

EXPECTING THE UNEXPECTED: EFFECTS OF DATA COLLECTION DESIGN CHOICES ON THE QUALITY OF CROWDSOURCED USER-GENERATED CONTENT¹

Roman Lukyanenko

Department of Information Technologies, HEC Montréal,
Montréal, QC CANADA H3T 2A7 {roman.lukyanenko@hec.ca}

Jeffrey Parsons

Faculty of Business Administration, Memorial University of Newfoundland,
St. John's, NL CANADA A1B 3X5 {jeffreyp@mun.ca}

Yolanda F. Wiersma

Department of Biology, Memorial University of Newfoundland,
St. John's, NL CANADA A1B 3X5 {ywiersma@mun.ca}

Mahed Maddah

Sawyer Business School, Suffolk University,
Boston, MA 02108 U.S.A. {mmaddah@suffolk.edu}

As crowdsourced user-generated content becomes an important source of data for organizations, a pressing question is how to ensure that data contributed by ordinary people outside of traditional organizational boundaries is of suitable quality to be useful for both known and unanticipated purposes. This research examines the impact of different information quality management strategies, and corresponding data collection design choices, on key dimensions of information quality in crowdsourced user-generated content. We conceptualize a contributor-centric information quality management approach focusing on instance-based data collection. We contrast it with the traditional consumer-centric fitness-for-use conceptualization of information quality that emphasizes class-based data collection. We present laboratory and field experiments conducted in a citizen science domain that demonstrate trade-offs between the quality dimensions of accuracy, completeness (including discoveries), and precision between the two information management approaches and their corresponding data collection designs. Specifically, we show that instance-based data collection results in higher accuracy, dataset completeness, and number of discoveries, but this comes at the expense of lower precision. We further validate the practical value of the instance-based approach by conducting an applicability check with potential data consumers (scientists, in our context of citizen science). In a follow-up study, we show, using human experts and supervised machine learning techniques, that substantial precision gains on instance-based data can be achieved with post-processing. We conclude by discussing the benefits and limitations of different information quality and data collection design choices for information quality in crowdsourced user-generated content.

Keywords: Crowdsourcing, user-generated content, citizen science, information systems design, information quality, information completeness, information accuracy, information precision, discovery, supervised machine learning

¹Andrew Burton-Jones was the accepting senior editor for this paper. Gove Allen served as the associate editor.

The appendices for this paper are located in the "Online Supplements" section of *MIS Quarterly*'s website (<https://misq.org>).

Introduction

Organizations are increasingly interested in *user-generated content* (UGC), information produced by members of the general public (the crowd), who are often unpaid and not affiliated with the organization (Doan et al. 2011; Love and Hirschheim 2017; Prpić et al. 2015; Surowiecki 2005). To harness UGC, organizations mine existing data, such as forums, blogs, social media, comments, and product reviews (e.g., Abbasi et al. 2018; Brynjolfsson et al. 2016; Gao et al. 2015; Kallinikos and Tempini 2014), or create new data collection processes through intermediaries such as Amazon's Mechanical Turk and CrowdFlower (Chittilappilly et al. 2016; Daniel et al. 2018; Deng et al. 2016; Garcia-Molina et al. 2016; Li et al. 2016).

Of great value is *crowdsourced UGC*, wherein organizations develop custom information systems (IS) to collect specific kinds of data from contributors external to the organization (see Kosmala et al. 2016; Lukyanenko et al. 2017; Prestopnik and Tang 2015; Prpić et al. 2015; Wiggins and Crowston 2011). As these projects tap into personal interests (e.g., civic responsibility, science, and health), successful projects attract reliable cohorts of contributors at little additional cost (Clery 2011; Nov et al. 2014). For instance, CitySourced (www.citysourced.com) is a U.S.-wide project that encourages people to report local civic issues (e.g., crime, public safety, environmental issues) and makes this data available to participating municipalities.

Crowdsourced UGC (hereafter simply UGC) often captures people's real-world experiences with public services, infrastructure, consumer products, or the natural environment. The potential unconstrained, experiential nature of UGC contrasts with traditional organizational data, which tends to be focused and transaction-based. Organizations can use UGC to better understand their customers, competitors, products and services, or the social and political environment (Brabham 2013; Goodman and Paolacci 2017; Khatib et al. 2011; Prpić et al. 2015). However, despite the opportunity to use UGC in innovative ways, many IS designed to collect it follow traditional design approaches common to organizational systems. Such designs frequently rely on predetermining the categories of data to be collected and present these categories as data entry choices that guide and constrain contributors in reporting data. This approach can interfere with the opportunity to capture contributors' unanticipated, unique, experiential interactions with the phenomena they observe and report via a crowdsourcing IS.

Researchers have argued that traditional processes of managing information quality (IQ), such as training and motivation, can break down in the context of UGC, making it

difficult to ensure data provided by online crowds is of sufficient quality to be used in organizational decision making (Lukyanenko et al. 2014). This problem of "crowd IQ" has attracted considerable research attention (for reviews, see Daniel et al. 2018; Kosmala et al. 2016; Lewandowski and Specht 2015; Li et al. 2016; Tilly et al. 2016), with much of the focus being on a single dimension of IQ: information accuracy (Daniel et al. 2018; Kilian 2008; Lukyanenko et al. 2014). Virtually nothing is known about how to capture data about unanticipated phenomena and foster discoveries from UGC. In particular, there is a strong need to understand how to use crowdsourced data to promote discovery without sacrificing traditional measures of IQ. In addition to accuracy, organizations want to make sure the data provided is as complete as possible and precise enough for the task at hand. In this paper, we examine ways in which to promote unconstrained data collection, while preserving a high level of IQ, focusing on the traditional dimensions of accuracy, completeness, and precision.

We examine the impact of IS design features (in particular, the data collection interface) on the ability to capture unanticipated data, and the impact of such features on crowd IQ. While organizations have less control over crowds than over employees, they do control the design of data collection in custom-built systems. We seek to answer the following research question:

How can an organization capture unanticipated data about phenomena from crowds in a way that maintains the information quality needed for anticipated uses of data?

To address this question, we begin by considering IS design approaches for UGC and consider how interface design decisions related to the categories of data collected can affect aspects of IQ in UGC. We then theorize about the effect of design choices on key indicators of data quality and the potential for capturing data that enables discovery. Subsequently, we implement alternative design choices based on this theory, evaluate them in complementary studies, and validate the practical utility of our work in an applicability check with practitioners. We conclude by discussing implications for future research and practice.

Information Quality and IS Design in Crowdsourced UGC

Traditional information quality research recognizes three stakeholders: (1) data consumers, who use data in decision making; (2) data custodians, who curate data; and (3) data

contributors, who produce data (Lee 2003; Pipino et al. 2002). In UGC settings, custom-built IS used for data collection generally incorporate and reflect the needs of data consumers (Kosmala et al. 2016; Lewandowski and Specht 2015). We develop a broader conceptualization of “crowd IQ” that incorporates characteristics of UGC contributors and the crowdsourcing IS.

Researchers have long taken the perspective that organizational data are collected for use by data consumers (Madnick et al. 2009; Strong et al. 1997; Wang and Strong 1996). Consequently, the prevailing approach to managing IQ in UGC is “consumer-centric” (see Lukyanenko et al. 2016; Wiggins et al. 2013). This perspective is understandable, as organizations develop UGC IS to achieve specific goals. The more UGC satisfies data consumers’ needs, the better the immediate return on effort and investment (Devillers et al. 2010; Engel and Voshell 2002; Louv et al. 2012).

The consumer-centric view is reflected in the prevailing definition of IQ as the *fitness-for-use* of data by consumers for specific purposes (Chittilappilly et al. 2016; Lee et al. 2006; Shankaranarayanan and Blake 2017; Tilly et al. 2016; Wang and Strong 1996). Consumer-centric IQ improvement strategies include vetting and selecting contributors, training them, and providing instructions and feedback (Lee 2003; Lee et al. 2006; Redman 1996). Under this view, IS design focuses on ways to better satisfy data consumers’ known information needs, which typically means constraining data collection based on the known and stable information requirements of data consumers (Browne and Ramesh 2002).

Consumer-Centric IQ and Class-Based Data Collection in UGC

Consumer-centric IQ management is manifested during the design of data collection processes by *class-based* approaches to data collection (Krogh et al. 1996; Parreiras and Staab 2010; Parsons and Wand 2000; Shanks et al. 2008). Class-based data collection involves *a priori* specification of the kinds of phenomena to be represented in an IS, and the relevant attributes and relationships among them (Chen 1976; Clarke et al. 2016; Olivé 2007; Wiggins et al. 2013). Using class-based interface design features, data consumers are able to specify exactly the kind of UGC they wish to obtain from crowds (e.g., via prespecified classes, attributes, or relationships). The data collected based on consumer-centric IQ are commonly at a high level of precision (e.g., biological species, product categories, landscape features) needed for expert analysis. Examples include eBird (www.ebird.org) and Citysourced (www.citysourced.com).

Traditional IQ research considers class-based data collection to be necessary for achieving high IQ (Wang et al. 1995). Concerns about the potential for crowds to generate hard-to-use “noisy” datasets (Brynjolfsson et al. 2016; Ipeirotis et al. 2010; Sheng et al. 2008) make it reasonable to constrain data collection in UGC settings (Kosmala et al. 2016; Wiggins et al. 2013). Finally, there may be specific questions that are highly focused and for which a class-based system is necessary. Once data consumers’ needs are captured in the design via classes, contributors can be trained how to report information using these classes and their attributes or relationships among them (Daniel et al. 2018; Kosmala et al. 2016; Wiggins et al. 2011).

Contributor-Centric IQ and Instance-Based Data Collection in UGC

Despite advantages, consumer-centric IQ and class-based data collection sometimes have limitations. In particular, prior research has shown that class-based data collection may have a negative impact on IQ in UGC settings, as nonexpert crowds may be incapable of providing *accurate* data at the level of precision (e.g., biological species) typically needed by data consumers (e.g., scientists) (Lukyanenko et al. 2014). There is a growing call to design data collection in UGC to be “as flexible and adaptable to the producers [of information] as possible, while expecting a variety of content” (Tilly et al. 2016, p. 8; see also McKinley et al. 2016; Parsons et al. 2011).

In contrast to organizational settings, UGC offers limited opportunities to exercise tight controls over data contributors, making traditional IQ management approaches, such as training, less effective. In many circumstances, the primary interaction data contributors have with the organization is through the design features and the content of the IS, the system that contributors use voluntarily and may abandon at any time (Deng et al. 2016; Gray et al. 2016; Wells et al. 2011). Moreover, data contributors might not be motivated or savvy enough to find a suitable workaround when a problem is encountered (e.g., by sending an email to a representative of the project or posting something in a discussion forum). Consequently, if the classes used in the design do not match the views of data contributors, they may abandon efforts at data entry or contribute data based on guesses.

Moreover, a significant part of the appeal of UGC is as a way of finding something unexpected and new. It has been long known that front-line employees, being in direct contact with day-to-day situations, are well-equipped at spotting unusual activity, manufacturing defects, or process failures (Tax and

Brown 1998; Trevino and Victor 1992; Tucker and Edmondson 2003). In our context, a notable ability of crowds is reporting individual experiences with objects of interest to data consumers. Many phenomena (e.g., consumer product malfunctions, urban crime, or drug side effects) may be observed only by contributors (Kallinikos and Tempini 2014; Lukyanenko, Parsons, and Wiersma 2016). Furthermore, “because citizens generally lack formal ... training, they view problems and issues in light of their own knowledge and interests, creating fertile ground for discoveries” (Lukyanenko, Parsons, and Wiersma 2016, p. 447). We thus observe the following paradox in UGC: nonexperts may be unable to provide data using categories determined in advance to be important for data consumers, but they have a remarkable ability to observe unusual and unexpected phenomena. Thus, to facilitate discovery, systems need to be open and flexible, rather than focused only on predetermined data consumers’ needs.

To support UGC projects interested in discovery, we introduce a complementary contributor-centric approach to crowd IQ management. While the consumer-centric approach seeks to maximize fitness-for-use, contributor-centric IQ management embraces heterogeneity and low domain expertise in the crowd. By capturing the perspectives of contributors, a system that collects UGC can be designed to be open to unexpected data, allow flexibility to collect a greater variety of data, and enable capturing data for unanticipated uses and discovery. We define **contributor-centric IQ management** as a *strategy that seeks to increase the heterogeneity and diversity of UGC by removing contribution barriers, including data input constraints and crowd selection mechanisms, without sacrificing traditional dimensions of IQ*. Contributor-centric IQ management recognizes that, when freed from predefined data entry constraints, crowds can still generate accurate data (at lower levels of precision). This approach is not intended to replace traditional IQ management predicated on fitness-for-use, but offers benefits when contributors are not equipped to provide data at the level of precision desired by project sponsors and when unanticipated data may be important for current or future uses.

To implement contributor-centric IQ, data collection needs to be flexible, supporting variable views and perspectives in heterogeneous crowds. Instance-based data collection has been suggested as a way to store data in UGC and empirical evidence shows that when it is used, nonexperts are able to provide accurate descriptions of instances using generic classes (e.g., bird, trees, fish) (Lukyanenko et al. 2014). Following these findings, in this paper we suggest implementing contributor-centric IQ through data collection driven by *instances*, based on the ontological view that instances exist independent of classes (Bunge 1977).

There has been increased interest in representations using instances and instance-based IS (e.g., Lukyanenko et al. 2019; Parsons and Wand 2000; Saghaei 2016; Samuel et al. 2018; Sekhavat and Parsons 2012). Building on that work, we operationally define an **instance-based IS** as a *system that captures information about instances (objects, individuals) via a data collection interface that allows contributors to describe observed instances in terms of any classes or attributes of interest at any level of precision*. An instance-based IS removes the requirement for contributors to understand and comply with predefined data collection options (which may not match the views of data contributors). Thus, rather than focusing on specific uses of information, quality improvements can be made by giving crowds flexibility to contribute data as they see it. This presupposes the ability to transform heterogeneous unclassified data to match classes of known interest to data consumers, an outcome that might be achievable by post-processing. Importantly, it opens data collection to capturing unexpected, idiosyncratic, and personalized perceptions of crowd members.

Instance-based and class-based data collection in UGC are complementary and each may be effective in different scenarios. In a spectrum of possibilities, we define two extremes: (1) *open*, with many unknowns, versus (2) *closed*, with few unknowns (see Table 1). In practice, projects may possess both closed and open features. Contributor-centric IQ and instance-based data collection are best pursued in open settings with weak data production controls, anonymous or nonexpert crowds, and an interest in capturing unanticipated phenomena or data. Conversely, class-based data collection is most appropriate when training is feasible, domains are relatively narrow in scope, and uses are established in advance and stable. Each approach comes with trade-offs. For example, instance-based data collection may open participation to wider audiences, but creates challenges for integrating data into traditional decision making and analysis (Table 1). Thus, the choice for a particular application depends on the relative importance of project dimensions.

The closed scenario is better understood as it builds upon a wealth of research on traditional corporate data production and use. In contrast, the potential and limitations of the open scenario are poorly understood. This lack of understanding may result in projects adopting ineffective IQ management strategies. We focus on the open setting with the aim of theoretically explicating and providing empirical support for the effectiveness of instance-based data collection as a means of enacting contributor-centric IQ.

Table 1. Data Collection Scenarios for Instance-Based Versus Class-Based Approaches

Project Dimension	Subdimension	Open with Many Unknowns	Closed with Few Unknowns
Project nature	Domain Scope	<i>Large, unbounded</i> (e.g., entire natural history of a region)	<i>Small, bounded</i> (e.g., tufted puffins in an area)
	Task	<i>Open-ended</i> (e.g., tell me anything about an object)	<i>Close-ended</i> (e.g., tag all pedestrians in a photo, transcribe text using finite symbols)
Data Contributors	Skills and abilities	<i>Open participation: nonexperts and experts in project domain</i> (e.g., ability to observe phenomena and describe it using own vocabulary)	<i>Closed Participation: experts in domain</i> (e.g., ability to identify instances of birds at species level of precision)
	Training	<i>Not required</i> (e.g., anyone can contribute data)	<i>Might sometimes be required</i> (e.g., users must pass tutorial to contribute data)
Data consumers	Uses	<i>Unknown, evolving, some known</i> (e.g., CitySourced.com collects civic reports; municipalities access data and use it in own ways)	<i>Known and well-understood</i> (e.g., detect occurrence of specific invasive species in a given area)
Suggested IQ Management and IS Design	IQ Management	<i>Contributor-centric</i>	<i>Consumer-centric (fitness-for-use)</i>
	Data collection	<i>Instance-based</i>	<i>Class-based</i>
	Post-processing	<i>Significant and advanced post-processing may be required</i> (e.g., machine learning may help infer species from contributed attributes of instances)	<i>Significant and advanced post-processing is generally not required</i> (e.g., classifications of galaxies may be aggregated by type)
Exemplar project		<i>iSpotNature</i> (www.ispotnature.org)—observations of wildlife worldwide	<i>eBird</i> (www.ebird.org)—classification of birds primarily into pre-specified species

Theoretical Propositions

We now consider psychological mechanisms that explain how class-based and instance-based data collection approaches affect IQ in open UGC settings. We first consider the traditional dimensions of accuracy, precision, and dataset completeness. We then propose a new IQ dimension: discovery (conceptualized as a facet of completeness).

Accuracy

Accuracy is central to assessment of IQ (Burton-Jones and Volkoff 2017; Redman 1996; Wand and Wang 1996). We define a statement $S(x)$ about a phenomenon x to be **accurate** if the *statement is true for x* (Lukyanenko et al. 2014; Redman 1996). For example, if a nonexpert observes a Mallard duck in the wild and records it as “Blue-winged teal,” this stored data will not be accurate, whereas labels such as “duck,” “bird,” “Mallard duck,” or “has webbed feet” will be accurate. For organizations hoping to make decisions based on crowd

data, accuracy is critical, and it is particularly challenging to determine the accuracy of observations, especially when they are being reported by nonexperts in the domain (Kosmala et al. 2016; Lewandowski and Specht 2015).

We expect accuracy in an instance-based IS to be higher than in a class-based one in open UGC projects for two reasons. First, there is a high likelihood of a mismatch between the classes familiar to a contributor (typically high-level or basic-level ones, which contributors tend to provide with high accuracy) and those defined in the IS (typically at a level of precision reflecting needs of expert data consumers). When required to conform to an IS-imposed class structure, a contributor may guess and classify an observation incorrectly. Research in cognitive psychology suggests that humans are more accurate when using classes with which they are familiar. The most familiar classes for nonexperts are *basic-level categories* (e.g., fly, snake, bird, bee, tree, fish), which typically lie in the middle of a taxonomical hierarchy (e.g., “bee” is a level higher than a specific species of bee, and lower than “insect”) (Lassaline et al. 1992; Rosch et al. 1976).

Note that accuracy is not necessarily higher when more general classes are used. Nonexperts may incorrectly classify at higher levels, while correctly classifying at intermediate ones (e.g., correctly classify an instance as a killer whale, but incorrectly classify it as a fish) (Bell 1981). Thus, when an IS requires information contributors to conform to a classification (driven by anticipated uses of data), accuracy will be lower when the classes are unfamiliar to contributors.

Second, when allowed to describe both familiar and unfamiliar objects, people are generally able to describe such objects using free-form attributes with high accuracy. This is because discriminating among attributes of objects is a prerequisite for classifying them. Classification is a fundamental function that allows humans to deal with the ever-changing diverse world (Berlin et al. 1966; Harnad 2005; Mervis and Rosch 1981). As discriminating objects by attributes is critical for successful interaction with the environment, humans have a well-developed and highly reliable ability to “describe any object” (Wallis and Bühlhoff 1999, p. 24). Thus, in the context of crowdsourcing, even when people may be reporting on an unknown object, we can expect the descriptions (e.g., attributes, textual comments) of such objects to be generally accurate. This leads to

Proposition 1: In open crowdsourced UGC projects, an instance-based IS will produce more accurate data than will a class-based IS.

Precision

As we consider UGC projects that satisfy specific data needs, the extent to which the resulting data are useful in addressing these needs is important. We thus consider another IQ dimension: *precision* (also known as level of detail or specificity) (Redman 1996; Wang and Strong 1996). Precision refers to the level in a knowledge hierarchy to which a concept belongs (e.g., species is more precise than genus) (Redman 1996, p. 250). In IQ research, precision is generally viewed as independent of accuracy (i.e., American Robin is more precise than bird, but inaccurate if the bird is a Blue Jay). The higher the precision, the more the resulting data is potentially useful to experts for particular uses, but only if it is also accurate (Boakes et al. 2010; Mayden 2002). For example, in projects that deal with nature, citizens are often asked to identify organisms by species, rather than generic classes such as bird or animal (Lewandowski and Specht 2015; McKinley et al. 2016; Sullivan et al. 2009). Indeed, projects often are interested in only some species and provide contributors with a list of desired species to select from (e.g., The Great Sunflower Project) (Callaghan and Gawlik 2015; Sullivan et al. 2009; Swanson et al. 2015).

Compared to an instance-based IS, we propose that a class-based IS will *positively* affect precision for two reasons: (1) a class-based IS may specifically require instances to be collected at a given level of precision, whereas an instance-based IS does not require this; (2) class-based IS are developed to satisfy specific data consumer needs, typically reflected in terms of specialized classes that nonexperts may struggle to provide, in contrast with an instance-based IS where such constraints are absent. Correspondingly, we propose

Proposition 2: In open crowdsourced UGC projects, a class-based IS will produce more precise data than will an instance-based IS.

Completeness

Another major IQ dimension is *completeness* (Batini et al. 2009; Redman 1996; Wang and Strong 1996), defined as “the degree to which a given data collection includes data describing the corresponding set of real-world objects” (Batini et al. 2009, p. 7). We adopt this definition and note that completeness can be assessed on a comparative basis, as different datasets can describe the “corresponding set of real-world objects” to varying degrees. We assume it is better to have some information about an object than no information at all. Considering this, the class-based and instance-based approaches can be compared in terms of two facets of completeness of UGC: (1) dataset completeness and (2) number of discoveries.

We define **dataset completeness** as the *extent to which an IS captures all phenomena of potential interest (for predetermined and emergent uses) to data consumers that data contributors are willing to provide (regardless of how detailed the data about instances may be)*. The phenomena of interest for a project can be specified by a (temporally or geographically bounded) superordinate category that guides potential contributors as to the project’s scope (e.g., birds).

We further define **number of discoveries** as the *number of instances captured of classes not anticipated during the design of an IS*. For example, if a contributor describes a relevant (to data consumers) unexpected kind of organism for which no class (or set of attributes) existed in the IS, we consider this a discovery. In our context, discovery need not mean something completely unknown to the knowledge community, although it does not preclude this possibility (as we demonstrate in the field experiment).

Compared to an instance-based IS, a class-based IS will negatively affect both dataset completeness and number of discoveries for several reasons: (1) in open UGC projects, it

may be impractical to determine in advance all relevant classes (regardless of whether participants are previously familiar with them), thereby deterring participants from reporting instances of these classes; (2) class-based interfaces lack direct technology affordances (perceived design features of IS matching human abilities (Leonardi 2011; Markus and Silver 2008; Norman 1983) to capture instances that do not match available classes, making it difficult to report such instances; (3) the predefined classes may act as an anchor (Gilovich et al. 2002; Tversky and Kahneman 1974), an initial starting condition that affects the information contributors subsequently provide by (intentionally or inadvertently) signaling to the contributor that only instances of the predefined classes are of interest; and (4) there is a possible mismatch between classes provided by the class-based IS and those familiar to data contributors, preventing contributors from reporting an instance in terms of the classes or attributes they are more familiar or comfortable with. Accordingly, we propose

Proposition 3: In open crowdsourced UGC projects, an instance-based IS will produce more complete data than will a class-based IS.

Figure 1 summarizes our research propositions (assuming no post-processing of the data).

Empirical Studies

To evaluate the relative merits of the consumer-centric and contributor-centric perspectives, we compare class-based and instance-based approaches in the context of an open UGC project. We conducted four empirical studies: (1) a field experiment to evaluate Propositions 2 and 3 (precision, dataset completeness, and discoveries); (2) a laboratory experiment (informed by the findings of the field experiment), where greater controls allowed us to evaluate all three propositions; (3) an applicability check in which we presented the two data collection approaches and our empirical evidence to potential data consumers and elicited their feedback on the applicability and usefulness of each approach for their work; and (4) a study with human experts and machine learning methods to investigate the potential usefulness of data generated by the instance-based approach.

Field Experiment

We conducted a field experiment in the context of citizen science in biology (Bonney et al. 2009; Silvertown 2009). This is an ideal setting for UGC IQ research, as there are established standards of quality (e.g., biological nomen-

clature), a well-defined cohort of data consumers (scientists), and relatively well-established information needs (e.g., collection of data at the species level of analysis) (Burgess et al. 2017; Levy and Germonprez 2017). As citizen science is a voluntary endeavor, an important challenge is how to induce data of acceptable quality while promoting discoveries. Finally, citizen science is becoming more prominent in the IS discipline, fueled by its societal importance (Goes 2014; Levy and Germonprez 2017; Lukyanenko et al. 2014). Despite a wealth of research, most studies on citizen science focus only on accuracy, with scant attention on how to foster discoveries (a key objective of science).

We argue that observations collected from citizen scientists in open settings can benefit from an instance-based approach. No study to date has compared the impact on IQ of this approach versus a traditional class-based IS. Consistent with Proposition 3, we predict

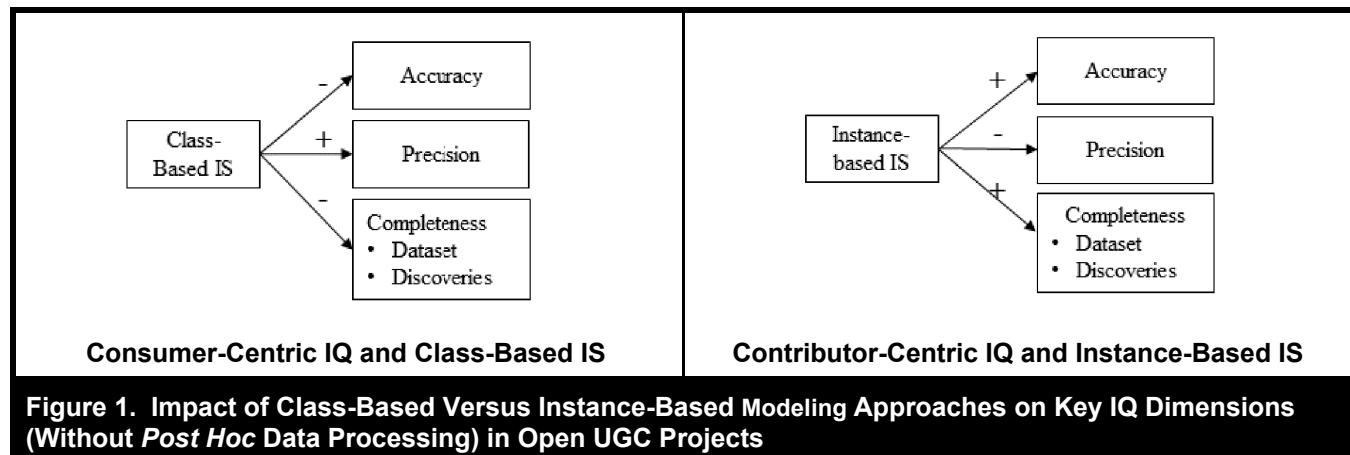
Hypothesis 1: Number of instances reported.

The number of observations reported using an instance-based IS (described using attributes or classes at any level) will be higher than the number of observations reported using a class-based IS (described as biological species).

At the same time, the focus on species in class-based IS guarantees a high level of precision (only species-level observations), whereas the instance-based IS is expected to deliver significantly fewer species-level classes, even if contributors are familiar with the organisms, as contributors are not directed toward this classification level. Although this might appear obvious (participants in the instance-based condition will be less likely to report at the species level when they have other options), it is not guaranteed. For example, participants might view the focus of the system to be on species and try to report accordingly (even if incorrectly). Furthermore, it is quite possible that people who voluntarily join the project have significant domain expertise and routinely conceptualize phenomena in terms of specialized classes. We hypothesize

Hypothesis 2: Information Precision. The proportion of species-level observations reported using a class-based IS will be higher than the proportion of species-level observations reported via an instance-based IS.

A field experiment also makes it possible to explore the potential for crowds to report information about unanticipated phenomena. Here, we compare class-based versus instance-based IS in terms of the extent to which they capture data about unanticipated phenomena. To ensure an equitable and



conservative comparison, we focus on *unanticipated* (i.e., not previously included in the project schema) *species-level classes*. This comparison is conservative, as species-level identification is the explicit focus of the class-based IS, but is not emphasized in the instance-based IS. Thus, one could argue that it should be more natural for the class-based IS to produce more unanticipated species. Following the arguments presented above, and consistent with Proposition 2, we hypothesize

Hypothesis 3: Number of instances of unanticipated species stored. The number of observations of unanticipated biological species reported using an instance-based IS will be higher than the number of observations of unanticipated biological species reported using a class-based IS (containing classes known to be useful to data consumers).

Independent Variables

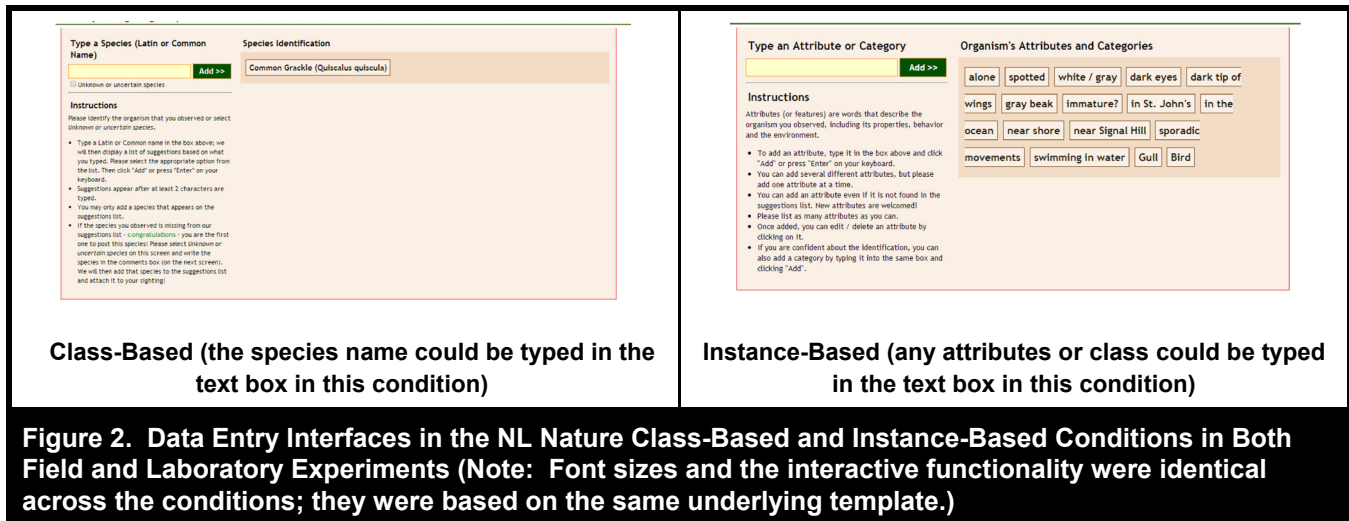
To evaluate the proposed hypotheses, we used data from a web-based citizen science project, NL Nature (www.nlnature.com). The project collects data on biodiversity in a region based on nature sightings (e.g., plants and animals). Importantly, the project was class-based for the four years preceding the experiment. We thus had an ecologically valid class-based project schema to compare with instance-based collection.

The decision to conduct the experiment was made one year prior to the start of data collection. The intervening time was spent in planning and development. Preceding the launch of the redesigned NL Nature, activity on the website was low. We redesigned NL Nature and launched the experiment in late spring, when people spend more time outdoors.

We developed two custom data collection interfaces: a class-based interface using species-level data entry, and an instance-based interface. The interfaces were designed to be visually similar and were dynamically generated from the same master template. Potential contributors (citizen scientists) were randomly assigned to one of the two data collection interfaces upon registration and remained in that condition for the duration of the experiment. The data entry form required authentication to ensure that contributors were not exposed to different conditions. All contributors received equal access to other areas of the project (e.g., internal messaging system, forum) and equal support from the project sponsors. This ensured equivalent *facilitating conditions* (Venkatesh et al. 2003) across the groups.

We first built a class-based interface following the common practice of asking contributors to select a species (i.e., class) from a predefined list.² As it is possible a contributor might not know or be confident in a species-level identification, we provided an explicit option (with clear instructions) to bypass the species-level classification by clicking on the “Unknown or uncertain species” checkbox below the data entry field (see Figure 2 left panel). We further instructed participants who checked the box to use the “comments” field to specify any class to which they believed the instance belonged. This was done to prevent what we believed might be a strong impediment to participation by nonexperts in the class-based condition (although it weakens our treatment compared to an inflexible class-based interface, thereby providing a conservative test of the effect of our treatment). However, consistent with the logic of consumer-centric IQ, we only counted species-level observations, in order to provide data at the level

²Examples of such projects include eBird.org, Citysourced.com and many projects on the Zooniverse.org platform (see Appendix C for more examples).



of specificity desired by target data consumers. In the instance-based condition, we instructed participants to provide attributes and, if possible, classes (Figure 2, right panel). This allowed contributors to report sightings even if they could not determine a class for the instance observed.

The initial list of species was developed by the third author, an ecology professor—an expert in local natural history—when the project was launched. During the four years preceding the field experiment, the list was updated periodically by the website members, who were encouraged to suggest new species (using the comments field available in the older version of NL Nature). Biologists also reviewed the list periodically and updated it as needed. By the time the experiment began, the species list was stable with very infrequent updates and contained 343 species-level classes. While we believe the list of classes was ecologically valid, it was also incomplete, as not every kind of organism which the crowd could potentially report was represented.

In both conditions, to see a list of options (classes or attributes) contributors were instructed to begin typing in the text box and click “Add” or press “Enter” when finished. As soon as more than two characters were entered, a suggestion box appeared with the available (stored) classes or attributes containing the string. In the class-based condition, participants were required to select an item from the list (or supply a new class in the comments). In the instance-based condition, a participant could select an item from the list or provide new attributes and classes via direct entry. Two sample records created in both interfaces are shown in Figure 3.

We used the traditional class-based condition as a template for the instance-based one, as it was more important to ensure

equivalence across conditions than to produce the most effective implementation of the instance-based interface.

Participation was voluntary and anonymous. To use NL Nature, participants first accepted a consent form outlining the nature of their interaction with the website. Those who didn’t were unable to contribute data. No incentives for participation were provided and no personally identifying information was collected. The existence of the experimental manipulation was not disclosed. We also monitored all contributor correspondence available through the website and social media that reference the project for evidence of contributors becoming aware of different interfaces; we found no such evidence.

Results

Our analysis is based on data provided by the website members who accepted the consent form after the manipulations outlined above took effect. Over a six-month period, 230 members accepted the consent form and began to participate. Contributors were randomly assigned into the two study conditions.³ Some participants registered, but never landed on the observation collection page and, hence, were not actually exposed to manipulation (determined by analyzing server logs). The final numbers of participants who visited one of the two observation collection interfaces at least once were 42 in the class-based condition and 39 in the instance-based con-

³Approximately 30% of the members who accepted the consent during the study period were randomly assigned to a third condition to test hypotheses outside the scope of this paper.



dition. The following analysis is based on the information provided by these 81 contributors.

While we did not require contributors to provide demographics, some volunteered this information. Fifteen participants indicated their age (mean 50.87, SD 15.54). Seventeen participants indicated how long they lived in the targeted geographic region (mean 18.85, SD 17.30). Fourteen participants provided the hours per week spent outdoors (mean 19.14, SD 15.54). In sum, those who provided demographics were mature and experienced.

To evaluate H1, we analyzed observations (i.e., one observation is one sighting of one or more organisms that could be classified or described using attributes, irrespective of how much data was provided) provided by 81 participants exposed to the two conditions. In the class-based condition, we removed non-species observations (i.e., cases where participants inserted a class in the comments box, but clicked “Unknown or uncertain species” in the data entry field) from the count of observations for these contributors, as they would not have been captured in a system only allowing species-level classification; we kept all the species classes (including seven entered using the comments box). Table 2 reports the number of contributions in each condition.

We tested the assumption of normality in the data using the Shapiro-Wilks test. In each condition, the distribution of observations per contributor significantly deviated from normal

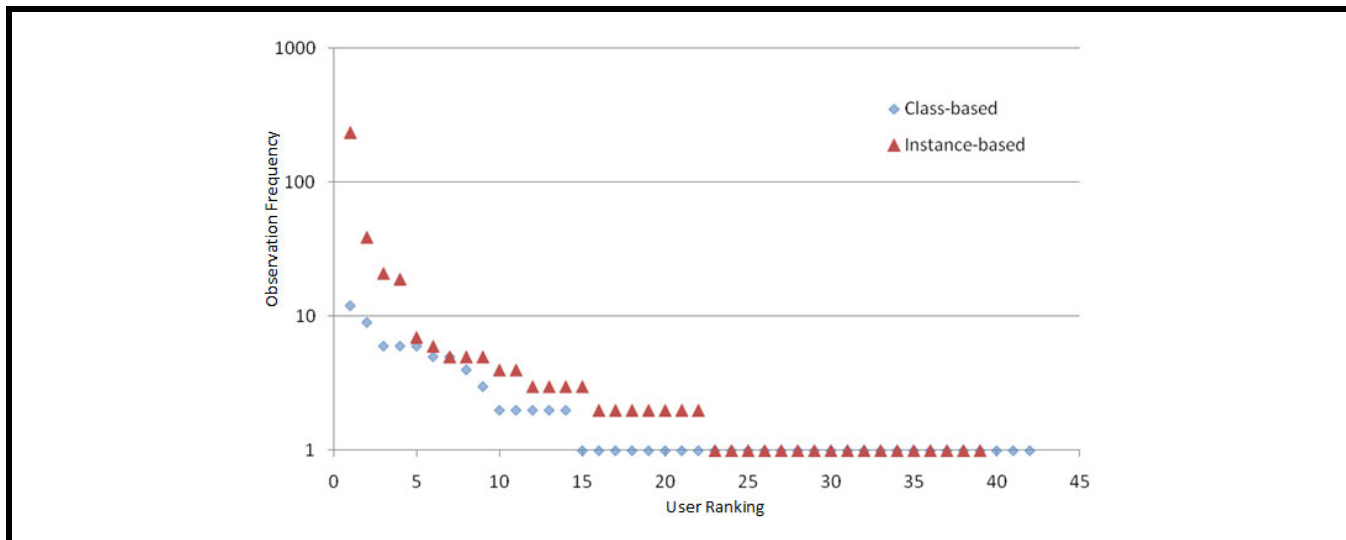
($W = 0.690$ and $p < 0.001$ for the class-based and $W = 0.244$ and $p < 0.001$ for the instance-based condition). We also note the presence of outliers in each condition.⁴ As seen from Table 2, in both cases the distributions are skewed and leptokurtic. This was confirmed using Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit statistics. Indeed, the top four contributors in the instance-based condition produced 80.8% of observations in that condition.

These results are not surprising: long-tail distributions have been observed consistently in other user-generated datasets (Brynjolfsson et al. 2016; Dewan and Ramaprasad 2012; Johnson et al. 2014; Meijer et al. 2009), including citizen science projects (Clow and Makriyannis 2011; Cooper et al. 2011). However, the instance-based condition has greater mean, variance, skewness and kurtosis (Table 2). Figure 4 shows that contributors in the instance-based condition tended to contribute a higher number of observations than those in the class-based condition, and fewer contributors in the instance-based condition contributed one or zero observations.

⁴We defined data points as outliers if they were 1.5*interquartile range above the third quartile or below the first quartile (Martinez et al. 2004). The following frequencies of observations per user are outliers in the instance-based condition: 236, 39, 21 and 19; and in the class-based condition 12, 9, 7, 7, 6, 6, 5, and 4. We also verified that the most extreme value is a significant outlier using Grubbs' test.

Table 2. Number of Observations by Condition

Experimental Condition	Number of Users	Observations				
		Total	Mean	St. Dev.	Skewness	Kurtosis
Class-based	42	87	2.07	2.56	2.08	4.23
Instance-based	39	390	10.0	37.83	5.47	29.66

**Figure 4. Observation Frequencies (y-axis) Ranked by User (x-axis) in Each Condition**

To determine if the difference in the number of observations per contributor is significant between conditions we used the exact permutation test, which diminishes the impact of absolute values (Good 2001) and is suitable when data are not normally distributed, sample sizes are low or medium, and outliers and ties (i.e., same values in two samples, as in Figure 4) are present. Based on the exact permutation test of observations per contributor between the two conditions, the p -value is 0.033, indicating that contributors in the instance-based condition provided significantly more observations than those in the species-based condition. As Figure 4 indicates, the instance-based condition produced a greater number of high and midrange contributors, and had a shorter tail. This supports Hypothesis 1.

As expected, the class-based interface produced significantly greater precision. Eighty-seven (93.5%) observations in this condition were at the species level (here, we included six additional observations that were reported at levels more general than species). In contrast, of the 390 observations in the instance-based condition, 179 (46%) were not classified at the species level ($\chi^2 = 49.44$, $p < 0.001$). This supports Hypothesis 2 and Proposition 2 (precision).

We further analyzed the categories and attributes provided to identify specific causes of lower performance in the class-based group. We observed three behavioral patterns contributing to lower dataset completeness. First, since the class-based interface constrains contributor input to predefined classes and attributes, contributors may not be able to record instances unless they provide classes that are congruent with the predefined structure in an IS. Evidence for this comes from a comparison of classes contributors attempted to enter in the dynamic textbox to the classes defined in the IS. While we specifically instructed contributors to provide species-level responses and identification at that level, the prevailing practice in natural history citizen science, some still attempted to provide classes at other levels. These were generally at higher levels in the classification hierarchy (e.g., “slug,” “earwig”), potentially reflecting classification uncertainty (e.g., due to conditions of observation) or lower levels of domain expertise.

The second observed pattern was that, when facing a structure incongruent with their own, some contributors changed the original submission. In several cases, this resulted in a failure to capture observations. For example, in one case a contrib-

utor began with typing “otter” (non-species level) and the entry was rejected by the system. This person then proceeded to record “Little Brown Bat” under the same observation ID. Another contributor began with “black bear scat,” and after two unsuccessful attempts, typed “Black Bear.” In all such cases, the original input was changed to conform to the classification choices available in the IS.

Finally, the lack of direct affordances to add new classes and attributes in class-based IS further hampered the ability to contribute. Our results offer some evidence for this. In six cases, contributors in the class-based condition bypassed species identification (presumably to provide a novel species level class), but then failed to provide any species-level labels (contributors might have become distracted). These cases were also excluded from the final count, further expanding the difference in the number of observations between the conditions.

These patterns were absent in the instance-based condition. Many classes provided in the instance-based condition were at taxonomic levels higher than species. As mentioned before, 179 (46%) of the observations in the instance-based condition were not classified at the species level. For these observations, contributors provided 583 classes and 69 attributes (222 distinct classes and 43 unique attributes). Among the classes provided, 110 were at the basic level (e.g., fly, bird, tree, fish, and spider). Thus, our field results are consistent with Proposition 1, as we expect basic level classes and attributes to be highly accurate (we further evaluate accuracy through a lab experiment, as described below).

Hypothesis 3 posited that a greater number of unanticipated species would be reported in the instance-based condition. Contributors in both conditions provided 997 attributes and classes including 87 in the class-based and 910 in the instance-based condition. Of these, 701 attributes and classes did not exist in the schema or data and were suggested additions. This was done directly by contributors in the instance-based condition and indirectly (via comments to an observation) by those in the class-based condition.

During the experiment, 126 unanticipated (i.e., not contained in the list of species available in the class-based condition) species-level classes were suggested, of which 119 came from contributors in the instance-based condition and seven from those in the class-based condition (see Table 3). In each condition, the distribution of unanticipated species per contributor significantly deviates from normal ($W = 0.430$ and $p < 0.001$ for the class-based condition and $W = 0.232$ and $p < 0.001$ for the instance-based condition). The distribution is long-tailed in the instance-based condition (using Kolmogorov–Smirnov

and Anderson–Darling goodness-of-fit) and uniform (Chi-squared = 47, Monte Carlo $p = 0.424$) in the class-based condition. Based on the exact permutation test, the number of unanticipated species is significantly greater in the instance-based condition ($p = 0.007$), supporting Hypothesis 3. This suggests that the instance-based approach to capturing data is more effective for capturing data about unanticipated phenomena of interest.

Contributors also provided interesting (and potentially useful to data consumers) attributes for some sightings. Many appeared to augment the classes provided and offered *additional information* not inferable from the classification labels:

- attributes describing behavior of the instances observed (e.g., mating, hopping)
- attributes describing something unusual about an instance (e.g., tagged, has one antler)
- attributes describing the environment/location of the instance (e.g., near highway, near bike trail)

As these attributes cannot be inferred from knowing the species, they constitute information beyond what could be collected in a traditional class-based interface.

Several sightings of biological significance were reported during the experiment. These included unexpected distributions of species (e.g., vagrant birds, fish, and insects). In addition, a species of mosquito new to the geographic area of the study was identified based on a reported sighting. The online sighting led to physical specimen collection by entomologists in the area where the sighting was reported. From this, a species of an Asian mosquito not previously recorded in this region of North America was confirmed (Fielden et al. 2015). Likewise, a possibly new (to the world) species of wingless wasp was reported, one with features not matching known species. Although this could not be confirmed from a web sighting, the sighting helps focus resources to obtain a specimen to confirm this discovery. Notably, all these occurred in the instance-based condition. Finally, some organisms suggested by the instance-based contributors belonged to groups underrepresented in the original project schema, such as insects.

Limitations of the Field Experiment and Further Hypothesis Development

The field experiment provided insights on the impact of different data collection approaches on information completeness. However, it had four notable limitations. First, despite efforts to attract members of the general public, the

Table 3. Number of Unanticipated Species Reported by Condition

Experimental Condition	Number of Users in Condition	Unanticipated Species				
		Total	Mean	St. Dev.	Skewness	Kurtosis
Class-based	42	7	0.17	0.44	2.53	5.96
Instance-based	39	119	3.05	13.17	5.35	28.51

majority of participants had a relatively high level of biology expertise. Much of the data provided was at the species level. This level of granularity is natural for domain experts, whereas novices are more comfortable with more generic classes (Lukyanenko et al. 2014).

Second, we could not determine what participants actually observed, making it impossible to assess observation *accuracy*. The greater number of observations reported in the instance-based condition during the field study might have come at the expense of accuracy.

Third, in the field setting we could not manipulate (or even measure confidently) information contributors' abilities. Specifically, a key assumption in citizen science UGC is familiarity with the domain of interest to the project. Empirical studies in UGC demonstrate that familiarity with the classes results in greater classification accuracy (Lewandowski and Specht 2015; Lukyanenko et al. 2014). Following the three propositions, we expect that when domain familiarity is high, both accuracy and completeness will be high, irrespective of the type of interface used. In contrast, when domain familiarity is low, we expect that accuracy and completeness will be high only when data contributors use an instance-based data collection interface. We hypothesize

Hypothesis 4: Accuracy. (a) Accuracy of observations in a class-based IS will be higher when a contributor has higher familiarity with the instances reported than when an observer has lower familiarity with the instances. (b) Accuracy of observations in an instance-based IS will be independent of the level of contributor familiarity with the instances reported.

While the list of classes in the field experiment was ecologically valid, it was also incomplete. We hypothesized the effects on completeness when schema is complete (i.e., has all the species classes for every instance observed), but participants may not be familiar with all classes. We thus predict that the number of instances reported will be higher in a class-based IS for those organisms highly familiar to contributors than for organisms unfamiliar to contributors. In contrast, we predict that in an instance-based IS, there will be no difference in the number of instances reported for familiar versus unfamiliar organisms. Hence, we propose

Hypothesis 5: Dataset completeness (number of instances reported). (a) The number of instances reported via a class-based IS will be higher when a contributor has higher familiarity with the instances reported than when a contributor has lower familiarity with the instances. (b) The number of instances reported via an instance-based IS will be independent of the level of contributor familiarity with the instances reported.

Finally, paralleling the field experiment, we hypothesize

Hypothesis 6: Information precision. The proportion of species-level observations reported using a class-based IS will be higher than the proportion of species-level observations reported via an instance-based IS.

Fourth, field settings offered limited ability to understand contributors' reactions to interfaces driven by different IQ approaches and explore psychological mechanisms that mediate the relationship between data collection interface features and IQ dimensions. A laboratory setting allows us to better understand the potential impact of class-based versus instance-based interfaces on self-reported measures of domain knowledge and behavioral intention (e.g., familiarity with local wildlife, self-reported biology expertise, ease-of-use of the interface, and probability of recording real sightings in the future using the IS). For example, lower dataset completeness in the class-based condition could be due to data contributors' perception of lower self-efficacy in the domain. It is reasonable to expect that, when forced to use unfamiliar classification choices (i.e., biological species) in the class-based condition, nonexperts felt inadequate in their ability to report information accurately. Indeed, we posit that imposing a class-based interface on data contributors undermines their perception of their own domain expertise in the domain. We also conjecture that a class-based interface will, in the context of low domain expertise, negatively affect usage intentions.

Laboratory Experiment

We conducted a follow-up laboratory study using the same NL Nature website employed in the field study. As in the

Table 4. Species Within Each of the Three Groupings in the Laboratory Experiment

High Familiarity	Moderate Familiarity	Low Familiarity
American Robin	Red-breasted nuthatch	Lapland Longspur
Common dandelion	Fireweed	Long's Braya
Blueberry	Crowberry	Bunchberry
Ladybird beetle	Common Black Beetle	Tiger Beetle
Blue Jay	Northern Flicker	Northern Shrike
Mallard Duck	Greater Scaup	Ring-necked Duck

field experiment, we randomly assigned study participants to either the class-based or the instance-based version of NL Nature.

Method

We created three sets of six images of plants and animals found in the region covered by NL Nature (Table 4). The organisms were selected by the third author, an ecology professor who is an expert in local natural history. Each set contained one insect, one flower, one berry, and three birds. Set 1 (high familiarity) included species we expected would be highly familiar and easy to identify accurately at the desired (species) level of precision. Set 2 (moderate familiarity) included species that were expected to be moderately familiar, in that they are highly abundant and commonly seen in the region, but which we expected nonexperts not to be as skilled in identifying. Set 3 (low familiarity) was comprised of unfamiliar species, either because they are extremely rare or not easily observed. We pretested our sets of images with 20 students and faculty to assess our groupings. In these pretests, we provided participants with randomly ordered photographs of organisms to be used in the study and asked them to sort them in three groups corresponding to our conditions. Participants consistently sorted the materials according to our expectations. We further verified familiarity in the manipulation check (discussed below).

Participants ($n = 108$) were students recruited from business classes to ensure lack of biology expertise. Each experimental session lasted approximately 20 minutes and took place in a computer lab. Participants were introduced to the NL Nature website and asked to set up an account. On first logging in, participants were randomly assigned to either the instance-based or class-based condition. In each session, one of the three sets of images (highly familiar, moderately familiar, unfamiliar) was used (i.e., we had six experimental conditions). Each image was shown on a screen for 40 seconds. This time was deemed adequate based on a pilot.

For each image, participants were asked to report the sighting using the NL Nature website. Following this, all six images were shown again on the screen and participants were asked to complete a questionnaire asking them how many of the organisms they felt they could identify and how many they were familiar with (i.e., they had seen before). They were also asked to indicate (on a seven-point Likert scale) their self-assessed familiarity and expertise in the natural history of the province, as well as their opinions of the website (ease of use, likelihood of contributing real sightings in the future). Finally, we asked a series of biographical questions (e.g., age, gender, university major). Table 5 summarizes the background of participants.

Our experimental groupings of the stimuli into three levels of familiarity were further validated by the self-reported results. When asked what number of organisms they felt they could identify, responses were in the predicted order (see Tables 6 and Table 7), with no significant difference between medium and low levels of familiarity. As well, the number of organisms deemed to appear "familiar" (i.e., they had seen them before) was consistent with our experimental groupings (Table 7), with significant differences between the three sets of stimuli, but not between experimental conditions (class versus instance-based). We also tested the interaction between the interface condition and the manipulation check variables and found no significant effects, indicating that manipulation was not affected by the data collection condition (e.g., participants perceived high familiarity higher than medium and low familiarity, irrespective of whether they were assigned to the instance-based or class-based IS).

Results

The data were cross-referenced to the images shown to assess accuracy. Significance testing for differences within and between experimental groupings and conditions was carried using MANCOVA with *post hoc* multiple comparisons using Tukey HSD, with $\alpha = 0.05$. To test our hypotheses, we

Table 5. Laboratory Experiment Participant Characteristics

Gender	Male	58
	Female	50
	Mean	Standard Deviation
Age	22.1	3.89
University-level Biology courses taken	0.3	0.66
Hours spent outdoor per week	11.1	9.80
Total years lived in local area	12.6	9.92
Familiarity with local wildlife (7 point scale)	4.0	1.77
Expertise in local biology (7 point scale)	2.2	1.36

Table 6. Main Effect Manipulation Check for Familiarity

Variable	F-Value (P-Value)	Data Entry Interface	Highly Familiar		Somewhat Familiar		Unfamiliar	
			Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Number of organisms can ID at species-level	20.366 (0.000)	Class-based	5.11	0.90	1.89	1.08	1.17	0.92
		Instance-based	5.06	1.31	1.89	1.41	1.47	1.63
Number of organisms seen before	13.182 (0.000)	Class-based	5.56	1.04	3.67	1.19	2.33	1.33
		Instance-based	5.33	0.91	3.56	1.50	1.88	1.58

Note: All demographic variables were used as covariates.

Table 7. Between Group Analysis of the Manipulation Check for the Familiarity Variable

Variable		CB-HF	CB-SF	CB-U	IB-HF	IB-SF	IB-U
Number of organisms can ID at species-level	CB-HF	5.11					
	CB-SF	***	1.89				
	CB-U	***		1.17			
	IB-HF		***	***	5.06		
	IB-SF	***			***	1.89	
	IB-U	***			***		1.44
Number of organisms seen before	CB-HF	5.56					
	CB-SF	***	3.67				
	CB-U	***	*	2.33			
	IB-HF		**	***	5.33		
	IB-SF	***			**	3.56	
	IB-U	***	**		***	**	1.89

Notes: CB: Class based condition; IB: Instance-based condition, HF: Highly familiar organisms group; SF: medium (somewhat) familiar organisms group; U: unfamiliar organisms group ±.

±Test based on *post hoc* Tukey analysis; diagonal cells show means.

*Significant at $p < 0.05$, **Significant at $p < 0.01$, ***Significant at $p < 0.001$.

analyzed all responses provided through the two interfaces across the three familiarity options. Accuracy was measured as the percentage of correct categorical responses, where a response was considered correct if it accorded with biological convention, irrespective of the taxonomic level at which the response was given. For example, “American Robin” and “bird” were both considered correct for American Robin, but a response of “American Robin” was coded as incorrect for Blue Jay. We excluded attributes from the comparison of the class-based versus instance-based condition as the former did not elicit attributes and, as we expected attributes to be mostly accurate (see results below), their inclusion would increase the difference between conditions even further. Completeness was measured as the percentage of times any response (correct or incorrect) was provided for a particular organism. In both interfaces, participants had the option to skip an image and not provide any classification labels. In the class-based condition, participants sometimes did not provide any data for organisms, as indicated in Table 8. Precision was measured as the percentage of classification responses (regardless of response correctness) at the species level.

The results provide strong evidence of the impact of data collection interface on accuracy, precision, and completeness (Table 8). The huge effect sizes further demonstrate the considerable impact of data collection choices on IQ. A comparison between the groups (Table 9) shows that high accuracy in the class-based interface was only attained in the highly familiar condition, and was extremely low (almost 0) in both moderately familiar and unfamiliar conditions. Similarly, dataset completeness was only high in the class-based interface for the highly familiar condition and declined for the other two conditions. In contrast, both accuracy and completeness were extremely high (close to 100%) in the instance-based IS across all levels of familiarity. Here, higher completeness and accuracy is a result of contributors being allowed to contribute data at the level at which they are familiar; many responses in the instance-based condition were at a generic (e.g., basic) level. It is also notable that the instance-based condition produced four times more responses than the class-based condition, where participants were limited to one classification label. However, as expected, the instance-based condition yielded few responses considered useful (i.e., with a high level of precision) to data consumers, and these were mostly for the familiar organisms.

Overall, the results provide support for all hypotheses and indicate a strong contingent relationship between accuracy, completeness, and information contributors’ domain familiarity for the class-based IS, but not the instance-based one. This indicates a greater fit of the instance-based IS with highly variable and heterogeneous capabilities of the crowd. At the same time, lower precision in this condition indicates

potential difficulty in making this data usable and useful for data consumers. As we argued using cognitive theory, humans are generally accurate when describing both familiar and unfamiliar objects using attributes. Our experiment offers strong support for this claim. To obtain accuracy of attributes, two of the authors of the paper independently coded 824 attributes reported in the three instance-based conditions as either correctly describing (1) or not (0) an organism shown on the image. The resulting Cohen’s kappa was 0.80, which is considered substantial agreement (Landis and Koch 1977). Inconsistencies were reconciled by an independent coder not familiar with the hypotheses.

Accuracy of attributes was 94.9% in the highly familiar, 98.4% in the somewhat familiar, and 98.4% in the unfamiliar conditions. This means that the attributes were generally accurate irrespective of the familiarity, mirroring the pattern observed for classes in the three instance-based conditions. This is in contrast with the class-based IS, where accuracy was strongly contingent on familiarity. We conclude that nonexperts can accurately describe familiar and unfamiliar phenomena using an instance-based IS via a combination of classes and attributes.

We also analyzed the exploratory variables. We did not find any evidence that class-based versus instance-based IS affected perceptions of the ease-of-use of the interface or the reported probability of using NL Nature in the future. This eliminates a potential alternative explanation for the overall findings in the laboratory and field experiments (differences in the usability of the two interfaces). However, the results somewhat surprisingly indicate that familiarity with the stimuli presented can affect perceptions of familiarity with the subject domain. In particular, there was a consistent, significant drop in the self-reported familiarity with wildlife and expertise in local biology between the highly familiar versus the somewhat and unfamiliar conditions. These results open new opportunities for future research.

Exploring the Usefulness of Instance-Based Data: An Applicability Check

In addition to determining whether attribute data could be transformed to a form useful to data consumers (species level classification for biologists, in our case), we conducted an applicability check (Rosemann and Vessey 2008) to explore perceptions about the potential uses and usefulness of data collected using an instance-based approach (versus a class-based approach) among potential consumers of UGC. We recruited participants from a university academic unit to which none of the authors belonged, and which also represented a discipline in which researchers are familiar with ap-

Table 8. MANCOVA Results for the Laboratory Experiment's Dependent Variables

Variable	F-test (p-value)	Data Entry Interface	Highly Familiar			Somewhat Familiar			Unfamiliar		
			Mean	Std. Dev.	Effect Size	Mean	Std. Dev.	Effect Size	Mean	Std. Dev.	Effect Size
Accuracy	156.567 (0.000)	Class-based	0.8	0.22	0.84	0.1	0.10	9.71	0.0	0.05	12.73
		Instance-based	1.0	0.04		1.0	0.06		1.0	0.07	
Completeness	20.560 (0.000)	Class-based	1.0	0.09	0.44	0.7	0.24	1.68	0.4	0.30	2.72
		Instance-based	1.0	0.00		1.0	0.00		1.0	0.00	
Precision	53.968 (0.000)	Class-based	1.0	0.00	4.19	1.0	0.00	13.82	1.0	0.00	12.28
		Instance-based	0.3	0.22		0.1	0.09		0.0	0.08	

Notes: Means and Standard Deviations are percentages scaled to 0-1. Effect sizes are corrected Cohen's D values, where < 0.2 is considered small, 0.5 is considered medium, 0.8 is considered large, and over 2 is considered huge (Sawilowsky 2009). Since samples $N < 50$ can inflate effect sizes, we applied a correction to compensate for this (Hackshaw 2008); the effect sizes remain very large after the correction.

Table 9. Between Group Analysis for the Laboratory Experiment's Dependent Variables±

Variable		CB-HF	CB-SF	CB-U	IB-HF	IB-SF	IB-U
Accuracy	CB-HF	0.83					
	CB-SF	***	0.12				
	CB-U	***		0.02			
	IB-HF	**	***	***	0.97		
	IB-SF	*	***	***		0.95	
	IB-U	*	***	***			0.96
Completeness	CB-HF	0.97					
	CB-SF	***	0.69				
	CB-U	***	***	0.40			
	IB-HF		***	***	1.00		
	IB-SF		***	***		1.00	
	IB-U		***	***			1.00
Precision	CB-HF	1.00					
	CB-SF		1.00				
	CB-U			1.00			
	IB-HF	***	***	***	0.33		
	IB-SF	***	***	***	***	0.05	
	IB-U	***	***	***	***		0.04

Notes: ±Test based on *post hoc* Tukey analysis; diagonal cells show means (scaled percentages); CB – class-based condition; IB-instance-based condition; HF – highly-familiar condition, SF – somewhat familiar condition, U – unfamiliar condition.

*Significant at $p < 0.05$; **Significant at $p < 0.01$; ***Significant at $p < 0.001$.

plying citizen science/crowdsourcing (a geography department). Participants were invited to attend a seminar and discussion forum that was advertised via departmental email lists, an events announcement webpage and posters in the building. Details on the participants and procedure are provided in Appendix A.

As part of the applicability check, we gave a short presentation in two parts: (1) a description of the class-based and instance-based approaches, and (2) a summary of our experiments and results. At the end of each part, we posed questions to participants to get feedback on two issues: the extent to which they perceived instance-based data collection to be relevant and applicable to the practice of citizen science, and the extent to which the results of the experiments were relevant and applicable to the practice of citizen science. In addition to oral comments, we distributed a questionnaire to get feedback on specific strengths and limitations of each data collection approach; 10 participants returned questionnaires (Table A1 of Appendix A contains both the questions and a summary of the responses).

Participants agreed that the instance-based data collection approach is relevant and applicable to the practice of citizen science (mean response of 6.0 on a seven-point scale, where seven was labeled “strongly agree”). Likewise, participants agreed that the results of the experiments we presented were relevant and applicable to the practice of citizen science (also 6.0 on a seven-point scale). These results indicate that participants viewed the instance-based approach as potentially valuable in the context of collecting citizen science data.

The discussion and questionnaire responses exposed further details about participants’ views on the data collection approaches and experimental findings. Specifically, participants noted the flexibility of the instance-based approach in accommodating unanticipated data. One respondent noted the approach would be useful “to document species diversity and new invasive species,” while another stated that it “can help you obtain data you never considered as important.” Participants also felt the instance-based approach encourages participation by people who are not familiar with the classes of interest to the researchers. One respondent stated that it “does not deter people who are not knowledgeable on the topic” and another indicated it will “allow nonexperts to contribute.” Two respondents commented on the applicability of the instance-based approach to projects they had worked on. In the first case, a researcher studying ocean plastics indicated that it was impossible to anticipate a complete set of classes for phenomena that citizen scientists might encounter on a beach. In the second case, a researcher studying deep sea sponges reflected on a failed project in which he worked with fishers who pulled sponges up with their fishing gear.

Despite being given keys and images to classify sponges by species, participants in that project were unable to provide high quality data due to the difficulty in using images to make positive species identification.

At the same time, concerns were expressed about the need for post-processing to make instance-based data useful for the goals of data consumers (scientists), and the likelihood that the data collected would be messy. As one respondent noted, “You might get a lot of data that are not very useful. You then might have to spend more time on data cleaning and post-processing.” Nonetheless, scientists recognized that this additional effort may be worthwhile considering the benefits (i.e., discovery, greater user participation) our approach may bring.

Exploring Usefulness of Instance-Based Data: Classification by Experts and Machines

An important concern arising from contributor-centric, instance-based data collection is the extent to which the resulting data are useful to data consumers. Since both the laboratory and field experiments show that precision suffers, it is important to know whether additional processing can produce accurate classifications at the desired level. To answer this question, we conducted a study in which biology experts were asked to infer classes based on attributes of observed organisms generated by citizen scientists. In addition, because domain expertise is scarce, we investigated the potential for classification using machine learning.⁵

To assess whether experts could accurately classify species based on attribute data, we conducted one-on-one interviews with local natural history experts. We selected 16 organisms from a dataset collected in a previous study and designed the interview as a “guessing game” (detailed procedure explained in Appendix B), in which attributes were revealed one at a time. After each attribute was revealed, participants were asked to identify the organism (if possible), indicate if they believed it was one of *N* species, and indicate their confidence in the level of precision provided (see Appendix B).

Overall, the experts were able to identify an average of 40.7% (± 10.3 s.d.) of the organisms based on the attributes provided (i.e., without being shown the photograph). Our experts tended to be experts in one taxonomic area (e.g., birds, but not plants), thus they could not always correctly classify the images (overall correct identification of specimens in the photos was 59.4% (s.d 14.7)). When we consider only cases

⁵Other approaches to moving from observations to classes of interest exist. For example, the classification task can be crowdsourced to other participants, as is done in iNaturalist (www.inaturalist.org).

when experts were able to positively identify the species after viewing the debriefing photograph, the experts were able to infer 59.79% (± 34.0 s.d.) of species from the attributes. The similarity in the correct classification scores based on attributes versus seeing the image suggests that when experts know a species group, they can classify it quite well based only on attributes provided by nonexperts. In addition, even for organisms for which experts had low to no correct classification, final precision was quite high, meaning that experts could come up with a limited list (usually less than five) of species that fit the attributes provided. Although perfect species-level identification may not always be possible based on attributes provided by nonexperts, a limited list (usually of closely related species) can still have utility for many ecological research questions. The results provide strong evidence of the utility of post-processing data generated by the instance-based approach for reducing classification uncertainty.

To prepare the data for machine learning (ML), we converted the attribute data set into an attribute matrix, where attributes were assigned “1” if a particular participant used that attribute to describe the species of interest and “0” otherwise. Each row represents attributes (with the associated basic-level category) provided by one of the 125 nonexpert data contributors for one of the organisms.

To ensure accessibility of our approach to data consumers, we applied a variety of common ML algorithms, including neural networks, support vector machines, random forests, boosting using decision trees, and naïve Bayes algorithms (Provost and Fawcett 2013). The average classification accuracy was above 70%. The top performing algorithm (Appendix B, Table B1) was a boosted decision tree classifier, which achieved an average F-measure of 0.76 (± 0.12 s.d.) across 16 species (based on 10-fold cross-validation and 50 boosting iterations).

A direct comparison between human and machine performance is not meaningful since ML worked with 16 finite targets (species), whereas experts had to draw from their knowledge of all possible organisms in a local area. However, the results suggest that, while the immediate data obtained from the instance-based approach may have low precision, it can be improved by human annotation and/or applying common, off-the shelf ML techniques.

Discussion

The four studies demonstrate advantages of contributor-focused IQ management supported by instance-based data collection for open data collection from nonexperts.

The instance-based approach appears to be more effective at capturing unanticipated phenomena. Contributors in the instance-based condition of the field experiment reported 17 times more observations of *unanticipated species* (new to the project) than those in the class-based condition, including the discovery of a new (to the region) species of mosquito and a possibly new (to science) species of wasp. Interestingly, some of the new instances logged by the instance-based contributors belonged to groups poorly represented in the project schema, including spiders, flies, and mosquitoes. These readily observable organisms were rarely reported in the four years the project operated prior to the experiment. A widely held assumption in natural history-focused citizen science holds that nonexperts mostly report “charismatic” organisms, fueling concerns that citizen science produces a distorted view of biodiversity (Boakes et al. 2010; Galloway et al. 2006). However, some of these distortions might be due to anchoring biases (Allen and Parsons 2010; Gigerenzer and Todd 1999; Gilovich et al. 2002) introduced during the data collection design. As we see from the class-based data, people stick closely to predefined options and fail to explore the domain of the project more fully. In contrast, open data collection that does not rely on a fixed class structure encourages discovery, as it does not present cues that might narrow the focus of data collection.

Importantly, the tendency of instance-based data collection to capture more unanticipated insights does not come at the expense of accuracy and dataset completeness. Indeed, the field and laboratory experiments strongly suggest that when given freedom, nonexperts are able to provide data with high accuracy and completeness. Moreover, when using instance-based data collection, accuracy is not contingent on expertise or familiarity, as nonexperts use a variety of generic classes and attributes to describe both familiar and unfamiliar organisms.

In contrast, the differences in accuracy and completeness within the class-based laboratory condition depended on familiarity with the organisms. This suggests that, when nonexperts are forced to classify at a level determined by scientific considerations (e.g., the species level), data quality suffers. Nonexperts can only contribute class-based data accurately when they are highly familiar with the classes in question. Hence, UGC projects that require contributors to classify at the level required by the project will either have to restrict participation to amateur experts (e.g., skilled bird watchers) or risk inaccurate or incomplete data, even when the participants are confident in the quality of their contributions.

Furthermore, being familiar with some domain objects does not guarantee familiarity with all of them. If the project is

collecting information on a broad domain (e.g., local nature, health, astronomy), it would be unlikely that even expert contributors would be familiar with every aspect of this domain (note our study with experts), especially when new objects and classes are likely to be observed. In such cases, an instance-based approach is advantageous. As we see from the laboratory results, accuracy in the familiar conditions is high in both approaches. However, it is low in the class-based unfamiliar condition, but remains high in the instance-based unfamiliar condition. As there is no guarantee that an object observed will be familiar even to the best experts, one can conclude that an instance-based approach can be effective in both cases, whereas a class-based approach is only appropriate for cases where people know what they are observing and the existing schema fully describes all the phenomena in the domain.

We also note limitations of the instance-based approach. The high accuracy of classification in the familiar condition of the laboratory experiment for both class-based and instance-based IS suggests that the instance-based approach holds no advantages when participants are familiar with the objects they are describing (and are able to use an existing project schema to communicate everything they wanted about these objects). A major limitation of the instance-based approach is low precision, resulting in the challenge of making data useful to experts. Nonetheless, in the study with experts and machines, we demonstrate that, in many cases, such data can be refined to a useful level of precision by post-processing.

Finally, the applicability check shows that potential data consumers are able to effectively assess the advantages and disadvantages of instance-based versus class-based data collection, consistent with the scenarios suggested in this paper (e.g., Table 1). This underscores the practical utility of our work for organizations interested in crowdsourcing.

Implications and Conclusions

This paper contributes to the theory and practice of UGC, crowdsourcing, information quality, and design of data collection instruments. To the best of our knowledge, ours is the first attempt to answer the question of how to design IS to capture unanticipated data about phenomena from crowds in a way that maintains IQ (e.g., accuracy, completeness, precision).

As UGC is being adopted in commercial, public, scientific, and medical domains (Andriole 2010; Culnan et al. 2010; Erickson et al. 2012; Hemsley and Mason 2012), persistent concerns about the quality of data generated by the crowd remain. We show that one way to address the research ques-

tion posed in this paper is by adopting a new perspective on IQ. A theoretical and practical contribution of this paper is the conceptualization of contributor-focused IQ management, which reconsiders traditional approaches to IQ (and IS design as a byproduct). This conceptualization embraces heterogeneity and low domain expertise in the crowds and views these characteristics as potential strengths that can be harnessed using innovative IS designs. Supported by empirical evidence of its advantages (and limitations), contributor-focused IQ advances the theory of IQ, a topic of core importance to IS (Larsen and Bong 2016, Appendix B; Petter et al. 2013) that until now has been dominated by a single perspective—fitness-for-use (Shankaranarayanan and Blake 2017; Wang and Strong 1996). Yet, this characterization of IQ is limited when information uses are evolving or unanticipated, and when discovery is desired. Since more information production now occurs in environments with weak controls (e.g., in social media, online, in crowdsourcing settings) and data repurposing is rampant (e.g., due to the rise of data analytics), the contributor-centric conceptualization of IQ is a timely addition to IS theory.

Our work further contributes to theory and practice by proposing boundaries for the class-based and instance-based IQ management and data collection. Specifically, the latter is most effective in projects that are open, with many unknowns, whereas the former is most suitable for closed projects with well-established and stable uses of data (see Table 1). We empirically show that each approach comes with trade-offs, which organizations can consider when choosing to implement a crowdsourcing IQ strategy and design a system to collect crowd input. The limitations of both instance and class-based approaches also serve as a motivation for future research to investigate ways of mitigating them, while leveraging the advantages of a given approach. We encourage researchers to explore new ways to implement contributor-centric IQ in IS development. While we examined several core dimensions of IQ, future research should also consider the implications of this new perspective on other IQ dimensions.

Our work shows that the traditional IQ approach manifested via class-based IS, while promoting information utility (e.g., desired level of input precision), may be problematic in open UGC, including scenarios beyond citizen science (Appendix C provides several examples). First, many contributors are nonexperts and thus may be unable to report on unfamiliar things in the domain, and when forced to do so, may resort to guessing, resulting in lower accuracy. Second, we showed that focusing on the views of data consumers may come at a cost of capturing less data and making fewer discoveries of unanticipated phenomena. These findings were consistent across field and laboratory settings. Many UGC projects wish to be more inclusive and promote wider participation. Yet,

technology can be a barrier to diversification of participation (Newman et al. 2012). We show that it is not only possible to maintain high accuracy, but accommodate broader audiences, thereby expanding the potential of UGC.

We also explore whether organizations can take advantage of the higher accuracy and completeness of instance-based data. We addressed this issue in a study in which domain experts in biology were asked to infer classes based on attributes of observed organisms generated by citizen scientists. In addition, because domain expertise is a scarce resource and does not scale in large datasets, we also investigated the potential for classification using machine learning. The results demonstrate that experts are able to leverage the attributes provided by nonexperts to infer classes that are more specific than those familiar to the nonexpert participants. The results point to the practical usefulness of asking nonexperts to provide data at the level with which they are comfortable, and subsequently inferring classes of interest to information consumers. Further, the potential to automatically infer classes can be used in a machine-to-contributor dialog, whereby an artificial agent may ask in real-time additional verification questions to increase the confidence in the ML-generated judgements and flag any unusual observations for further analysis. Future research should also examine challenges that data consumers (e.g., scientists) face when interpreting and analyzing instance-based data, and how they can better take advantage of this form of data.

According to a recent *MIS Quarterly* editorial, a major IS challenge is “addressing the rigidity of data categorization schemas” (Rai 2016, p. vi) to allow for unanticipated phenomena to be captured, while preserving stability of the overall system. In addition to uses in crowdsourcing, organizations can leverage the insights from our work to supplement their traditional processes for data collection within the organization. This is of growing importance as organizations are becoming more agile in their processes, and encourage employees to be more proactive (Gebauer and Schober 2006; Tax and Brown 1998).

Acknowledgments

We wish to thank Daria Lukyanenko for her assistance in conducting the field and laboratory experiments. We also wish to thank the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, GEOIDE Network of Centres of Excellence Canada, The Institute for Data Valorisation (IVADO), and Memorial University's Harris Centre for providing funding in support of this project. Finally, we wish to thank the anonymous contributors of sightings to the NL Nature project.

References

- Abbasi, A., Zhou, Y., Deng, S., and Zhang, P. 2018. “Text Analytics to Support Sense-Making in Social Media: A Language-Action Perspective,” *MIS Quarterly* (42:2), pp. 1-38.
- Allen, G., and Parsons, J. 2010. “Is Query Reuse Potentially Harmful? Anchoring and Adjustment in Adapting Existing Database Queries,” *Information Systems Research* (21:1), pp. 56-77.
- Andriole, S. J. 2010. “Business Impact of Web 2.0 Technologies,” *Communications of the ACM* (53:12), pp. 67-79.
- Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. 2009. “Methodologies for Data Quality Assessment and Improvement,” *Computing Surveys* (41:3), pp. 1-52.
- Bell, B. F. 1981. “When Is an Animal, Not an Animal?,” *Journal of Biological Education* (15:3), pp. 213-218.
- Berlin, B., Breedlove, D. E., and Raven, P. H. 1966. “Folk Taxonomies and Biological Classification,” *Science* (154:3746), pp. 273-275.
- Boakes, E. H., McGowan, P. J., Fuller, R. A., Chang-Qing, D., Clark, N. E., O'Connor, K., and Mace, G. M. 2010. “Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data,” *PLoS Biology* (8:6), pp. 1-11.
- Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., and Shirk, J. 2009. “Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy,” *BioScience* (59:11), pp. 977-984.
- Brabham, D. C. 2013. *Crowdsourcing*, Cambridge, MA: MIT Press.
- Browne, G. J., and Ramesh, V. 2002. “Improving Information Requirements Determination: A Cognitive Perspective,” *Information & Management* (39:8), pp. 625-645.
- Brynjolfsson, E., Geva, T., and Reichman, S. 2016. “Crowd-Squared: Amplifying the Predictive Power of Search Trend Data,” *MIS Quarterly* (40:4), pp. 941-961.
- Bunge, M. 1977. *Treatise on Basic Philosophy: Ontology I: The Furniture of the World*, Boston: Reidel.
- Burgess, H., DeBey, L., Froehlich, H., Schmidt, N., Theobald, E., Ettinger, A., HilleRisLambers, J., Tewksbury, J., and Parrish, J. 2017. “The Science of Citizen Science: Exploring Barriers to Use as a Primary Research Tool,” *Biological Conservation* (108), pp. 113-120.
- Burton-Jones, A., and Volkoff, O. 2017. “How Can We Develop Contextualized Theories of Effective Use? A Demonstration in the Context of Community-Care Electronic Health Records,” *Information Systems Research* (28:3), pp. 468-489.
- Callaghan, C. T., and Gawlik, D. E. 2015. “Efficacy of eBird Data as an Aid in Conservation Planning and Monitoring,” *Journal of Field Ornithology* (1:1), pp. 1-7.
- Chen, P. 1976. “The Entity-Relationship Model—Toward a Unified View of Data,” *ACM Transactions on Database Systems* (1:1), pp. 9-36.
- Chittilappilly, A. I., Chen, L., and Amer-Yahia, S. 2016. “A Survey of General-Purpose Crowdsourcing Techniques,” *IEEE Transactions on Knowledge and Data Engineering* (28:9), pp. 2246-2266.

- Clarke, R., Burton-Jones, A., and Weber, R. 2016. "On the Ontological Quality and Logical Quality of Conceptual-Modeling Grammars: The Need for a Dual Perspective," *Information Systems Research* (27:2), pp. 365-382.
- Clery, D. 2011. "Galaxy Zoo Volunteers Share Pain and Glory of Research," *Science* (333:6039), pp. 173-175.
- Clow, D., and Makriyannis, E. 2011. *ISpot Analysed: Participatory Learning and Reputation*, in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, New York: ACM Press, pp. 34-43.
- Cooper, A. K., Rapant, P., Hjelmager, J., Laurent, D., Iwaniak, A., Coetzee, S., Moellering, H., and Dören, U. 2011. "Extending the Formal Model of a Spatial Data Infrastructure to Include Volunteered Geographical Information," in *Proceedings of the 25th International Cartographic Conference*, Paris, July 3-8, Paper ID CO-120.
- Culnan, M. J., McHugh, P. J., and Zubillaga, J. I. 2010. "How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value," *MIS Quarterly Executive* (9:4), pp. 243-259.
- Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., and Allahbakhsh, M. 2018. "Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions," *ACM Computing Surveys* (51:1), Article 7.
- Deng, X., Joshi, K., and Galliers, R. D. 2016. "The Duality of Empowerment and Marginalization in Microtask Crowdsourcing: Giving Voice to the Less Powerful through Value Sensitive Design," *MIS Quarterly* (40:2), pp. 279-302.
- Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., and Shi, W. 2010. "Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities," *Transactions in GIS* (14:4), pp. 387-400.
- Dewan, S., and Ramaprasad, J. 2012. "Research Note—Music Blogging, Online Sampling, and the Long Tail," *Information Systems Research* (23:3, Part 2), pp. 1056-1067.
- Doan, A., Ramakrishnan, R., and Halevy, A. Y. 2011. "Crowdsourcing Systems on the World-Wide Web," *Communications of the ACM* (54:4), pp. 86-96.
- Engel, S., and Voshell, R. 2002. "Volunteer Biological Monitoring: Can It Accurately Assess the Ecological Condition of Streams?," *American Entomologist* (48:3), pp. 164-177.
- Erickson, L., Petrick, I., and Trauth, E. 2012. "Hanging with the Right Crowd: Matching Crowdsourcing Need to Crowd Characteristics," in *Proceedings of the 19th Americas Conference on Information Systems*, Seattle, Washington, August 9-12.
- Fielden, M. A., Chaulk, A. C., Bassett, K., Wiersma, Y. F., Erbland, M., Whitney, H., and Chapman, T. W. 2015. "Aedes Japonicus Japonicus (Diptera: Culicidae) Arrives at the Most Easterly Point in North America," *The Canadian Entomologist* (147:6), pp. 737-740.
- Galloway, A. W. E., Tudor, M. T., and Haegen, W. M. V. 2006. "The Reliability of Citizen Science: A Case Study of Oregon White Oak Stand Surveys," *Wildlife Society Bulletin* (34:5), pp. 1425-1429.
- Gao, G. G., Greenwood, B. N., Agarwal, R., and McCullough, J. S. 2015. "Vocal Minority and Silent Majority: How Do Online Ratings Reflect Population Perceptions of Quality?," *MIS Quarterly* (39:3), pp. 565-589.
- Garcia-Molina, H., Joglekar, M., Marcus, A., Parameswaran, A., and Verroios, V. 2016. "Challenges in Data Crowdsourcing," *IEEE Transactions on Knowledge and Data Engineering* (28:4), pp. 901-911.
- Gebauer, J., and Schober, F. 2006. "Information System Flexibility and the Cost Efficiency of Business Processes," *Journal of the AIS* (7:3), p. 8.
- Gigerenzer, G., and Todd, P. M. 1999. *Simple Heuristics That Make Us Smart*, New York: Oxford University Press.
- Gilovich, T., Griffin, D., and Kahneman, D. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge, UK: Cambridge University Press.
- Goes, P. B. 2014. "Editor's Comments: Design Science Research in Top Information Systems Journals," *MIS Quarterly* (38:1), iii-viii.
- Good, P. I. 2001. *Resampling Methods: A Practical Guide to Data Analysis. Second Edition*, Berlin, Germany: Springer Science & Business Media.
- Goodman, J. K., and Paolacci, G. 2017. "Crowdsourcing Consumer Research," *Journal of Consumer Research* (44:1), pp. 196-210.
- Gray, M. L., Suri, S., Ali, S. S., and Kulkarni, D. 2016. "The Crowd Is a Collaborative Network," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, New York: ACM Press, pp. 134-147.
- Hackshaw, A. 2008. "Small Studies: Strengths and Limitations," *European Respiratory Journal* (32), pp. 1141-1143.
- Harnad, S. 2005. "To Cognize Is to Categorize: Cognition Is Categorization," in *Handbook of Categorization in Cognitive*, H. Cohen and C. Lefebvre (eds.), Amsterdam: Elsevier Science, pp. 20-45.
- Hemsley, J., and Mason, R. M. 2012. "The Nature of Knowledge in the Social Media Age: Implications for Knowledge Management Models," in *Proceedings of the 45th Hawaii International Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society Press, pp. 3928-3937.
- Ipeirotis, P. G., Provost, F., and Wang, J. 2010. "Quality Management on Amazon Mechanical Turk," in *Proceedings of the ACM SIGKDD Workshop on Human Computation*, New York: ACM Press, pp. 64-67.
- Johnson, S. L., Faraj, S., and Kudaravalli, S. 2014. "Emergence of Power Laws in Online Communities: The Role of Social Mechanisms and Preferential Attachment," *MIS Quarterly* (38:3), pp. 795-808.
- Kallinikos, J., and Tempini, N. 2014. "Patient Data as Medical Facts: Social Media Practices as a Foundation for Medical Knowledge Creation," *Information Systems Research* (25:4), pp. 817-833.
- Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., and Popović, Z. 2011. "Crystal Structure of a Monomeric Retroviral Protease Solved by Protein Folding Game Players," *Nature Structural & Molecular Biology* (18:10), pp. 1175-1177.
- Kilian, D. 2008. "When it Absolutely Has to Be Accurate, Don't Trust the Crowd," ECT News Network.

- Kosmala, M., Wiggins, A., Swanson, A., and Simmons, B. 2016. "Assessing Data Quality in Citizen Science," *Frontiers in Ecology and the Environment* (14:10), pp. 551-560.
- Krogh, B., Levy, S., Dutoit, A., and Subrahmanian, E. 1996. "Strictly Class-Based Modelling Considered Harmful," in *Proceedings of the 29th Hawaii International Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society Press, pp. 242-250.
- Landis, J. R., and Koch, G. G. 1977. "The Measurement of Observer Agreement for Categorical Data," *Biometrics* (33:1), pp. 159-174.
- Larsen, K., and Bong, C. H. 2016. "A Tool for Addressing Construct Identity in Literature Reviews and Meta-Analyses," *MIS Quarterly* (40:3), pp. 529-552.
- Lassaline, M. E., Wisniewski, E. J., and Medin, D. L. 1992. "Basic Levels in Artificial and Natural Categories: Are All Basic Levels Created Equal?," in *Percepts, Concepts and Categories: The Representation and Processing of Information (Volume 93)*, *Advances in Psychology*, B. Barbara (ed.), Amsterdam: North-Holland, pp. 328-378.
- Lee, Y. W. 2003. "Crafting Rules: Context-Reflective Data Quality Problem Solving," *Journal of Management Information Systems* (20:3), pp. 93-119.
- Lee, Y. W., Pipino, L. L., Funk, J. D., and Wang, R. Y. 2006. *Journey to Data Quality*, Cambridge, MA: MIT Press.
- Leonardi, P. 2011. "When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies," *MIS Quarterly* (35:1), pp. 147-167.
- Levy, M., and Germonprez, M. 2017. "The Potential for Citizen Science in Information Systems Research," *Communications of the AIS* (40), Article 2.
- Lewandowski, E., and Specht, H. 2015. "Influence of Volunteer and Project Characteristics on Data Quality of Biological Surveys," *Conservation Biology* (29:3), pp. 713-723.
- Li, G., Wang, J., Zheng, Y., and Franklin, M. 2016. "Crowd-sourced Data Management: A Survey," *IEEE Transactions on Knowledge and Data Engineering* (28:9), pp. 2296-2319.
- Louv, R., Dickinson, J. L., and Bonney, R. 2012. *Citizen Science: Public Participation in Environmental Research*, Ithaca, NY: Cornell University Press.
- Love, J., and Hirschheim, R. 2017. "Crowdsourcing of Information Systems Research," *European Journal of Information Systems* (26:4), pp. 315-332.
- Lukyanenko, R., Parsons, J., and Samuel, B. M. 2019. "Representing Instances: The Case for Reengineering Conceptual Modeling Grammars," *European Journal of Information Systems* (28:1), pp. 68-90.
- Lukyanenko, R., Parsons, J., and Wiersma, Y. 2014. "The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content," *Information Systems Research* (25:4), pp. 669-689.
- Lukyanenko, R., Parsons, J., and Wiersma, Y. 2016. "Emerging Problems of Data Quality in Citizen Science," *Conservation Biology* (30:3), pp. 447-449.
- Lukyanenko, R., Parsons, J., Wiersma, Y. F., Wachinger, G., Huber, B., and Meldt, R. 2017. "Representing Crowd Knowledge: Guidelines for Conceptual Modeling of User-Generated Content," *Journal of the AIS* (18:4), pp. 297-339.
- Lukyanenko, R., Parsons, J., Wiersma, Y., Sieber, R., and Maddah, M. 2016. "Participatory Design for User-Generated Content: Understanding the Challenges and Moving Forward," *Scandinavian Journal of Information Systems* (28:1), pp. 37-70.
- Madnick, S. E., Wang, R. Y., Lee, Y. W., and Zhu, H. 2009. "Overview and Framework for Data and Information Quality Research," *Journal of Data and Information Quality* (1:1), pp. 1-22.
- Markus, M. L., and Silver, M. S. 2008. "A Foundation for the Study of IT Effects: A New Look at DeSanctis and Poole's Concepts of Structural Features and Spirit," *Journal of the AIS* (9:10/11), pp. 609-632.
- Martinez, W. L., Martinez, A., and Solka, J. 2004. *Exploratory Data Analysis with MATLAB*, New York, NY: Taylor & Francis.
- Mayden, R. L. 2002. "On Biological Species, Species Concepts and Individuation in the Natural World," *Fish and Fisheries* (3:3), pp. 171-196.
- McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., Evans, D. M., French, R. A., Parrish, J. K., and Phillips, T. B. 2016. "Citizen Science Can Improve Conservation Science, Natural Resource Management, and Environmental Protection," *Biological Conservation* (208), pp. 15-28.
- Meijer, A., Burger, N., and Ebbens, W. 2009. "Citizens4Citizens: Mapping Participatory Practices on the Internet," *Electronic Journal of E-Government* (7:1), pp. 99-112.
- Mervis, C. B., and Rosch, E. 1981. "Categorization of Natural Objects," *Annual Review of Psychology* (32:1), pp. 89-115.
- Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., and Crowston, K. 2012. "The Future of Citizen Science: Emerging Technologies and Shifting Paradigms," *Frontiers in Ecology and the Environment* (10:6), pp. 298-304.
- Norman, D. A. 1983. "Some Observations on Mental Models," *Mental Models* (7:112), pp. 7-14.
- Nov, O., Arazy, O., and Anderson, D. 2014. "Scientists@ Home: What Drives the Quantity and Quality of Online Citizen Science Participation," *PloS One* (9:4), pp. 1-11.
- Olivé, A. 2007. *Conceptual Modeling of Information Systems*, Berlin: Springer Science & Business Media.
- Parreiras, F. S., and Staab, S. 2010. "Using Ontologies with UML Class-Based Modeling: The TwoUse Approach," *Data & Knowledge Engineering* (69:11), pp. 1194-1207.
- Parsons, J., Lukyanenko, R., and Wiersma, Y. 2011. "Easier Citizen Science Is Better," *Nature* (471:7336), pp. 37-37.
- Parsons, J., and Wand, Y. 2000. "Emancipating Instances from the Tyranny of Classes in Information Modeling," *ACM Transactions on Database Systems* (25:2), pp. 228-268.
- Petter, S., DeLone, W., and McLean, E. R. 2013. "Information Systems Success: The Quest for the Independent Variables," *Journal of Management Information Systems* (29:4), pp. 7-62.
- Pipino, L. L., Lee, Y. W., and Wang, R. Y. 2002. "Data Quality Assessment," *Communications of the ACM* (45:4), pp. 211-218.
- Prestopnik, N. R., and Tang, J. 2015. "Points, Stories, Worlds, and Diegesis: Comparing Player Experiences in Two Citizen Science Games," *Computers in Human Behavior* (52), pp. 492-506.

- Provost, F., and Fawcett, T. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, Sebastopol, CA: O'Reilly Media, Inc.
- Prpić, J., Shukla, P. P., Kietzmann, J. H., and McCarthy, I. P. 2015. "How to Work a Crowd: Developing Crowd Capital through Crowdsourcing," *Business Horizons* (58:1), pp. 77-85.
- Rai, A. 2016. "Editor's Comments: Synergies Between Big Data and Theory," *MIS Quarterly* (40:1), iii-ix.
- Redman, T. C. 1996. *Data Quality for the Information Age*, Norwood, MA: Artech House.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyesbraem, P. 1976. "Basic Objects in Natural Categories," *Cognitive Psychology* (8:3), pp. 382-439.
- Rosemann, M., and Vessey, I. 2008. "Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks," *MIS Quarterly* (32:1), pp. 1-22.
- Saghafi, A. 2016. "Ontological and Cognitive Principles on Information Systems Modelling," Doctoral Dissertation, University of British Columbia.
- Samuel, B. M., Khatri, V., and Ramesh, V. 2018. "Exploring the Effects of Extensional Versus Intentional Representations on Domain Understanding," *MIS Quarterly* (42:4), pp. 1186-1209.
- Sawilowsky, S. S. 2009. "New Effect Size Rules of Thumb," *Journal of Modern Applied Statistical Methods* (8:2), pp. 467-474.
- Sekhavat, Y., and Parsons, J. 2012. "Sliced Column-Store (SCS): Ontological Foundations and Practical Implications," in *Proceedings of the 31st International Conference on Conceptual Modeling*, Florence, Italy, October 15-18, pp. 102-115.
- Shankaranarayanan, G., and Blake, R. 2017. "From Content to Context: The Evolution and Growth of Data Quality Research," *Journal of Data and Information Quality* (8:2).
- Shanks, G., Tansley, E., Nuredini, J., Tobin, D., and Weber, R. 2008. "Representing Part-Whole Relations in Conceptual Modeling: An Empirical Evaluation," *MIS Quarterly* (32:3), pp. 553-573.
- Sheng, V. S., Provost, F., and Ipeirotis, P. G. 2008. "Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers," in *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp. 614-622.
- Silvertown, J. 2009. "A New Dawn for Citizen Science," *Trends in Ecology & Evolution* (24:9), pp. 467-471.
- Strong, D. M., Lee, Y. W., and Wang, R. Y. 1997. "Data Quality in Context," *Communications of the ACM* (40:5), pp. 103-110.
- Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., and Kelling, S. 2009. "EBird: A Citizen-Based Bird Observation Network in the Biological Sciences," *Biological Conservation* (142:10), pp. 2282-2292.
- Surowiecki, J. 2005. *The Wisdom of Crowds*, New York: Anchor Books.
- Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., and Packer, C. 2015. "Snapshot Serengeti, High-Frequency Annotated Camera Trap Images of 40 Mammalian Species in an African Savanna," *Scientific Data* (2), Article 150026.
- Tax, S. S., and Brown, S. W. 1998. "Recovering and Learning from Service Failure," *MIT Sloan Management Review* (40:1), pp. 75-88.
- Tilly, R., Posegga, O., Fischbach, K., and Schoder, D. 2016. "Towards a Conceptualization of Data and Information Quality in Social Information Systems," *Business & Information Systems Engineering* (59:1), pp. 3-21.
- Trevino, L. K., and Victor, B. 1992. "Peer Reporting of Unethical Behavior: A Social Context Perspective," *Academy of Management Journal* (35:1), pp. 38-64.
- Tucker, A. L., and Edmondson, A. C. 2003. "Why Hospitals Don't Learn from Failures: Organizational and Psychological Dynamics That Inhibit System Change," *California Management Review* (45:2), pp. 55-72.
- Tversky, A., and Kahneman, D. 1974. "Judgment under Uncertainty: Heuristics and Biases," *Science* (185:4157), pp. 1124-1131.
- Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly* (27:3), pp. 425-478.
- Wallis, G., and Bülhoff, H. 1999. "Learning to Recognize Objects," *Trends in Cognitive Sciences* (3:1), pp. 22-31.
- Wand, Y., and Wang, R. Y. 1996. "Anchoring Data Quality Dimensions in Ontological Foundations," *Communications of the ACM* (39:11), pp. 86-95.
- Wang, R. Y., Reddy, M. P., and Kon, H. B. 1995. "Toward Quality Data: An Attribute-Based Approach," *Decision Support Systems* (13:3-4), pp. 349-372.
- Wang, R. Y., and Strong, D. M. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems* (12:4), pp. 5-33.
- Wells, J. D., Valacich, J. S., and Hess, T. J. 2011. "What Signals Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions," *MIS Quarterly* (35:2), pp. 373-396.
- Wiggins, A., Bonney, R., Graham, E., Henderson, S., Kelling, S., LeBuhn, G., Litauer, R., Lots, K., Michener, W., and Newman, G. 2013. "Data Management Guide for Public Participation in Scientific Research," DataOne Working Group.
- Wiggins, A., and Crowston, K. 2011. "From Conservation to Crowdsourcing: A Typology of Citizen Science," in *Proceedings of the 44th Hawaii International Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society Press.
- Wiggins, A., Newman, G., Stevenson, R. D., and Crowston, K. 2011. "Mechanisms for Data Quality and Validation in Citizen Science," in *Proceedings of the Computing for Citizen Science Workshop, IEEE eScience Conference*, Stockholm, pp. 14-19.

About the Authors

Roman Lukyanenko is an assistant professor in the Department of Information Technologies at HEC Montreal, Canada. Roman obtained his Ph.D. from Memorial University of Newfoundland. His research interests include conceptual modeling, ontological foundations of information systems, information quality, citizen science, crowdsourcing, machine learning, design science research, and research methodology (research validities, instantiation validity, and artifact sampling). In addition to *MIS Quarterly*, Roman's work has been published in *Nature*, *Information Systems Research*,

Journal of the AIS, and *European Journal of Information Systems*, among others. Roman served as a Vice President of the AIS Special Interest Group on Systems Analysis and Design.

Jeffrey Parsons is University Research Professor and Professor of Information Systems in the Faculty of Business Administration at Memorial University of Newfoundland. His research interests include conceptual modeling, crowdsourcing, information quality, and recommender systems. His work has appeared in many outlets, including *MIS Quarterly*, *Management Science*, *Information Systems Research*, *ACM Transactions on Database Systems*, *IEEE Transactions on Knowledge and Data Engineering*, and *Nature*. Jeff is a senior editor for *MIS Quarterly*, a former senior editor for *Journal of the AIS*, and has served as program co-chair for a number of major information systems conferences.

Yolanda F. Wiersma is Professor of Biology at Memorial University of Newfoundland. Her research area is landscape ecology

with applications focused on forestry, wildlife management, and protected areas. She also conducts interdisciplinary research on citizen science. Her research has been published in many venues, including *Landscape Ecology*, *Ecology Letters*, *Biological Conservation*, *Biodiversity and Conservation*, *Conservation Biology*, and *Nature*. She is an associate editor for *Journal of Applied Ecology* as well as a coordinating editor for the journal *Landscape Ecology*.

Mahed Maddah is an assistant professor in the Information Systems and Operations Management Department of the Sawyer Business School, Suffolk University. Mahed obtained his Ph.D. from the Information Systems and Business Analytics department at Florida International University. In addition to *MIS Quarterly*, Mahed's work has been published in the *Scandinavian Journal of Information Systems* as well as leading information systems and business conferences. His research interests include data quality, user-generated content, social media analytics, human– computer interaction, design science research and cognitive psychology.

EXPECTING THE UNEXPECTED: EFFECTS OF DATA COLLECTION DESIGN CHOICES ON THE QUALITY OF CROWDSOURCED USER-GENERATED CONTENT

Roman Lukyanenko

Department of Information Technologies, HEC Montréal,
Montréal, QC CANADA H3T 2A7 {roman.lukyanenko@hec.ca}

Jeffrey Parsons

Faculty of Business Administration, Memorial University of Newfoundland,
St. John's, NL CANADA A1B 3X5 {jeffreyp@mun.ca}

Yolanda F. Wiersma

Department of Biology, Memorial University of Newfoundland,
St. John's, NL CANADA A1B 3X5 {ywiersma@mun.ca}

Mahed Maddah

Sawyer Business School, Suffolk University,
Boston, MA 02108 U.S.A. {mmaddah@suffolk.edu}

Appendix A

Applicability Check Details

This appendix describes our applicability check in more detail. The purpose of the applicability check (Rosemann and Vessey 2008) was to determine whether attribute data could be transformed to a form (in this case, species level classification) useful to data consumers (in this case, biologists). We also used the applicability check to explore perceptions that biologists in a university setting held about the potential uses and usefulness of data collected using an instance-based approach (versus a class-based approach). The applicability check is discussed briefly in the main manuscript; here, we provide details about the method we used to collect data, and the feedback we received from participants.

Method

We tested the applicability of an attribute-based data collection approach to users of UGC via an interactive seminar presentation. We made the presentation as part of a seminar series in the Department of Geography at Memorial University of Newfoundland, as geography is a field in which there is considerable interest in crowdsourced UGC (referred to by geographers as volunteered geographic information) and none of the authors are affiliated with the department. We developed a questionnaire, which was distributed in paper form, on the tables where audience members sat. The questionnaire asked six open-ended questions about the perceived benefits and limitations of both instance-based versus class-based approaches, as well as about potential applications of the instance-based approach to the respondent's own research. In addition, there were two questions asking respondents to rank (on a seven-point Likert scale) their agreement with two statements, one about the relevance and applicability of the instance-based data collection approach and the other about the relevance and applicability of the experimental findings we presented. The questionnaire also included some biographical questions (gender, position, research field, highest degree obtained).

The third author introduced the topic and format of the seminar and outlined how the website NL Nature had been harnessed by the research team to examine questions about data quality in citizen science. The second author then proceeded to outline the concept of instance-based versus class-based approaches to citizen science data collection. About 20 minutes into the presentation, a slide with the question “From your perspective (please state that perspective), in what situations (if any) would each approach be useful for data collection?” was shown and the presenter opened the floor to discussion/feedback. We then proceeded to present detailed results from our suite of experiments, followed by a slide with the question “Do the results we summarized change anything about your original perceptions of class-based versus instance-based data collection?,” which prompted further discussion. The third author took notes on the discussion, and the entire discussion was audio recorded.

Results

The seminar was attended by 21 people. The majority (18) were from the Geography Department. Three members of the Biology Department also attended. The audience was a mix of faculty members, graduate students, and visiting researchers. Ten people returned questionnaires.

In response to our question about the extent to which participants agreed that the instance-based data collection approach is relevant and applicable to the practice of citizen science, the mean response was 6.0 on a seven-point scale (where seven was labeled “strongly agree”). Likewise, the mean response to our question about the extent to which participants agreed that the results of the experiments we presented were relevant and applicable to the practice of citizen science was also 6.0 on a seven-point scale. *These results provide a clear picture that participants viewed the instance-based approach as potentially valuable in the context of collecting citizen science data.*

Turning to the results of the open-ended questions we asked, Table A1 summarizes the responses we received to each question.

From our perspective, the feedback from the presentation reaffirmed our findings about the advantages and limitations of instance-based data collection. Specifically, participants saw the flexibility of the instance-based approach in accommodating unanticipated data, encouraging participation by people who are not familiar with the classes of interest to the researchers, and recognizing the potential to capture nonstandard forms of data. At the same time, concerns were expressed about the need for post-processing of data to make it useful for the goals of data consumers (scientists), and the likelihood that the collected data would be messy.

Table A1. Responses to Applicability Check Questions

Question	Summary of Responses	Examples Provided by Participants
Describe potential uses or applications of instance-based data collection	<ul style="list-style-type: none"> • When classes cannot be predetermined • When phenomena cannot reliably be classified • When participants knowledge (including traditional knowledge) is very different from that of researchers 	<ul style="list-style-type: none"> • In a project on toxic waste, land users noted changes in the color of flames generated by burning wood as evidence a site was toxic. Researchers had not anticipated this.
(How) Could you make use of instance-based data collection in your research?	<ul style="list-style-type: none"> • To obtain new insights • To collect (more) information about organisms difficult to identify visually • To integrate nontraditional sources of knowledge 	<ul style="list-style-type: none"> • A botanist was engaged in a project to classify plants, but was unable to in many cases, only realizing afterward that, because it was fall, the plants looked different than during summer. Identifying attributes would have been much easier.
Advantages of instance-based data collection	<ul style="list-style-type: none"> • Collect organism-specific data • Easy to use • Greater accuracy • Capture unanticipated data • Enhance reuse • Capture citizen scientists' categories • Capture subjective data • Allow nonexperts to participate 	<ul style="list-style-type: none"> • In a project in which fishers reported on ocean sponges, one participant indicated the IB approach would generate much richer data than "coral, other" or "sponge, unidentified" that frequently showed up in their data, due to difficulty in identifying sponge species
Limitations of instance-based data collection	<ul style="list-style-type: none"> • Difficulties post-processing data to a standard format • No common measures across instances • Complete dataset unlikely to be useful to an individual project • Hard to implement where classification/standardization is very important 	<ul style="list-style-type: none"> • In a project on caribou behavior, the interest is in a single class and additional instances might be considered noise
Advantages of class-based data collection	<ul style="list-style-type: none"> • Uniform data requires less post-processing and supports easy analysis • Useful with "informed" citizens who know what they are collecting • Useful for specific goals/questions • Tractable quantities of data • Data is uniform 	<ul style="list-style-type: none"> • Standardization/classification makes data more accessible and readable across the scientific community
Limitations of class-based data collection	<ul style="list-style-type: none"> • Categories must be known and well-defined • Low accuracy • More training required for amateurs • May miss relevant phenomena • Does not accommodate unanticipated data 	<ul style="list-style-type: none"> • In a project working with local residents, class-based data collection imposes the view or schema of the researchers on the contributors, thereby missing potential valuable perspectives of those closest to the phenomena of interest.

Appendix B

Expert and Machine Classification Study Details

Exploring Usefulness of Instance-Based Data: Classification by Experts and Machines

To answer the question whether additional processing can produce accurate classifications of attribute-based data at a level useful to data consumers, we conducted a study in which domain experts in biology were asked to infer classes based on attributes of observed organisms generated by citizen scientists. In addition, because domain expertise is a scarce resource, we also investigated the potential for classification using machine learning. We used a large dataset collected previously in a controlled laboratory setting.¹ In this previous study, we elicited free-form attributes from 390 nonexperts in biology (business students). The participants were shown images of plants and animals from the local region and were asked to describe those using attributes and classes.

To assess whether experts in natural history could accurately classify species based on attribute data, we conducted one-on-one interviews with local natural history experts having strong knowledge of the flora and fauna of the region (e.g., professors of biology at Memorial University of Newfoundland, members of the natural history society). We selected 16 organisms, a reasonable maximum for participants before risking expert fatigue (determined via a pretest). We designed the interview as a guessing game. For each organism, we had tallied the top 11 attributes provided by the nonexperts in the study in which the data were collected, along with the top most-frequent class (always a basic-level category, such as bird, tree, or fish) that had been provided by the research participants in that earlier experiment. The cutoff at 11 attributes did not affect our results, as all experts reached one of the stopping rules (see below) before the eleventh attribute.

After informed consent was obtained, the interviewer read out the instructions. To start, the interviewer presented the first sheet of the questionnaire to the natural history expert with the basic-level category and revealed only the first (most common) attribute. The expert was asked to provide a “best guess” of what the organism was. The expert was then asked to rank how well s/he could narrow down the list of possible species the organisms could be (as a measure of precision), as well as the confidence in the answer at the precision level given, based on the values given on the cards. For example, if the first attribute shown made the natural history expert think of a possible five species but s/he was unable to narrow it further, the precision was ranked “a.” If the expert had sufficient attributes that matched only one species (correct or incorrect) in their mental map, then precision was coded as “e.” Following the initial response, the interviewer revealed the second attribute. The experts were given the opportunity to revise their answer and were again asked to provide precision and confidence rankings. This procedure continued until one of the following stopping rules (not revealed to participants) was reached: (1) expert remained at precision level a, b, or c for three attributes; (2) expert reached precision level (d) or (e), even if incorrect; or (3) expert correctly identified the species. Once the stopping rule was achieved, the interviewer revealed the image and asked the natural history expert to identify it. This was followed by a short debriefing. Then the interviewer moved to the second organism on the questionnaire and the process continued for approximately one hour. After completing this process, participants were asked to complete a short exit questionnaire to collect biographical information. The entire interview was recorded on a digital recorder and later transcribed and a second researcher sat in on the interview (but did not speak) and took notes/recorded responses (see Figure B1 for the experimental setup).

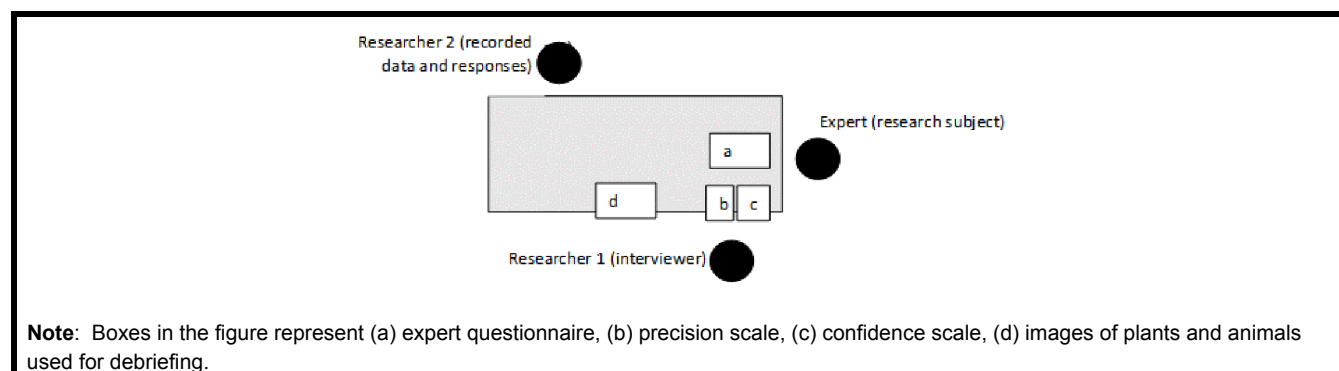


Figure B1. Set-Up for Expert Classification Experiment (Aerial View of Table and Participants)

¹Our study only considers a subset of the data obtained through this experiment; the full experiment is beyond the scope of the paper and is reported elsewhere (Lukyanenko et al. 2014).

The contents of each of the sheets shown to the expert are described below. (a) is a copy of the attribute list, the example below is for a single species—there was one sheet per species; the list was initially covered with an opaque card and the attributes revealed to the expert one at a time. (b) is a copy of the “precision scale”; the expert was asked to give a rank for their answer each time a new attribute was revealed. (c) is a copy of the “confidence scale”; the expert was asked to give a rank for their answer with each guess. (d) is an envelope of 8.5×11 ” color photographs of each species that was revealed to the expert when they reached one of the stopping rules (see text).

A. Experts Questionnaire

Sample Item [Common Tern, *Sterna hirundo*]

[Attributes sequentially revealed]:

White; Orange beak; Black head; Orange feet; Large wings; Long wings; Long tail; Orange legs; Black top of head; Pointy beak; Grey

B. Precision Scale

Based on the attributes given thus far, could you narrow down this species to one of a possible?

More than 10 species?

5–10 species?

3 or 4 species?

2 species?

1 species?

C. Confidence Scale

Based on the attributes given thus far, how confident are you that you can identify the species to the specificity level you indicated?

Use a scale from 1 to 5:

1 = Not at all confident

5 = Completely confident

D. Sample Photograph (Common tern—source: Wikimedia Commons)



We prepared questionnaires by printing a list of the 11 attributes for each organism on a single sheet of paper, as well as a sheet with precision rankings and one with confidence rankings. These were laid out on the interview table in full view of the expert participating (Figure B1).

The attributes were ordered based on the reported frequency by the nonexperts (most frequent first).² At the top of each page, the basic-level category for the organism was given. The order in which organisms (but not attributes) were presented was randomized across participants. We prepared 8.5x11" color photos of the same images that had been shown on screen to the nonexperts in the experiment that generated the attribute data used here, but kept these hidden from view initially (Figure B1). Each interview session lasted approximately one hour.

Responses were coded as correct (scored as 1) if they matched the common names or the genus/species. We assigned a score of 0.5 for answers containing the correct general name (e.g., "orchid" as final response for "Calypso orchid"). If the final response included two or three possible species, including the correct one, we also coded this as 0.5. Overall scores were tallied for each expert, including partial scores (thus a participant could score 9.5 out of 16 responses, for example).

The 16 natural history experts had a mean of 28 years (s.d. 20 years) of experience in natural history and 33.5 years (s.d. 17.5 years) living in the region. Self-identified areas of expertise varied, but included fish, mushrooms, birds, plants, and mammals, covering all the kinds of organisms used in the study. We further quantified expertise based on how well participants were able to identify the organism from the photograph (shown after the stopping rule was reached). When shown the photograph, our interview participants had a mean number of correct classifications of the item shown (at the genus/species-level) of 59.4% (s.d. 14.7). This is considerably higher than that generally achieved by nonexperts (e.g., in our lab experiment).³

Participants' ability to identify organisms varied based on the attributes (see Table B1). There was a high correlation between the confidence level reported with the final guess and the percentage of times the guess was correct (Spearman's $\rho = 0.68$, $p < 0.01$). There was also a high correlation between the precision reported with the final guess and the percentage of times the item was guessed correctly (Spearman's $\rho = 0.87$, $p < 0.001$). However, even for organisms for which experts had low to no correct classification, final precision was quite high, meaning that experts could come up with a limited list (usually less than five) of species that fit the set of attributes provided. While perfect species-level identification may not always be possible based on attributes provided by nonexperts (in taxonomy the diagnostic attributes to discriminate closely related species can be cryptic), a limited list (usually of similar species) can have utility for many ecological research questions, even if the true species-level identity is unknown. The results provide strong evidence of the utility of the instance-based approach for reducing the classification uncertainty from the basic level, which is the level at which nonexperts in general can accurately classify. Our set of species encompassed a range of taxonomic groups (plants, birds, mammals) and not all natural history experts are necessarily well-versed in all taxa.

To prepare the data for machine learning (ML), we converted the attribute data set into a matrix of attributes where the attributes provided in the study were assigned 1 if a particular participant used that attribute to describe the species of interest and 0 otherwise. The resulting matrix contained 119 columns and 1,839 rows with 5,129 (2.34%) of the attributes coded as "1" and 213,712 (97.66%) coded as "0." Each row represents attributes (with associated basic-level category) provided by one of the 125 nonexpert data contributors who were asked to describe the organisms.

As expected, the attribute data are sparse, making it potentially challenging for the machine learning algorithm to discover patterns and classify species correctly based on the attributes. The sparsity of the dataset is consistent with other research findings showing that crowds generate "noisy" data (Brynjolfsson et al. 2016; Sheng et al. 2008).

To ensure accessibility of our approach to data consumers, we applied a variety of common ML algorithms, which are available in popular ML software, including neural networks, support vector machines, random forests, boosting using decision trees and naïve Bayes algorithms (Provost and Fawcett 2013). In each case, the average classification accuracy was above 70%. The top performing algorithm (Table C1) was a boosted decision tree classifier, which achieved an average F-measure (a widely-used machine learning accuracy metric) of 0.76 (± 0.12 s.d.) across 16 species (based on 10 fold cross-validation and 50 boosting iterations).

A direct comparison between human and machine performance is not meaningful since ML worked with 16 finite targets (species), whereas experts had to draw from their knowledge of all possible organisms in a local area. The results, however, suggest that, while the immediate data obtained from the instance-based approach may have low precision, it can indeed be improved by human annotation and/or applying common, off-the-shelf ML techniques.

²Note that this procedure might inflate the apparent ability of experts to identify species relative to data collected in a field setting, as there would be no way to assess attribute frequency in the latter case. However, the goal in this study is simply to assess the feasibility of using attribute data for identification, and we determined this was a reasonable way to order the attributes as it reflected the likelihood such attributes would appear in practice.

³A less than perfect identification accuracy by experts demonstrates the limit of expertise in broad domains. High expertise is usually narrow in scope (e.g., a single subordinate category). For example, a person who owns a collie could be considered a "collie expert," but not expert in other dog breeds (Tanaka and Taylor 1991). Yet in practice, data consumers frequently have to make decisions on broader issues (e.g., all local birds).

Table B1. Classification Accuracy of Experts and F-Measure Obtained by Machines

Species	% Correct responses (humans) corrected for expertise based on ability to identify species in photo	Mean confidence on scale 1 (low) to 5 (high) by humans	Mean precision on scale 1 (low, can think of 5–10 species) to 5 (high, can think of 1 species) by humans	F-measure – harmonic mean (boosted tree)*
American Robin	65.60	3.69	3.34	0.88
Blue Jay	78.10	3.93	4.06	0.86
Blue Winged Teal	50.00	3.07	3.33	0.85
Calypso Orchid	71.40	2.87	3.07	0.64
Caribou	85.70	4.43	4.71	0.89
Caspian Tern	42.90	3.60	3.53	0.63
Common Tern	45.50	3.53	3.20	0.70
Fireweed	0.00	3.21	2.93	0.62
Greater Yellowlegs	100.00	3.80	3.80	0.84
Indian Pipe	10.00	3.15	2.62	0.82
Lung lichen	33.30	3.27	2.82	0.80
Mallard Duck	80.80	3.60	4.27	0.85
Moose	93.30	3.93	4.47	0.91
Old Man's Beard	100.00	3.50	3.57	0.68
Sheep laurel	0.00	3.31	2.38	0.50
Spotted sandpiper	100.00	3.64	2.86	0.71
Mean	59.79	3.53	3.44	0.76

*The results were obtained from a boosted decision tree classifier. The tree implemented a Chi-square automatic interaction detection algorithm (Geurts et al. 2006). The trees were boosted using an adaptive boosting (AdaBoost) classifier that combined the outputs of decision trees such that the trees were sequentially modified to reduce misclassification rates of the predecessor trees (Freund and Schapire 1997). The metrics were evaluated based on 10-fold cross-validation and 50 boosting iterations. AdaBoost algorithm is commonly available in such data mining software packages as RapidMiner (rapidminer.com) or SAS Enterprise Miner (sas.com).

Finally, we highlight several challenges of applying machine learning in open crowdsourcing.

First, for projects that are starting from scratch, there could be difficulties in obtaining a suitable training sample from which to build decision models. The key difficulty is that unlike applications of machine learning, where previous decisions by humans can be leveraged, in open crowdsourcing, data sets produced by crowds can be unique and equivalent data for training may be difficult or impossible to obtain. At the same time, several strategies can be used to generate a training set that approximates the actual crowdsourced data. For example, one could use a “verified” subset of the data created by experts. Another approach is to use uploaded photographs as a way to positively identify down to the desired level of precision, and then use these labels as the training set. Finally, one could also conduct a laboratory procedure as we did above by asking participants to observe sample instances and provide free-list attributes and classes, and then use of these for training. The three approaches can also be combined. There is ongoing research on learning from insufficient training data in computer science and related fields (e.g., Sommer and Paxson 2010; Webb et al. 2001; Wuest et al. 2016). Our paper provides an additional use case for these efforts, which, with more progress, can better support practical application of our ideas.

Second, we also note the difficulty in handling multiple target classes using machine learning. Typical machine learning activities involve predicting a small subset of classes (e.g., yes/no) (Provost and Fawcett 2013). However, some domains in crowdsourcing may contain hundreds or even thousands of classes that need to be learned and automatically predicted based on sparse crowdsourced UGC. Here, we anticipate that progress in machine learning algorithms will increase the ability to handle and predict large numbers of classes. We hope our paper motivates research in machine learning on improving ability to make a large number (i.e., tens, hundreds, or thousands) classification decisions at once (see Ou and Murphy 2007; Shalaginov and Franke 2016; Vincent and Hansen 2014).

We also note, however, that although a domain may contain a large number of classes, for a given task, data consumers are typically interested

in a small subset of these classes (e.g., instances of wildfires in California, earthquakes in Japan, specific invasive species in Queensland). As we show, off-the-shelf machine learning techniques are quite effective at handling few (in our case 16) classification categories. Furthermore, there are established strategies for zeroing in on a single class out of many in machine learning (see Bishop 2006).

Many of the challenges related to the use of machine learning in support of the contributor-centric IQ may not be severe for all projects and should become easier to address over time.

Appendix C

Additional Use Cases for Instance-Based Data Collection

We discuss additional use cases for instance-based data collection (beyond the citizen science context used throughout the paper).

Case 1: Urban Sensing in Smart Cities

As more people are living in urban areas, research shows that human well-being in cities is dependent on the quality of urban infrastructure and municipal services (Clarkson and Kirby 2016; Hartig and Kahn 2016). The development of “smart cities” helps urban planners, managers and decision makers collect a range of environmental and human-use data related to urban life (Cardone et al. 2013; Hivon and Titah 2017; Kalay 2017; Ramaswami et al. 2016).

Sensors such as traffic counters and air, noise, and water monitoring devices can be deployed to gather data in urban spaces. In addition, smart cities increasingly benefit from *human sensors*. Unlike automated sensors, human sensors have the capacity to interpret real-world events and act upon them. In addition to reporting on typical things and events, humans can make sense of unanticipated phenomena that would get coded as “errors” or “outliers” by most electronic sensors or not captured at all. However, to take advantage of this ability, more flexible data collection may be needed.

Potential data consumers in this case include municipal police, municipal public health and safety agencies, city planners, architects, planning consultants, local businesses and organizations, and citizens themselves.

Municipalities and urban agencies increasingly develop or subscribe to platforms that collect UGC from urban sensors. For example, a popular crowdsourcing project, CitySourced.com, asks ordinary citizens to provide reports based on predefined categories, as shown in Figure C1. However, this predefined schema may be inadequate to capture all phenomena of potential interest to municipalities. Consider the example of an overturned ammonia truck on a city road. Potentially affected residents might experience and report different phenomena, “I’m stuck in traffic but I don’t know why,” “I saw an overturned tanker truck,” “I was working in my garden and suddenly had difficulty breathing,” and “I saw clouds billowing from the other side of my backyard.” All of these may be manifestations of the incident. However, it would be virtually impossible to anticipate all such events in advance, or to incorporate them in the schema of an app such as CitySourced.com.



Figure C1. Categories of Urban Phenomena Provided by CitySourced.com (as of October 20, 2017)

Case 2: Disaster Management and Response

User-generated content is becoming a major source of information on natural and anthropogenic disasters. Disaster management and response applications are increasingly used by people in areas affected by natural and man-made disasters, as well as by volunteers outside the affected areas who wish to assist in recovery and restoration efforts (Goodchild 2007; Goodchild and Glennon 2010). Frequently, the observations made by ordinary people provide real-time input for disaster response and management agencies and result in more efficient and timely rescue and relief efforts (Johnson and Sieber 2012; Majchrzak and More 2011; Pultar et al. 2009).

Data consumers in this domain include governments at different levels, relief agencies, first responders, affected and interested local public and private organizations, citizens in the affected areas, as well as scientists, engineers, and planners.

In this context, timeliness, accuracy, and completeness of the report on an unfolding natural disaster are of the essence. It is, therefore, vital to ensure that the data collection processes facilitate, rather than impede, the ability of people to provide their observations. Many existing platforms designed to capture UGC related to disaster response rely on predefined categories. Consider a prominent project of this kind, the U.S. Federal Emergency Management Agency Disaster Reporter, which seeks “to crowdsource and share disaster-related information for events occurring within the United States” (<https://www.fema.gov/disaster-reporter>). Based on the predefined schema for the app (see Figure C2), we see that people may have difficulties reporting on anthropogenic events, as there are virtually no categories for human-made events (e.g., chemical explosion) provided. Similarly, every event requires an observer to have a positive classification, which may not be possible in all circumstances. For example, people may observe wilted plants and dried-up creeks, but may not necessarily conclude that drought is the cause, and thus may not post an otherwise valid observation. Finally, underscoring the difficulties in creating predefined classes, we note that there are multiple categories for the same event (e.g., “Home Fires” and “Floods” may be caused by “Wildfires” and “Tsunamis,” respectively). This may create difficulty for observers in using appropriate categories when reporting (as well as integrating related observations).



Figure C2. Categories of Disaster Events Provided by FEMA Reporter (as of October 20, 2017)

Case 3: Product Improvement and Customer Service Based on Customer Feedback

Part of the appeal of crowdsourced UGC is finding something unexpected and new. It has been long known that front-line employees, being in direct contact with day-to-day situations, are well-equipped at spotting unusual activities, manufacturing defects, or process failures (Tax and Brown 1998; Trevino and Victor 1992; Tucker and Edmondson 2003). In a UGC setting, a notable ability of contributors is that they can report individual experiences with objects of interest to data consumers. Increasingly, companies are taking advantage of UGC to seek customer feedback on consumer product impressions, malfunctions, general usage, and suggestions for improvement or ideas for future products (Abbasi et al. 2018; Ordenes et al. 2014; Stelzer et al. 2016; Voss et al. 2004). Data consumers in this domain are mainly within the organization and include product designers and engineers, marketers, production line managers, customer service specialists, business analysts, and top executives responsible for shaping company's strategy and new product development.

Customers should be able to provide feedback by communicating their experiences as seamlessly, accurately, and completely as they would like. Having observed several existing interfaces for collecting customer product feedback, we noticed that they employ predefined categories as primary units of data collection. Although relying on predefined categories can be helpful in focusing consumers on the aspects of the products or services of most interest to the organizational data consumers, and can enable auto-directing of feedback to the unit in the organization that can act on it promptly, we expect lower accuracy and completeness (including discovering something new or unexpected) to occur in this setting.

To illustrate, Apple's Customer Feedback website (<https://getsupport.apple.com>) offers a selection of predefined categories for getting hardware help with an iPhone, as shown in Figure C3. While these categories cover many of the common issues, a perusal of iPhone online communities and media reports reveals many more categories not that do not fit the current predefined schema, such as a cracked camera lens, debris under the screen, camera foam misalignment, missing camera lenses, and iPhone getting hot.⁴ Considering the multitude of potential interactions with an iPhone, it would be difficult, if not impossible, to predict and include every possible issue that customers might encounter. Naturally, the option to report another topic is also provided, but having the predefined choices may dissuade consumers from communicating something that does not fall into these categories.

⁴For example, <https://unlockgadget.com/blog/post/computer-repair/different-types-of-phone-damage-and-ways-to-repair-phones/36>.

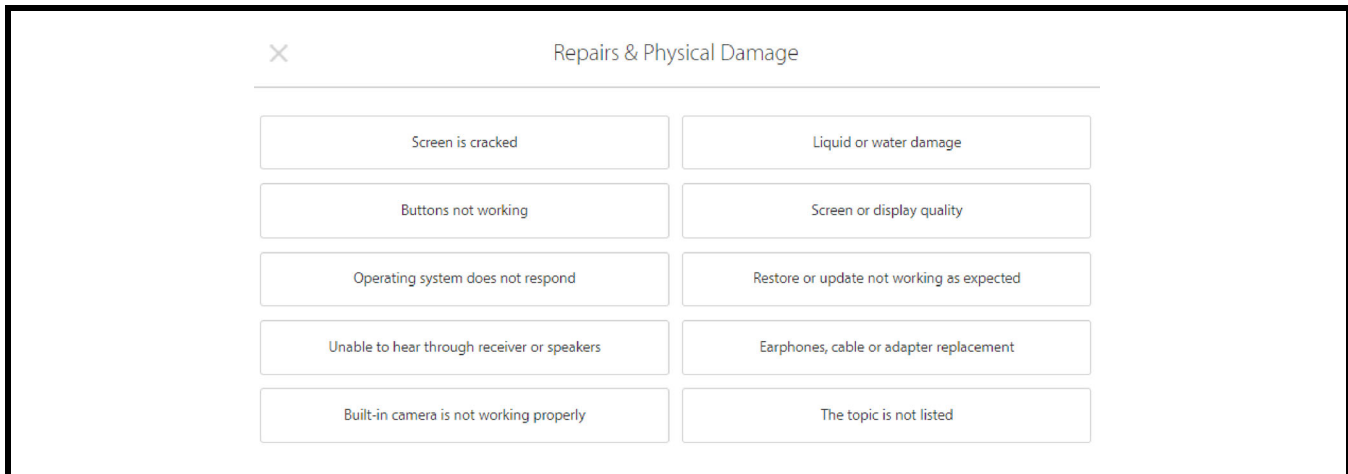


Figure C3. Categories of Hardware Issues by Apple Inc. (as of January 12, 2018)

References

- Abbasi, A., Zhou, Y., Deng, S., and Zhang, P. 2018. "Text Analytics to Support Sense-Making in Social Media: A Language-Action Perspective," *MIS Quarterly* (42:2), pp. 1-38.
- Bishop, C. 2006. *Pattern Recognition and Machine Learning*, New York : Springer-Verlag.
- Brynjolfsson, E., Geva, T., and Reichman, S. 2016. "Crowd-Squared: Amplifying the Predictive Power of Search Trend Data," *MIS Quarterly* (40:4), pp. 941-961.
- Cardone, G., Foschini, L., Bellavista, P., Corradi, A., Borcea, C., Talasila, M., and Curtmola, R. 2013. "Fostering Participation in Smart Cities: A Geo-Social Crowdsensing Platform," *IEEE Communications Magazine* (51:6), pp. 112-119.
- Clarkson, B. D., and Kirby, C. L. 2016. "Ecological Restoration in Urban Environments in New Zealand," *Ecological Management & Restoration* (17:3), pp. 180-190.
- Freund, Y., and Schapire, R. E. 1997. "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences* (55:1), pp. 119-139.
- Geurts, P., Ernst, D., and Wehenkel, L. 2006. "Extremely Randomized Trees," *Machine Learning* (63:1), pp. 3-42.
- Goodchild, M. 2007. "Citizens as Sensors: The World of Volunteered Geography," *GeoJournal* (69:4), pp. 211-221.
- Goodchild, M. F., and Glennon, J. A. 2010. "Crowdsourcing Geographic Information for Disaster Response: A Research Frontier," *International Journal of Digital Earth* (3:3), pp. 231-241.
- Hartig, T., and Kahn, P. H. 2016. "Living in Cities, Naturally," *Science* (352:6288), pp. 938-940.
- Hivon, J., and Titah, R. 2017. "Conceptualizing Citizen Participation in Open Data Use at the City Level," *Transforming Government: People, Process and Policy* (11:1), pp. 99-118.
- Johnson, P. A., and Sieber, R. E. 2012. "Situating the Adoption of VGI by Government," in *Crowdsourcing Geographic Knowledge*, D. Sui, S. Elwood, and M. Goodchild (eds.), Dordrecht, Netherlands: Springer, pp. 65-81.
- Kalay, Y. E. 2017. "How Smart Is the Smart City? Assessing the Impact of ICT on Cities," in *Proceedings of the Agent Based Modelling of Urban Systems: First International Workshop*, M.-R. Namazi-Rad, L. Padgham, P. Perez, K. Nagel, and A. Bazzan (eds.), Singapore, May 10, 2016, Springer International Publishing Switzerland, pp. 189-207.
- Lukyanenko, R., Parsons, J., and Wiersma, Y. 2014. "The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content," *Information Systems Research* (25:4), pp. 669-689.
- Majchrzak, A., and More, P. H. B. 2011. "Emergency! Web 2.0 to the Rescue!," *Communications of the ACM* (54:4), pp. 125-132.
- Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., and Zaki, M. 2014. "Analyzing Customer Experience Feedback Using Text Mining: A Linguistics-Based Approach," *Journal of Service Research* (17:3), pp. 278-295.
- Ou, G., and Murphey, Y. L. 2007. "Multi-Class Pattern Classification Using Neural Networks," *Pattern Recognition* (40:1), pp. 4-18.
- Provost, F., and Fawcett, T. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, Sebastopol, CA: O'Reilly Media, Inc.
- Pultar, E., Raubal, M., Cova, T. J., and Goodchild, M. F. 2009. "Dynamic GIS Case Studies: Wildfire Evacuation and Volunteered Geographic Information," *Transactions in GIS* (13:s1), pp. 85-104.

- Ramaswami, A., Russell, A. G., Culligan, P. J., Sharma, K. R., and Kumar, E. 2016. "Meta-Principles for Developing Smart, Sustainable, and Healthy Cities," *Science* (352:6288), pp. 940-943.
- Rosemann, M., and Vessey, I. 2008. "Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks," *MIS Quarterly* (32:1), pp. 1-22.
- Shalaginov, A., and Franke, K. 2016. "Towards Improvement of Multinomial Classification Accuracy of Neuro-Fuzzy for Digital Forensics Applications," in *Proceedings of the 16th International Conference on Hybrid Intelligent Systems*, A. Abraham, A. Haqiq, A. M. Alimi, G. Mezzour, N. Rokbani, and A. K. Muda (eds.), Springer International Publishing Switzerland, pp. 199-210.
- Sheng, V. S., Provost, F., and Ipeirotis, P. G. 2008. "Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers," in *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp. 614-622.
- Sommer, R., and Paxson, V. 2010. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in *Proceedings of the IEEE Symposium on Security and Privacy*, Los Alamitos, CA: IEEE Computer Society Press, pp. 305-316.
- Stelzer, A., Englert, F., Hörold, S., and Mayas, C. 2016. "Improving Service Quality in Public Transportation Systems Using Automated Customer Feedback," *Transportation Research Part E: Logistics and Transportation Review* (89), pp. 259-271.
- Tax, S. S., and Brown, S. W. 1998. "Recovering and Learning from Service Failure," *MIT Sloan Management Review* (40:1), pp. 75-88.
- Tanaka, J. W., and Taylor, M. 1991. "Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder?," *Cognitive Psychology* (23:3), pp. 457-482.
- Trevino, L. K., and Victor, B. 1992. "Peer Reporting of Unethical Behavior: A Social Context Perspective," *Academy of Management Journal* (35:1), pp. 38-64.
- Tucker, A. L., and Edmondson, A. C. 2003. "Why Hospitals Don't Learn from Failures: Organizational and Psychological Dynamics That Inhibit System Change," *California Management Review* (45:2), pp. 55-72.
- Vincent, M., and Hansen, N. R. 2014. "Sparse Group Lasso and High Dimensional Multinomial Classification," *Computational Statistics & Data Analysis* (71), pp. 771-786.
- Voss, C. A., Roth, A. V., Rosenzweig, E. D., Blackmon, K., and Chase, R. B. 2004. "A Tale of Two Countries' Conservatism, Service Quality, and Feedback on Customer Satisfaction," *Journal of Service Research* (6:3), pp. 212-230.
- Webb, G. I., Pazzani, M. J., and Billsus, D. 2001. "Machine Learning for User Modeling," *User Modeling and User-Adapted Interaction* (11:1-2), pp. 19-29.
- Wuest, T., Weimer, D., Irgens, C., and Thoben, K.-D. 2016. "Machine Learning in Manufacturing: Advantages, Challenges, and Applications," *Production & Manufacturing Research* (4:1), pp. 23-45.