and findings related to the topic of interest (30) (31). Within this context, a set of sources was consulted and analyzed to meet the objectives set out in the study, thus characterizing the challenges in metadata documentation (32) (33) (34). Previous experiences with standards such as Darwin Core and Dublin Core were also analyzed.

In addition, this study was supported by propositional research, which is a research approach that focuses on generating solutions, proposals or recommendations to address specific problems or improve a given situation (35). In that sense, this is a research approach that focuses on generating solutions, proposals or recommendations to address specific problems or improve a given situation. Unlike descriptive or exploratory research, whose main objective is to understand and describe phenomena, propositional research is oriented towards action and the practical application of the knowledge acquired. Therefore, a proposal of standards for documenting metadata in citizen science projects is developed.

In this study, the documentary analysis technique was used, which implies an intellectual process through which the information in the document that is relevant for its representation is selected (36) (37). For the literature review, classic and new sources related to the subject were considered; in addition, the file and the computer and its storage units were used as instruments (38) (39).

3. Results

Current Status of Metadata Documentation in Citizen Science

The exponential expansion of Citizen Science over the past decade has resulted in an extraordinary diversity of projects, ranging from biodiversity observation to environmental monitoring. "Citizen Science has experienced exponential growth, with hundreds of projects in diverse subject areas capturing observations from thousands of volunteers." (40). This wealth of initiatives has generated a vast dataset, fuelled by the active participation of passionate individuals and engaged citizens (41) (42). "This multifaceted set of initiatives has generated a massive stream of data generated by dedicated citizens."(43) However, "the transformative potential of this unprecedented amount of information is limited by inconsistencies in metadata documentation."(44) The lack of standardization in metadata documentation proves to be a critical issue that overshadows the full potential of these projects (45) (46) (47) (48).

Heterogeneity in data description, which includes information such as location, date, and context, has created a complex landscape for Citizen Science, as "the lack of uniformity in metadata has created a patchwork of information that reduces the possibility of combining data from different initiatives."(49) This diversity, while reflecting the breadth and vitality of projects, introduces substantial challenges for interoperability and effective data integration across projects, as "the heterogeneity of contextual data has created a 'tower of Babel' that hinders interoperability" (10) and "the diversity in metadata capture currently makes it difficult to integrate information from independent projects".(50)

Current Challenges and the Need for Standards

The lack of clear standards for metadata documentation in citizen science projects poses multifaceted challenges. "Lack of metadata consistency limits the ability to connect and contrast relevant data across initiatives." (17) First, it makes it difficult to compare and combine data across projects (51) (52), eroding the ability of researchers to conduct comprehensive analyses that transcend the boundaries of individual projects, as "the lack of homologation hinders retrospective work that progressively integrates results from different origins."(10) This obstacle compromises the ability to gain broader, contextualized insights, as "the inherent complexity of data interoperability makes it difficult to extend conclusions beyond the scale of single projects." (53)

Second, the lack of standards directly impacts effective data reuse, a fundamental principle in contemporary scientific research (54). "Data access and reuse is key to the progress of science, but is undermined by the lack of documentation and standards."(55) Reuse makes the most of past collective efforts, avoiding unnecessary duplication of effort and promoting faster and more efficient progress
in research. However, "the lack of standardized metadata creates substantial barriers to the goal of using data secondarily and deriving new knowledge". (56) The lack of consistent standards in metadata documentation stands as a central obstacle to achieving these fundamental objectives (57) (58), since "by standardizing complementary information we facilitate comparing activities and reviewing results holistically". (59)

Past Experiences and Lessons Learned

Previous experiences in defining standards for metadata documentation, especially in scientific and technological contexts (60), have provided valuable lessons that cannot be overlooked. "International standards such as Darwin Core have facilitated the publication and download of millions of biological records." (61) Notable initiatives such as Darwin Core for biodiversity and Dublin Core for web-based resource description have demonstrated the effectiveness of having well-established standards to ensure consistency and interoperability. For example, "The EML standard has made it possible to publish metadata for thousands of ecological assemblages, making it easier to locate and reuse environmental data." (62) In addition, "Standards such as Ecological Metadata Language (EML) promote the sustainable exchange of ecological information". (63) Successful experiences such as OBIS demonstrate the benefits of this approach by allowing massive indexing of records.

Past Experiences and Lessons Learned

In order to respond to this objective of the study, the diagnosis and need to develop a proposal for standards to document metadata in citizen science projects in order to improve the interoperability and reuse of data was investigated. For this purpose, the theoretical basis for such a proposal was explored.

Within this context, when a document is converted into digital format to become part of a collection, it is usually done for one of two fundamental reasons: to preserve or distribute the documentary material (64). On the one hand, various institutions such as libraries, archives and museums have as their primary objective the preservation of their documentary collections for future generations. Their goal is to ensure that the material will endure over time and be accessible to readers for years or even centuries to come. On the other hand, some institutions, even if their primary function is not preservation, wish to make certain documentary material available to their user communities, reaching a growing number of readers, in more remote locations and for as long as possible (65) (66) (67) (68).

Metadata is data that describes and provides information about other data. In essence, it is "data about data". Metadata can help organize, understand, search and manage datasets, digital resources or any other type of information (69) (70). In that sense, metadata makes it possible for an individual to find and understand data by providing information needed to identify which datasets are available for a specific geographic location, or to assess whether a dataset is suitable for particular purposes (71) (72). It is also useful when an already identified dataset needs to be retrieved or acquired, as well as processed and used.

The reasons for implementing metadata and its usefulness are as follows (73):

- Guarantee the protection of documents and ensure their accessibility and availability over time.
- Simplify the understanding of documents.
- Contribute to ensure the authenticity, reliability and integrity of documents.
- Support the management of access, privacy and intellectual property rights for each document.
- Support interoperability strategies through the official incorporation of documents generated in different administrative and technical environments into the system, ensuring their maintenance for the required time.
- Provide the foundation for an effective search.
- Establish logical connections between documents and their context of origin, keeping them structured, reliable and understandable.
- Facilitate the identification of the technological environment in which the digital documents were created or integrated, as well as the management of that technological environment during their maintenance, ensuring their faithful reproduction as authentic documents when necessary.
- Assist in the efficient and successful transfer of electronic documents from one system or platform to another, as well as any possible alternative for preserving them.

For its part, a proposal for standards to document metadata in citizen science projects in order to improve interoperability and data reuse should have the following characteristics:

Clarity and consistency: Standards should be clear and consistent in terms of the metadata elements required, their format and structure. This ensures that the data is easy to understand and use by different users and systems.

Adaptability: Standards should be adaptable to a variety of citizen science projects in different fields and disciplines. They should be flexible enough to meet the specific needs of each project without compromising overall consistency.

Compatibility: Standards should be compatible with other existing standards and protocols in the field of citizen science and data management. This facilitates integration and interoperability between different systems and platforms.

Inclusiveness: They should be inclusive and take into account the diverse needs and perspectives of participants in citizen science projects, including researchers, citizen volunteers and local communities (74) (75) (76) (77) (78).

Ease of implementation: Standards should be practical and feasible to implement in practice, taking into account the technical and resource constraints that citizen science projects may face (79).
Full documentation: They must be accompanied by full documentation that describes in detail each metadata element, its purpose and its use. This helps ensure consistent implementation and a clear understanding of the standards (80).

Updating and maintenance: They should be dynamic and subject to periodic updates to keep up with technological advances, changes in citizen science practices and emerging user needs (81) (82).

Support and training: Resources should be available to provide support and training to participants in citizen science projects on how to effectively comply with metadata documentation standards.

Within this context, the development of the proposed standards for documenting metadata in citizen science projects in order to improve interoperability and data reuse, had seven stages, in which each one contains phases or activities to be developed, as described below:

**Identification of the Context**
This stage involves understanding and defining the environment in which the proposal will be developed. This includes:
- **Project Context:** Understand the purpose and scope of the citizen science project.
- **Needs Identification:** Identify specific metadata documentation needs to improve interoperability and data reuse.
- **Regulatory and policy framework:** Investigate the regulatory and policy framework related to citizen science and scientific data management. This may include government policies, international regulations, data protection regulations, open access policies, among other relevant aspects.

**Review of Existing Standards**
This stage involves analyzing and evaluating the standards and regulations already established in the field of science, technology and data management. This review aims to identify those standards that may be applicable or adapted to improve interoperability and data reuse in citizen science projects. Some key aspects of this stage include:
- **Previous Standards Review:** Research existing standards and best practices in metadata documentation for citizen science projects.
- **Identify Gaps:** Identify areas where existing standards may not fully address project needs.

**3. Definition of Relevant Metadata**
At this stage, the types of information that must be captured and recorded to adequately describe the data generated in these projects must be identified and specified. Metadata is data that provides information about other data, and is critical to understanding, interpreting, and effectively using scientific datasets (83). Some key aspects of this stage include:
- **Requirements Gathering:** Consult with experts and stakeholders to determine what metadata is critical for interoperability and data reuse in the project.
- **Metadata Prioritization:** Prioritize the most relevant and useful metadata to include in the proposal.

**4. Development of the Standards Proposal**

This phase involves the concrete creation of the standards that will be used to collect, organize and describe the relevant metadata of the data generated in these projects. This stage generally follows the definition of relevant metadata and may include the following steps:
- **Format and Structure:** Define the format and structure of the metadata documentation (e.g., XML, JSON, CSV).
- **Metadata Elements:** Specify required metadata elements, such as project title, data description, participant identification, etc.
- **Nomenclature Standards:** Establish nomenclature standards to ensure consistency and understanding of metadata.
- **Translation of requirements into standards:** Metadata requirements identified during the definition stage are translated into specific standards that determine how information related to citizen science project data will be structured and presented (84) (85). This may include the selection of existing metadata formats, controlled vocabularies, ontologies and metadata schemas that best suit the needs of the project (86) (87).
- **Metadata schema development:** Detailed metadata schemas are developed that specify what information should be collected and how it should be organized (88). This may involve the creation of mandatory and optional fields, the definition of hierarchical structures for metadata, and the specification of rules for metadata validation and exchange.

**5. Validation and Feedback**
This phase involves testing the proposed standards and gathering comments and suggestions from users and experts in the field. This stage is critical to ensure that the standards are effective, practical, and appropriately tailored to the needs of the citizen science community. Some key aspects of this stage include:
- **Pilot testing:** Proposed standards are pilot tested using real or simulated data sets. This allows to evaluate the effectiveness of the standards in practice and to identify possible problems or areas for improvement.
- **Expert Review:** Request expert review of citizen science and metadata documentation to validate the proposal.
- **Obtaining Feedback:** Obtain feedback from key stakeholders such as researchers, platform developers and project participants.
- **Identification of problems and areas for improvement:** Any problems or areas for improvement found during testing and feedback are identified and documented. This may include identifying additional metadata fields needed, clarifying instructions or reviewing the structure of the standards.

**6. Implementation and Dissemination**
This phase involves the practical application of the developed standards and the promotion of their use within the scientific and citizen community. This stage is fundamental to ensure that the standards are widely adopted and contribute effectively to improve interoperability and data reuse. Some key aspects of this stage include:
Integration with existing platforms: Metadata standards are integrated into existing software platforms and tools used in citizen science projects. This facilitates the adoption of the standards by seamlessly integrating them into users' workflows and processes (89).

Proposal Deployment: Implement the metadata documentation standards proposal in the citizen science project.

Training and Guidance: Provide training and guidance on how to comply with metadata documentation standards (90) (91).

Outreach: Promote the adoption of the standards through publications, conferences and other outreach activities (92) (93) (94).

7. Continuous Evaluation
This stage involves regular review and monitoring of the performance and effectiveness of the implemented standards (95). This phase is critical to ensure that the standards remain relevant, up-to-date and adequate to meet the changing needs of the citizen science community. Some key aspects of this stage include:

Monitoring and Evaluation: Continually monitor compliance with metadata documentation standards and gather feedback for future improvements (96).

Iteration: Make adjustments and improvements to standards as needed to address new needs and challenges.

Communication of results: Ongoing assessment results are communicated to the citizen science community and other stakeholders to inform progress and changes in standards. This may include publishing assessment reports, organizing workshops or conferences, and participating in professional networks and communities (97) (98) (99) (100).

This proposal should be the result of a collaborative and carefully planned process involving all relevant stakeholders (101) (102). By following these steps and considerations, a solid proposal can be created that improves the interoperability and reuse of data in citizen science projects, which in turn contributes to the advancement of scientific research and knowledge.

Based on lessons learned and a clear understanding of the current need, an initial framework of standards for metadata documentation in citizen science projects is presented. As noted by Fernández-Llamazares and Cabeza note, "defining flexible but consistent standards makes it possible to optimize the value of voluntary data" (103). This proposal seeks to provide not only a flexible but also a robust basis for consistent and meaningful documentation. Key elements addressed by this proposal include project description, geographic location, observation dates and participant information, as recommended by Follett and Strezov (104).

4. DISCUSSION
Citizen science projects have proven to be a powerful tool for large-scale data collection and public participation in scientific research. However, the lack of clear standards for metadata documentation in these projects can pose significant challenges in terms of interoperability, data quality, and reproducibility of results (105). Therefore, the development of standards for metadata documentation in citizen science projects is an important step in addressing these concerns and improving the effectiveness and utility of this research approach (106) (107) (108).

This study focused on the development of a proposal for standards for metadata documentation in citizen science projects, with the objective of improving interoperability and data reuse. The process of developing these standards was divided into seven stages, each of which involved specific phases and activities that were carried out in sequence: identification of the context, review of existing standards, definition of relevant metadata, development of the standards proposal, validation and feedback, implementation and dissemination, and ongoing evaluation (109).

This proposal aims to enable users and researchers to download and publish records linked to citizen science, such as international standards like Darwin Core, EML, and OBIS.

However, one of the main challenges in developing standards for metadata documentation in citizen science projects is the diversity of approaches and types of data that these projects may involve (52) (110) (111). For example, while some projects may focus on wildlife observation in a specific area, others may involve environmental data collection or monitoring of astronomical phenomena (112). Therefore, it is important that the standards are flexible enough to accommodate a wide range of research contexts and data types (113) (114) (115).

In addition, it is critical to involve the citizen science community in the process of developing these standards (116) (117) (118). Participants in citizen science projects can provide valuable insight into their needs and preferences for metadata documentation, which can help ensure that the standards developed are practical, relevant, and accepted by the community (119) (120) (121).

In summary, the development of standards for metadata documentation in citizen science projects is an important step to improve the quality and utility of the data collected, as well as to promote transparency, collaboration, ethical considerations and scientific advancement in this emerging field of research (122) (123) (124) (125) (126). However, for these standards to be effective, a number of technical, practical, ethical, and legal challenges need to be addressed, and it is essential to involve the citizen science community in this process (127) (128).

5. CONCLUSION
In the course of this article, the current reality of metadata documentation in Citizen Science projects has been explored, outlining the palpable challenges that the community faces in this crucial area. The lack of standardization and persistent heterogeneity in metadata documentation have emerged as substantial obstacles,
imposing significant barriers to interoperability and efficient data reuse, fundamental links to scientific progress.

The initial standards proposal presented here was conceived in response to these challenges, and is a flexible but robust scaffolding. By incorporating fundamental elements such as project description, geographic location and participant information, we seek to lay the groundwork for documentation that is not only consistent, but also rich in meaning. This proposal is not intended to be a panacea, but rather a strategic and thoughtful starting point. However, it is humbly acknowledged that this proposal is merely the beginning of a broader and more complex journey. The active participation and contributions of the vibrant Citizen Science community are imperative to refine and enrich these standards, ensuring their utility in practice. The inherent diversity of the projects and the direct involvement of engaged citizens demand a collaborative and adaptive approach that evolves as rapidly as the changing needs of the scientific community.

An open and enthusiastic invitation is extended to researchers, developers, Citizen Science project participants, and anyone interested in joining this collaborative initiative. Continued collaboration will not only ensure that these standards adjust and refine with the changing demands of the community, but will also provide a solid foundation for effective data management in the dynamic context of Citizen Science. Together, we can chart the path to stronger, more consistent and effective Citizen Science.

References


12. Ramírez Montes LA. El desarrollo de la ciudadanía en los adolescentes: la lucha entre la influencia de la familia, el grupo de pares y la escuela pública. Región Científica. 2024;2024251. https://doi.org/10.58763/rc2024251


19. Riganti MaV. The brothel system as a social network: an approach to the sexual exploitation of women from a perspective of feminism and social network analysis. AWARI. 2023;4. https://doi.org/10.47909/awari.46


32. Gómez Cano CA, Sánchez Castillo V. Knowledge Structure in Rehabilitation within and beyond the Medical Field: Bibliometric Perspectives of the Categories “Physical Therapy, Sports Therapy and Rehabilitation” and “Rehabilitation”. Interdisciplinary Rehabilitation / Rehabilitacion Interdisciplinaria. 2022;2:22. https://doi.org/10.56294/rti202222


34. Mvr Almeida R, Fontes-Pereira AJ. Availability of retracted Covid-19 papers on Internet research-sharing platforms. Seminars in Medical Writing and


41. Prevall Vinent IM. Corporal Expression as a broad spectrum psycho-pedagogical resource. Seminars in Medical Writing and Education. 2024;3:48. https://doi.org/10.56294/mw202448


45. Noguera Cardona JS, Sánchez Castillo V. Misleading advertising of the food industry as a strategy to incentivize consumption even in the face of the enactment of the junk food law. Southern perspective / Perspectiva austral. 2024;2:26. https://doi.org/10.56294/pa202426


52. Cherradi M. Data lake management using topic modeling techniques. Seminars in Medical Writing and Education. 2024;3:282. https://doi.org/10.56294/dm2024282


71. Villa JAT. The official household public service companies and their obligatory to publish their actions in the SECOP. AG Managment. 2023;1:19–19. https://doi.org/10.62486/agma202319
72. Pérez Gamboa AJ. Contribution of graph theory to the understanding of social dynamics. AWARI. 2023;4. https://doi.org/10.47909/awari.51


74. Lepez CO, Simeoni IA. Pedagogical experience with Public Health campaigns from the design of socio-educational projects with insertion in the local territory. Community and Interculturality in Dialogue. 2023;3:74. https://doi.org/10.56294/cid202374


76. Álvarez García D, Cruz Cerón DY, Sánchez Castillo V. Analysis of farmers’ imaginary around the transition and adoption of the new livestock reconversion model in the municipality of Cartagena del Chaírá. Southern perspective / Perspectiva austral. 2024;2:27. https://doi.org/10.56294/pa202427


88. Alarcon JA. Risk management model for information security. DecisionTech Review. 2023;3:1–6. https://doi.org/10.47909/dtr.05


93. Rivera TC, Santa Cruz DIL. The needs to foment values in professional nursing technical training. AG Salud. 2024;2:47–47. https://doi.org/10.62486/agsalud202447


112. Ángel-Gaviria IS. The environmental effects of traditional livestock. AG Multidisciplinar. 2023;1:18–18. https://doi.org/10.62486/agusalud202315


