Modeling Observational Crowdsourcing

Arturo Castellanos[1], Roman Lukyanenko[2], and Veda C. Storey[3]

[1] Baruch College, CUNY, New York, NY United States
[2] HEC Montreal, Montreal, QC, Canada
[3] Georgia State University, Altanta, GA United States
arturo.castellanos@baruch.cuny.edu, roman.lukyanenko@hec.ca, vstorey@gsu.edu

Abstract. Crowdsourcing is an efficient way to engage the general public in making contributions to the production of goods and services. Studies have shown that observational crowdsourcing, as a continuous activity, has many potential benefits to society. However, a major challenge is how to model a crowdsourced activity. In this research, we provide guidelines for modeling observational crowdsourcing, focusing on user interfaces, data collection, data sharing, and interoperability. These guidelines are represented and illustrated by the application of a systemist ontology and the implications of doing so discussed.

Keywords: Observational Crowdsourcing · Crowdsourcing · Upper Ontology · General Systemist Ontology

1 Introduction

The last decade has seen the rise of *crowdsourcing*, whereby an organization, or even an individual sponsor, enlists members of the general public (the *crowd, contributors*) to produce data, goods or services [1], [2]. One of the most popular types of crowdsourcing is *observational crowdsourcing;* that is, "continuous, on-going process that involves observing or sensing the broader environment" [3, p. 3]. Notable examples include: eBird.org, a bird reporting project and one of the first online crowdsourcing initiatives founded in 2002; iSpotNature.org, a global platform to collect sightings of wildlife; and Rocksolid.com, a citizen engagement platform to support urban environment improvements. Observational crowdsourcing is widely used to support general management of global problems, such as pandemics (including COVID-19 [4]), climate change, overexploitation, invasive species, land use change, and pollution [5], [6].

Despite its potential, a major challenge is designing platforms is to engage crowds in this type of crowdsourcing. There is a much recognized need "to support observational crowdsourcing with innovative data management solutions" [3, p. 10] that would ensure decisions informed by available data, even when the insights are gained from the general public, and to do so in an innovative and cost-effective way.

The objective of this research is to propose a conceptual modeling approach to observational crowdsourcing to address the challenges associated with modeling crowdsourcing activities and to derive guidelines for modeling such activities. To do so, we apply principles from ontology to take a system view of the problem. The resulting guidelines focus on user interfaces, data collection, data sharing, and interoperability. The guidelines are presented using an upper-level ontology [7], the General Systemist Ontology (GSO), which describes the most general, domain-independent categories of reality. We conclude by discussing the implications of our work and suggesting future research opportunities.

2 Background: Observational Crowdsourcing and Ontology

Observational crowdsourcing continues to suffer from many challenges related to the design, development, and effective use of crowdsourcing platforms. We focus on three issues widely accepted as central challenges in observational crowdsourcing [8]–[10]:

- *Crowd Information Quality Challenge*— or **CrowdIQ** [11]. A major challenge in observational crowdsourcing is ensuring that the data produced by online crowds are of high-quality for subsequent analysis and action. From a modeling perspective, it means having conceptual structures that facilitate accurate and complete data reporting, storage and retrieval.
- *Crowd Usability Challenge*. To ensure crowdsourcing engages wide and diverse audiences, a major challenge is designing easy-to-use interfaces [12], which we call "**Crowd Usability**". From a modeling perspective, this translates to developing representations that are both intuitive and accessible to many people.
- *Crowd Interoperability Challenge*. Many crowdsourcing projects, especially of a scientific nature (online citizen science), seek to be maximally transparent and open to allow for active data sharing, which we call "**Crowd Interoperability**." From a modeling perspective, it requires creating structures that promote interoperability.

Research increasingly seeks solutions to these challenges. Notable approaches to *Crowd Information Quality*, for example, include restricting user participation to experts within crowds [13]; developing standardized protocols that can be distributed to volunteers, along with tutorials and instructions [8]; and leveraging redundancy, for example, by asking multiple participants to observe on the same phenomena and aggregating the results [14]. Limitations of these approaches include expert crowds suffering from "tunnel vision," strictly adhering to the task at hand, and ignoring valuable, unusual details [15]. Furthermore, restricting participation to experts misses an opportunity to engage with broader audiences.

Research is also exploring ways to improve *Crowd Usability*. One approach is to design interfaces and collection protocols to be as simple as possible; that is, "to design for dabblers" [12] or use very few and simple data collection options, referred to as "basic classes," such as bird, tree, or fish in the collection of wildlife observations [16]. Another modeling solution is to use novel instance-based modeling, which collects information in terms of unique instances and their attributes [17]. Potential limitations of these approaches include limited expressivity of data based on simple domain models and difficulty in using sparse and heterogeneous instance-based data.

To address *Crowd Interoperability* issues, there are efforts to recommend adoption of standardized protocols for data collection [18], but these remain contested [19], and often have different interfaces and approaches to collecting the same types of phenomena [20].

Additional progress on the three observational crowdsourcing challenges could be made by adopting ontological foundations that can represent observation crowdsourcing. Various ontologies have been adopted or developed within the IT research community, including DOLCE [21], Unified Foundational Ontology (UFO) [22], social ontology of Searle [23], General Formal Ontology [24], and the Bunge-Wand-Weber (BWW) [25], [26]. Prior research has already used select notions from an ontology to support crowdsourcing (e.g., [17], [19] which borrowed the notion of things and attributes from Bunge Wand Weber (BWW) [27]). However, no systematic attempt has been made to ground observational crowdsourcing into a specific ontology.

3 General Systemist Ontology - GSO

General Systemist Ontology (GSO) [28] is a new ontology based on the more recent ideas of Mario Bunge [29]. GSO appears to be especially well-suited for modeling observational crowdsourcing, as it explicitly deals with observations and systems [28]. GSO claims the world is made of *systems:* "everything is a system or a component of a system" [30, p. 23]. (See Table 1 for some of GSO's constructs). When systems interact, they transfer energy from one another. This leads to change in states, as they acquire or lose their properties. This produces *events* (a single change from one state to another). Multiple events form *processes*: defined as "a sequence, ordered in time, of events and such that every member of the sequence takes part in the determination of the succeeding member" [31, p. 172].

Table 1. Selected constructs of the General Systemist Ontology, from [28]

Construct	Definition	Source
Fact	whatever is the case, i.e., anything that is known or assumed - with some ground - to belong to reality	[28], p. 171
Object	whatever is or may become a subject of thought or action	[28], p. 174
Observation (direct observation)	purposeful and enlightened perception: purposeful or deliberate because it is made with a given definite aim; enlightened because it is somehow guided by a body of knowledge.	[28], p. 181
Observation (indirect observation)	hypothetical inference employing both observational data and hypotheses.	[28], p. 181
Phenomenon	is an event or a process such as it appears to some human subject: it is a perceptible fact.	[28], p. 173
System	complex object every part or component of which is connected with other parts of the same object in such a manner that the whole possesses some features that its components lack – that is, emergent properties.	[27], p. 205

Events, processes, phenomena, and concrete systems are instances of the mental concept of *fact.* Facts are kinds of *objects:* "whatever is or may become a subject of thought or action" [31, p. 174]; that is, "known or assumed - with some ground - to belong to reality" [31, p. 171]. Thus, through the notions of facts, GSO connects the fundamental ideas about the composition of reality to the mental world of humans. This makes *observation* (direct or indirect) a central construct at the nexus of ontology and epistemology. To further increase objectivity, especially when observations are those of humans, Bunge suggests that "observation results of *the same kind* should be reproducible by qualified observers; otherwise it should be kept in suspense. Exact duplication is desirable but not always attainable." [31, p. 186]. GSO introduces phenomenological considerations as well as the path from phenomena to human theories and mental models about the world. This is a notable departure from BWW, which focuses on the physical composition of reality (with the exception of the notions of classes and attributes, which are also part of GSO).

4 Guidelines for Supporting Observational Crowdsourcing

GSO is based on an ontological primitive, which is a "system." This is a fundamental construct which permeates the entire set of beliefs captured by this ontology. The focus on systems is primarily expected to improve usability and interoperability. Having identified and modeled the systems of interest to a given project, designers can then request that online crowds observe and report on these systems. The users can then enjoy the flexibility of either treating the objects of their observation as a complex phenomenon (i.e., a system), or as an individualized object. Both should be accommodated by the interfaces built based on GSO. This contrast, for example, with modeling approaches that focus on individual objects. As systems accommodate both perspectives, data generated on GSO-based platforms become compatible among them.

Some systems are still "individualized" entities (simply meaning we can abstract away their systemic properties). An observer can point to a system and describe it using its specific, attributes (which purport to represent the underlying properties). This ontological primitive can be used in many crowdsourcing initiatives which focus on mapping, tracking and representing individualized phenomena. For example, birds, fish, animals, consumer products, are all accommodated through the notion of a system from this point of view. This has been the focus of ontological studies [19], [32] that deal with crowdsourcing projects that focus on the identification of plants and animals.

At the same time, the notion of *system* in GSO goes beyond "individual" objects. Indeed, a notable limitation of thinking about the world from the point of view of individual things, is the difficulty in applying this notion to crowdsourcing projects that deal with phenomena without a clear identity or clear boundaries [33].

Thus, the notion of systems accommodates projects which are interested in clouds, algae, waves, sounds, wind, among others [34]. It is more "natural" to think of forces or fields as systems, rather than things [31]. Indeed, the systemic approach provides a broader foundation for crowdsourcing, seemingly accommodating most projects. Based on our observations, we introduce four guidelines that support observational crowdsourcing:

- **Guideline 1 – Design for *systems*:** *Consider the basic element of an application to be modeled as a system.*

GSO postulates that any system can be understood in terms of *composition*, *environment*, *structure* and *mechanism (CESM)*. We refer to this as the *CESM model* of CSO. The CESM model of GSO suggests specific ontological constructs for describing systems. These can be directly used in conceptual modeling. A conceptual model of a domain can contain elements consistent with CESM and propagate these into data collection and interface design choices. For example, a project collecting observations on lichens, can ask participants about the structure of the lichen observed, the host to which the lichen is attached (its environment), the individual strands that make up a collection of lichen, as well as the properties of these individual strands. With such an approach, projects can thus collect more complete data on systems of interest, by adopting CESM, thus increasing the value of such data for insights and actions.

Projects focused on interoperability can tag the elements of data collection as belonging to one or more of these basic elements. This would support the integration of data across different projects. Effectively, the CESM model can then become a structuring device for either creating data collection interfaces or sharing the data across different platforms.

- **Guideline 2 – Model based on CESM:** *Consider CESM elements as a device for obtaining information on the reported system.*

A large variety of crowdsourcing projects request contributors to take a "snapshot" (picture) observation of phenomena. For example, a contributor could be asked to describe an observed whale, with the user interface simply requesting the date and time of the observation. While this may be suitable and pragmatic for many projects, GSO suggests that the crowdsourcing community consider change in systems more deeply. As GSO postulates, all concrete things change, with change as a fundamental property. For GSO, change is understood in terms of the change of properties (attributes) of systems. These may be the properties of the system itself, its environment, or its subsystems. Note that GSO explicitly understands "observation" as either an observation of state (i.e., static representation of system) or observation of an event or a process (i.e., the change in the system). The latter two have been underutilized in crowdsourcing projects. Furthermore, very few crowdsourcing projects adopt a process view. The vast majority are "spot" observations [3], [34]. This denies a valuable opportunity to better understand how phenomena evolve, limiting the amount of data about the system and the kinds of inferences and insights that could be drawn. It is unlikely that there will be improvements to usability from following this aspect of GSO. However, there can be important implications for information quality. As more information is collected, and more opportunities emerge that can improve accuracy through better understanding of the phenomena, interoperability can improve by providing a common basis for modeling and representing change.

- **Guideline 3 – Design sensitive to change:** *Consider a system as something not static but under constant change.*

GSO provides an explicit conceptualization of the **"observation" construct**. Specifically, an observation of some observable fact about an underlying system. For GSO, facts can be mental conceptions of events, processes, phenomena or concrete systems (i.e., states). An observation describing the fact also includes an observer, the circumstances of observation and observation tools, making a given observation a *4-tuple* (e.g., "*w* observes *x* under *y* with the help of *z*"). This formalization of the observation objects is directly applicable to observational crowdsourcing and stands to benefit all three crowd challenges.

Modeling Guidelines and the Crowd Challenges Addressed

Guideline	Description	Crowd Challenges Addressed
G1: **Design for systems**	It is natural to think of forces or fields as systems, rather than things. Parts or components of systems are systems themselves.	Crowd Usability Crowd Interoperability
G2: **Model based on CESM**	In GSO, any system can be understood in terms of its *composition*, *environment*, *structure* and *mechanism (CESM), suggesting ontological constructs for describing systems.*	Crowd IQ Crowd Usability Crowd Interoperability
G3: **Design sensitive to change**	GSO provides the vocabulary and a way to conceptualize change [of properties of the system itself, its environment, or its subsystems].	Crowd IQ Crowd Interoperability
G4: **Model the observation construct**	In GSO, an observation describing a fact includes the observer, the circumstances of observation, and the observation tools	Crowd IQ Crowd Usability Crowd Interoperability

First, GSO reminds crowdsourcing projects that an observation made by a contributor is not a direct projection of the underlying system (because most facts are not directly observable). Rather, this is what an observer, from their own point of view, and with the help of their own observational tools, has detected and wishes to convey. For example, an observer watching from afar using a spotting scope a group of Kittiwakes (birds) sitting on an iceberg, close to a glacier. This perspective emphasizes the importance of understanding the observational tools; that is, knowledge and physical equipment used in making observations. Data resulting from crowdsourcing projects should be analyzed and interpreted by being cognizant of the knowledge and the equipment available (or assumed to be available) to the contributors. It is well-recognized that, to improve information quality, understanding the context of data capture is paramount [35], [36]. This also means that crowdsourcing projects may benefit from collecting additional information on the knowledge and equipment used to contribute observations, which is something that rarely occurs in projects. Rather, crowdsourcing projects are often narrowly focused on collecting data on the phenomenon of interest.

Second, the formalized notions of GSO can be used for standardization of interfaces and data collection, increasing interoperability and data exchange among projects. Third, GSO insists on the need for greater transparency and reliability of observations. Bunge asserts that multiple observations of the same system, including by multiple observers, as well as the same observer over time, should all be conducted in a public and transparent manner. This is needed for ensuring that the observational data faithfully depicts the underlying observed system. This idea is not new to observational crowdsourcing; for example, redundancy can be exploited within crowds by asking multiple people to report on the same thing, a frequently employed strategy. GSO provides theoretical justification and a conceptual framing for pursuing this strategy.

- **Guideline 4—Formalization and standardization of the observation object:** *Collect additional information about the observer, the circumstances of an observation, and observation tools*.

Table 2 summarizes the Modeling Guidelines based on GSO and the primary crowd challenges addressed by the guidelines. Overall, these guidelines provide an ontological foundation to observational crowdsourcing and have the potential to address the three focal challenges of this domain.

5 Conclusion

Observational crowdsourcing is now a prolific practice, engaging millions of people and thousands of organizations globally. Despite the mounting evidence of successes, it has numerous challenges due, at least in part, to the lack of a common theoretical foundation. This research leverages a new high-level ontology, General Systemist Ontology [28], as a basis for modeling observational crowdsourcing. Based on our analysis, the General Systemic Ontology appears to have the potential in improving user interfaces, data collection processes, data sharing, and interoperability in observational crowdsourcing. Nevertheless, future research is necessary to fully synthesize and empirically evaluate the benefits and limitations of GSO for observational crowdsourcing.

References

[1] D. C. Brabham, *Crowdsourcing*. Cambridge, MA: MIT Press, 2013.

[2] J. Prpić, P. P. Shukla, J. H. Kietzmann, and I. P. McCarthy, “How to work a crowd: Developing crowd capital through crowdsourcing,” *Business Horizons*, vol. 58, no. 1, pp. 77–85, 2015.

[3] R. Lukyanenko and J. Parsons, “Beyond Micro-Tasks: Research Opportunities in Observational Crowdsourcing,” *Journal of Database Management (JDM)*, vol. 29, no. 1, pp. 1–22, 2018.

[4] K. Sun, J. Chen, and C. Viboud, “Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study,” *The Lancet Digital Health*, 2020.

[5] E. J. Theobald *et al.*, “Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research,” *Biological Conservation*, vol. 181, pp. 236–244, 2015.

[6] R. Bonney *et al.*, “Citizen science: a developing tool for expanding science knowledge and scientific literacy,” *BioScience*, vol. 59, no. 11, pp. 977–984, 2009.

[7] G. Guizzardi, G. Wagner, J. P. A. Almeida, and R. S. Guizzardi, “Towards ontological foundations for conceptual modeling: the unified foundational ontology (UFO) story,” *Applied ontology*, vol. 10, no. 3–4, pp. 259–271, 2015.

[8] M. Kosmala, A. Wiggins, A. Swanson, and B. Simmons, "Assessing data quality in citizen science," *Frontiers in Ecology and the Environment*, vol. 14, no. 10, pp. 551–560, 2016.
[9] R. Lukyanenko, A. Wiggins, and H. K. Rosser, "Citizen Science: An Information Quality Research Frontier," *Information Systems Frontiers*, vol. 22, no. 1, pp. 961–983, 2019, doi: https://doi.org/10.1007/s10796-019-09915-z.
[10] N. Prestopnik and K. Crowston, "Citizen science system assemblages: understanding the technologies that support crowdsourced science," 2012, pp. 168–176.
[11] R. Lukyanenko, J. Parsons, and Y. Wiersma, "The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-generated Content," *Information Systems Research*, vol. 25, no. 4, pp. 669–689, 2014.
[12] A. Eveleigh, C. Jennett, A. Blandford, P. Brohan, and A. L. Cox, "Designing for dabblers and deterring drop-outs in citizen science," 2014, pp. 2985–2994.
[13] V. de Boer *et al.*, "Nichesourcing: Harnessing the Power of Crowds of Experts," vol. 7603, A. ten Teije, J. Völker, H, S. schuh, H. Stuckenschmidt, M. d'Acquin, A. Nikolov, N. Aussenac-Gilles, Hern, and N. ez, Eds. Springer Berlin / Heidelberg, 2012, pp. 16–20.
[14] R. Bonney *et al.*, "Next steps for citizen science," *Science*, vol. 343, no. 6178, pp. 1436–1437, 2014.
[15] S. Ogunseye and J. Parsons, "Can Expertise Impair the Quality of Crowdsourced Data?," 2016.
[16] A. Castellanos, M. Tremblay, R. Lukyanenko, and B. Samuel, "Basic Classes in Conceptual Modeling: Theory and Practical Guidelines," *Journal of the Association for Information Systems*, vol. 21, no. 4, pp. 1001–1044, 2020.
[17] R. Lukyanenko, J. Parsons, Y. F. Wiersma, G. Wachinger, B. Huber, and R. Meldt, "Representing Crowd Knowledge: Guidelines for Conceptual Modeling of User-generated Content," *Journal of the Association for Information Systems*, vol. 18, no. 4, pp. 297–339, 2017.
[18] A. Wiggins *et al.*, "Data management guide for public participation in scientific research," 2013.
[19] R. Lukyanenko and J. Parsons, "Conceptual modeling principles for crowdsourcing," in *International Workshop on Multimodal Crowdsensing*, Maui, Hawaii, USA, 2012, pp. 3–6.
[20] M. Maddah, R. Lukyanenko, D. VanderMeer, and B. Samuel, "Data Collection Interfaces in Online Communities: The Impact of Data Structuredness and Nature of Shared Content on Perceived Information Quality," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, Maui, Hawaii, USA, 2020, pp. 1–10.
[21] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, "Sweetening ontologies with DOLCE," in *Knowledge engineering and knowledge management: Ontologies and the semantic Web*, Springer, 2002, pp. 166–181.
[22] G. Guizzardi, G. Wagner, J. P. A. Almeida, and R. S. Guizzardi, "Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story," *Applied ontology*, vol. 10, no. 3–4, pp. 259–271, 2015.
[23] S. T. March and G. N. Allen, "Toward a social ontology for conceptual modeling," *Communications of the AIS*, vol. 34, 2014.

[24] H. Herre, "General Formal Ontology (GFO): A foundational ontology for conceptual modelling," in *Theory and Applications of Ontology: Computer Applications*, Springer, 2010, pp. 297–345.

[25] Y. Wand and R. Weber, "Toward a theory of the deep structure of information systems," in *International Conference on Information Systems*, Copenhagen, Denmark, 1990, pp. 61–71.

[26] Y. Wand and R. Weber, "An ontological analysis of some fundamental information systems concepts," *Proceedings of the Ninth International Conference on Information Systems*, vol. 1988, pp. 213–226, 1988.

[27] Y. Wand and R. Weber, "Mario Bunge's Ontology as a formal foundation for information systems concepts," P. Weingartner and G. Dorn, Eds. Rodopi, 1990, pp. 123–150.

[28] R. Lukyanenko, V. C. Storey, and A. Castellanos, "Introducing GSO: A General Systemist Ontology," in *ER Forum 2020*, Vienna, Austria, 2020, pp. 1–8.

[29] R. Lukyanenko, "A Journey to BSO: Evaluating Earlier and More Recent Ideas of Mario Bunge as a Foundation for Information and Software Development," in *Exploring Modeling Methods for Systems Analysis and Development (EMMSAD 2020)*, Grenoble, France, 2020, pp. 1–15.

[30] M. A. Bunge, "Systems everywhere," in *Cybernetics and applied systems*, London England: CRC Press, 2018, pp. 23–41.

[31] M. A. Bunge, *Philosophy of Science: Volume 2, From Explanation to Justification*. New York NY: Routledge, 2017.

[32] R. Lukyanenko, J. Parsons, Y. Wiersma, and M. Maddah, "Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-generated Content," *MISQ*, vol. 43, no. 2, pp. 634–647, 2019.

[33] B. Smith and D. M. Mark, "Do mountains exist? Towards an ontology of landforms," *Environment and Planning B: Planning and Design*, vol. 30, no. 3, pp. 411 – 427, 2003.

[34] A. Wiggins and K. Crowston, "From Conservation to Crowdsourcing: A Typology of Citizen Science," in *44th Hawaii International Conference on System Sciences*, Jan. 2011, pp. 1–10.

[35] Y. W. Lee, "Crafting Rules: Context-Reflective Data Quality Problem Solving," *Journal of Management Information Systems*, vol. 20, no. 3, pp. 93–119, 2003.

[36] G. Shankaranarayanan and R. Blake, "From Content to Context: The Evolution and Growth of Data Quality Research," *Journal of Data and Information Quality (JDIQ)*, vol. 8, no. 2, p. 9, 2017.