Visitor Studies

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uvst20

Seeing the Forest, Not the Trees – Crowdsourced Data Collection Methods for Sector-Wide Research

John Voiklis, Kate Flinner, Shaun Field, Rupanwita Gupta, John Fraser, Joseph de la Torre Dwyer, Shelley Rank & Kathryn Nock

To cite this article: John Voiklis, Kate Flinner, Shaun Field, Rupanwita Gupta, John Fraser, Joseph de la Torre Dwyer, Shelley Rank & Kathryn Nock (2023): Seeing the Forest, Not the Trees – Crowdsourced Data Collection Methods for Sector-Wide Research, Visitor Studies, DOI: 10.1080/10645578.2023.2167404

To link to this article: https://doi.org/10.1080/10645578.2023.2167404

Published online: 01 Feb 2023.

VISITOR STUDIES
https://doi.org/10.1080/10645578.2023.2167404

Routledge
Taylor & Francis Group

Seeing the Forest, Not the Trees – Crowdsourced Data Collection Methods for Sector-Wide Research

John Voiklis[a], Kate Flinner[b], Shaun Field[a], Rupanwita Gupta[b], John Fraser[a], Joseph de la Torre Dwyer[a], Shelley Rank[c] and Kathryn Nock[a]

[a]Knology, New York City, USA; [b]Independent Researcher; [c]Wildlife Conservation Society, New York City, USA

ABSTRACT

Research that involves a large and broad sample of museums can produce a representative picture of the entire museum sector and lead to global insights that may not be attainable through a more local lens. However, many museum research projects use a small sample of museums, meant to represent the entire field. We propose a research method that distributes data collection across a broad swath of museums to provide local detail that can be used to assemble a collective picture on a topic of interest to the field. This method, called *crowdsourced data collection*, was used in a yearlong study of zoos and aquariums in North America, in which 95 institutions were asked to collect data for one to two survey modules per month. We hoped this approach would produce data comparable to data gathered with conventional methods and reduce burden on participating institutions. We found the method replicated nationally representative studies with two validated scales. While only one third of the institutions completed all modules, institutions typically did 8-9 modules, with only slight decreases in the probability of completing the study over time. These results suggest researchers can use crowdsourced data collection to reliably study the museum sector. We also discuss the challenges of this method for researchers and institutions participating as data collection sites.

ARTICLE HISTORY
Received 17 February 2021
Revised 30 August 2021
Accepted 6 January 2023

KEYWORDS
Attrition; crowdsource; data collection; methods; replication; sector study

Introduction

In 2018 and 2019, our team of researchers partnered with a professional association in the museum sector to study how a crowdsourced data collection method could engage a large portion of institutions in research, to collectively produce an aggregated and accurate data set. The sector—in this case, the zoo and aquarium field in the United States, Mexico, and Canada—had not undertaken a collaborative and coordinated data collection initiative in the past. The study was originally designed to answer questions about public perceptions and understanding of the role of zoos and aquariums in STEM (science, technology, engineering, and mathematics) learning, as part

CONTACT John Voiklis JohnV@knology.org Knology, New York City, USA.

of the Why Zoos and Aquariums Matter research initiative. During the study, the research team also realized that documenting and presenting this methodological approach could advance applied social science research in the museum sector, as well as other sectors. This paper provides a description of the literature, the crowdsourcing data collection methods, and two types of analysis to test the value of this method.

Crowdsourcing was coined by a contributing editor at Wired magazine (Howe, 2006) to describe a then novel variation on the practice of outsourcing products and services, where institutions would split complex tasks into many simple tasks each of which was completed by many independent contractors and then reassembled into a single, optimal product or service. While the coinage and widespread use of crowdsourcing are recent, a precedent for using crowdsourced data collection dates, at least, to 1890 and the establishment of the National Weather Service Cooperative Observer Program (COOP). The goal of COOP is to assemble a unified dataset needed to study the climate of the United States, as a whole (National Weather Service, (n.d.).

For the purposes of this study, we define crowdsourced data collection as an approach in which multiple institutions or sites collect new data from audiences or other stakeholders using standardized measures, instruments, and protocols. This method produces a unified dataset that enables researchers to study a sector, as a whole. This sector-wide objective is what makes the crowdsourcing approach distinct from collaborative multi-site studies and research designed to develop and share research tools to multiple sites. With the latter two approaches, the goal is often comparative case studies about a particular topic, where collecting data from individual sites has primacy over collecting data from across the sector. This primacy of the individual site remains true even in the cases where the ultimate goal is a meta-analysis of results across multiple sites. Put simply, crowdsourced data collection is a collaboration among multiple sites from across a sector to collect data about the sector, as a whole.

Literature review

There is substantial precedent for crowdsourcing data in behavioral research. Social scientists have long used Amazon's Mechanical Turk to gather data from individuals, and have documented the benefits, limitations, and need for more research on these methods. Zhou and Fishbach (2016) describe a problem with crowdsourced data collection that is specifically relevant to the current study: attrition and its role in skewing the data. In spite of the many concerns surrounding crowdsourced data, studies have validated the method through replicating classic behavioral experiments that were originally conducted in controlled laboratory settings (e.g., Gureckis et al., 2016; Crump et al., 2013). These results at the individual level suggest that crowdsourced institution-level data could yield accurate results as well.

While social science research has focused on crowdsourcing individual-level data, there has been little research on crowdsourcing data with organizational membership entities to gain sector-level results, particularly in the museum field. As a result of these apparent omissions in the research, the literature to inform sector-level data crowdsourcing appears to be limited. To understand the literature currently available,

we reviewed the archive of *Visitor Studies* and grey literature in InformalScience.org. For *Visitor Studies*, we reviewed article titles and abstracts appearing from 2007 to 2020, and could not find many articles that addressed sector-level data collection efforts. We did find a study that used data collected at two natural history museums (Kisiel & Ancelet, 2009). A more recent study—which was also part of the authors' WZAM initiative—used case studies at multiple institutions in the US (Gupta et al., 2019). On InformalScience.org, we used search terms like "collaborative data collection" and "sector research in museums." These searches identified collaborative institutional projects where multiple institutions come together for a grant, or researchers looked at data from many institutions. For example, they brought up Philadelphia Informal Science Education Collaborative (PISEC), research partnerships between museums and community centers (e.g., Collaborative Research: Tinkering and Making Strategies to Engage Children and Families in Creating with Code), and syntheses of research on museums (e.g., Impact Evaluation of Museums, Archives, and Libraries: Available Evidence Project). Across these sources, we could not find literature specifically about methods for sector-level research and data collection. Notably, the Association of Science and Technology Centers is already moving in this direction with the Collaboration for Ongoing Visitor Experiences Studies (COVES) initiative, which uses shared evaluation tools and aggregates data across institutions.

We looked for examples of these data collection methods in fields similar to museums and found a promising study of informal STEM learning at science festivals, in which 25 science festivals collected data from 30,000 participants (Peterman & Gathings, 2019). While our literature review focused primarily on *Visitor Studies* and InformalScience.org, the lack of studies involving crowdsourced data collection for research suggested this approach has been limited in the museum field. Nevertheless, similar studies in adjacent fields indicate that there may be studies that this review failed to uncover.

In spite of the potential gap in research precedent, professional associations in museums and potentially other nonprofit sectors appear to have the procedural capability to play a role in sector-level crowdsourced data collection for the purposes of research. Many professional associations already poll their institutional members—that is, a leader or administrator representing an institution or organization provides information on the organization's behalf. Associations poll organizations on a regular basis to inform collective understanding of sectors or sub-sectors, often with a focus on comparable data like income, fundraising and compensation. For example, the American Alliance of Museums surveys the sector on salary and compensation. The Association of Children's Museums collects a wide range of institutional data through ACM Trends, an ongoing series of research publications (Flinner et al., 2020). These association-level studies in the nonprofit sector appear to focus on information that organizations already have in hand—such as financial outcomes, consumer use and engagement, and other metrics important to the sector—but they rarely ask organizations to collect novel data for the purpose of research involving the public. For example, the *Impact Evaluation of Museums, Archives and Libraries: Available Evidence Project* examined reports to identify the social, educational, and economic outcomes of institutions from across the United Kingdom (Wavell et al., 2002). This study did

not include research on the public, such as attitudes, understanding, and values related to institutions.

While there seems to be relatively few examples of sector-level research that use crowdsourced data for research, professionals, organizations, the museum sector, and other sectors stand to benefit from participating in research at this scale. Both researchers and professionals have long called for integration of research into applied work, but have consistently documented barriers across sectors (e.g., Guthrie et al., 2011; Helmsley-Brown & Sharp, 2003). For decades, social science researchers have responded to this challenge by using and refining collaborative strategies like Participatory Action Research and Action Research, but data collection still tends to focus on individuals or sampling from a small number of organizations as proxies for the larger group. More research is needed to understand whether data collection at the sector level is effective, and if so, how it can benefit integrated research approaches and professional practice. If data can be successfully collected and aggregated across organizations, we anticipate that the risk of errors will be lowered and researchers will be better prepared to capture an accurate view of a sector. This work may benefit nonprofit sectors, particularly those focusing on social good outcomes, by more closely demonstrating impact in a variety of contexts.

Study overview: STEM communications surrounding zoos & aquariums

The overall study aimed to deepen our understanding of how zoos and aquariums advance STEM learning in North America. We asked: What are the public's perceptions of zoos and aquariums as part of non-formal and informal STEM learning ecology?, and What relative authority does the public confer on zoos and aquariums about STEM topics outside the zoo and aquarium experience? The research focused on the many ways that these institutions participate in public life—from print and online communications, through conversations and presentations at their facilities, to informal discussions personnel have with family and friends—and how those interactions shape public perceptions of zoos and aquariums.

To study these issues, we devised a crowdsourcing approach to data collection, in order to assemble a representative data set of zoos and aquariums across the United States, Mexico, and Canada. We partnered with the Association of Zoos & Aquariums (AZA), an association that accredits and acts as the collective representative for the major zoos and aquariums in the sector in the United States and other countries, to promote and support the study. Next, we designed the study with the goal of collecting a few responses from each institution, and, in aggregate, to build a substantial database that captured the perspectives of professionals, volunteers, and visitors. Our aim was to capture sufficient data to describe the heterogeneity of the sector, and the regional variations across North America.

To produce a rich dataset that minimized the burden on individual institutions, we split the survey into modules that were dispatched over time, so that institutions facilitated only a small survey study each month over the course of one year. Each module was estimated to involve a minimal amount of staff time each month, and no more than ten minutes for any participant to complete their responses.

Simultaneously, the research team hoped that institutional participation in this study could serve as a capacity building exercise for staff conducting research related to their specific programming to understand personnel and visitors' experiences. To our knowledge, this kind of crowdsourcing effort for data collection has only happened at a small scale in the zoo and aquarium sector in the US (Fraser et al., 2009) or for single-project interventions (e.g. Geiger et al., 2017a, 2017b, 2019; Moss et al., 2017).

The research team had several hypotheses related to the crowdsourced data collection method:

Hypothesis 1: Given that crowdsourcing spreads burdens across participants, we anticipated that the likelihood of completing the yearlong study would remain within the range of completion/attrition rates reported by national survey research organizations (e.g., Pew Research Center).

Hypothesis 2: Given that previous studies have used crowdsourcing methods to replicate laboratory studies, we anticipated that we could replicate nationally representative survey studies with a crowdsourcing method focused on sector-wide data.

Research design

The research team conducted the study from June 2018 to July 2019. Recruitment began in June 2018. The AZA sent an email to their membership base to share information about the study. The email contained a link to a questionnaire, where institutions could indicate their interest, provide contact information, and describe whether or not their staff had been certified in Protections of Human Subjects Research. Staff and volunteers who were not currently certified were invited to a free webinar training later that month to begin the certification process (and later complete the training on their own). The researchers also wanted to ensure the sampling strategy represented a range of regions, types of institution (i.e., zoos, aquariums, bioparks), and budget sizes. To accomplish this goal, additional institutions were recruited via personal email from the research team to senior leadership contacts in the sector, as well as in person at the AZA annual conference held in Seattle, Washington in September 2018. As a result of this recruitment approach, the participating institutions were all AZA-accredited, a valued status in the zoo and aquarium sector. They also represented the field's variation in scale and location; the Representation & Participation section below discusses this in further detail.

Starting in July 2018, the research team emailed the designated institutional representatives each month with a new part of the survey, called a "Quick Facts Ask." Monthly emails included both the new survey module information, and a quick summary of the initial results from the previous month. The kinds of information requested in the surveys were data on historical admission prices, STEM learning experiences and opportunities during a zoo or aquarium visit, how STEM topics are incorporated into communications with the public, ways people think about wildlife after visiting a zoo or aquarium, and more. Survey modules were programmed into an online survey platform (Qualtrics).

The institutional representative was responsible for distributing each month's survey module to the intended participants, or those who then collected the data from others.

Institutions were responsible for collecting one to ten responses for each module, depending on the nature of the questions. For instance, modules that asked about institutional partnerships required one response to represent the organization. For modules that studied visitor perspectives, institutions collected five to ten responses. Depending on the type of participant intended for each module, the survey was taken in different formats. Staff and volunteers filled out their survey modules on office computers or personal devices. Visitors filled out their modules on tablets that were distributed by staff or volunteers who had completed their human subjects research training. Researchers also made print versions of the survey modules available to institutions that did not have access to tablets for their visitors; for these print surveys, the researchers also did the data entry. For the types of survey responses we report here—closed-ended, multiple-choice selections and rating scales—differences between print and on-screen surveys tend to be negligible (for review see Noyes & Garland, 2008; also Crump et al. 2013). Further discussion of differences between print and on-screen data collection are beyond the scope of this present comparison of crowdsourced data collection to nationally representative online surveys.

Institutions could complete a module at a later date if they could not complete data collection targets in one month or if they entered the study at a later date. All data collection finished by July 31, 2019. We offered each institution $100 USD, disbursed in two installments, so that institutions could purchase thank-you gifts for the survey modules intended for visitors. Institutions gave gifts after visitors completed their survey modules. Some participating institutions did not accept the participant support money.

All research protocols involving human subjects were reviewed and approved by Solutions Institutional Review Board (IRB00008523, IRB Type: OHRP/FDA, Protocol #2018/05/12).

The research team developed a training for representatives from participating institutions to become certified in human subjects research, which included all research protocols specific to this study. The study required each participating institution's representative(s) to complete the training and sign a consent form, called a Certification of Collaborator form. Data collectors provided research participants with information about their participation before they initiated the survey modules; at that point, they were given the option not to go forward.

We conducted two analyses to understand the effectiveness of this crowdsourced data collection approach: 1) Testing study attrition, and 2) Testing reliability and replicability of crowdsourced data.

Representation & participation

Overall, 95 of the AZA's 240 accredited institutions participated in the study, representing approximately 40% of the total membership of the AZA at that time (see Table 1 for the number of participating institutions in each AZA budget size category). These institutions were located in 36 states in the United States, 1 state in Mexico, and 1 province in Canada. There were at least five organizations in each region of the US, as defined by the Bureau of Economic Analysis's delineation of regions (BEA, 2020). Participating institutions represented a variety of organizational sizes. As a proxy for

Table 1. Number of participating institutions in each AZA budget size category (AZA, 2018).

	Participating institutions	Total institutions in category (2018)
$200,000 - $1,999,999	20	48
$2,000,000 - $6,999,999	24	67
$7,000,000 - $25,999,999	37	69
$26,000,000 and Larger	11	35

Table 2. Schedule and participation rates for each module.

Name of ask / module	Month of ask	Type of ask	Number of organizations who received ask*	Number of organizations that participated in ask	Percent participation	Number of responses collected
Historical admissions prices	Aug-18	Personnel	95	89	94%	89
STEM learning experiences – visitor perspectives^	Sep-18	Visitors	94	N/A	–	697
STEM throughout the day – personnel perspectives	Oct-18	Personnel	90	79	88%	806
Institutional collaborations	Nov-18	Personnel	93	75	81%	75
Communications – social media	Dec-18	Personnel	93	67	72%	67
Communications – PR/ marketing	Dec-18	Personnel	93	73	78%	73
STEM learning opportunities	Jan-19	Personnel	93	80	86%	80
Conservation perceptions	Feb-19	Personnel & visitors	93	80	86%	682
My Z/A compared to other ISEIs	Mar-19	Visitors	93	58	62%	319
Everyday conversations	Apr-19	Personnel	92	75	82%	597
My Z/A compared to other Z/As	May-19	Visitors	91	53	58%	397
Wildlife values orientations	Jun-19	Visitors	84	47	56%	358
STEM learning ecology	Jul-19	Visitors	89	52	58%	755

size, we used the Association of Zoo and Aquarium's 2018 categorization of annual operating budgets, broken down into four groups according to USD (AZA, 2018). In our sample, institutions represented every budget size category, though fewer institutions were a part of the highest category (Table 2).

All data (without identifying information, such as demographic responses) for all WZAM studies, including those reported here, are available for download (https://bit.ly/2EgNLNN). In addition, Field, de la Torre Dwyer, Gupta et al. (2019; https://bit.ly/2EgNLNN) provides a summary of results from all crowdsourced modules that the WZAM team reported to the National Science Foundation and AZA. All 95 institutions participated in at least one module. On average, institutions participated in 8 or 9 of the 13 modules developed for this study.

Study 1: testing study attrition

The primary aim of Study 1 was to estimate the probability of continuing participation for institutions in the study and compare that estimate to hypothetical values obtained from prominent survey research organizations, such as the Pew Research Center (Hypothesis 1).

Participants

Over the 12-month study period, some participating institutions dropped out of the study. Approximately one third (31) of the institutions completed all 13 modules of the study. The primary reasons that institutions left the study early were: lack of staffing support to coordinate data collection, turnover of staff members who were the data collection contacts, and management electing to skip or omit survey modules that they perceived as logistically challenging (e.g., the organization was already conducting its own survey of a similar topic). A power analysis indicated that the number of participating institutions ($N = 95$) that started the study exceeded the minimum ($N = 53$) to detect the observed effects (Yung & Liu, 2020).

Analysis

We performed a survival analysis, which predicts the probability of continued participation at each of the twelve time-steps (months), given several factors: Type of institution, Region (as defined by the US Bureau of Economic Analysis, [2020]), Annual Budget, Number of Full-Time Employees, Number of Part-Time Employees, and Number of Volunteers. In order to test Hypothesis 1, we compared the probability of continued participation in this study to ten simulated "survival" data sets.

We simulated the data sets based on the one-time and longitudinal attrition rates–1%-30%–observed by national polling and research organizations, such as the Pew Research Center (2019a, November; 2019b, December; 2020, April), RAND Corporation (Pollard & Baird, 2017), and the General Social Survey (Smith & Son, 2010). While these organizations generally experience attrition rates smaller than 15%, we included the higher rates observed by Zhou and Fishbach (2016) for single-session crowdsourced studies. We used a statistically simple procedure to create the 10 simulated data sets (Nielsen, 2019), one for each of 10 levels of attrition from 1% to 5% (in increments of 1%) and from 10% to 30% (in increments of 5%). For each data set, we assumed 100 notional institutions starting as participants in a 12-month notional study. For each month, we used the attrition rate (e.g., 1%) as the probability (e.g., $p = 0.01$) that any of the participating institutions would drop out of the study and, then, removed the dropouts from the participant count before proceeding to the next month. For each data set, we repeated this procedure 1,000 times and, then, calculated the average survival of each notional institution.

Results

Overall, the probability of continued participation decreased slowly over the 12-month period, reaching a potential minimum range of 76%–95% chance of continued participation at month 12 (Figure 1). Continued participation did not depend on any of the factors mentioned above (all $p \geq 0.34$). Institutions that ended the data collection early appear to have dropped out for their own particular reasons and not because of any systematic constraints related to type, region, budget or staffing (The X axis shows the value of responses on the so-called standard or z scale. The raw scores on different

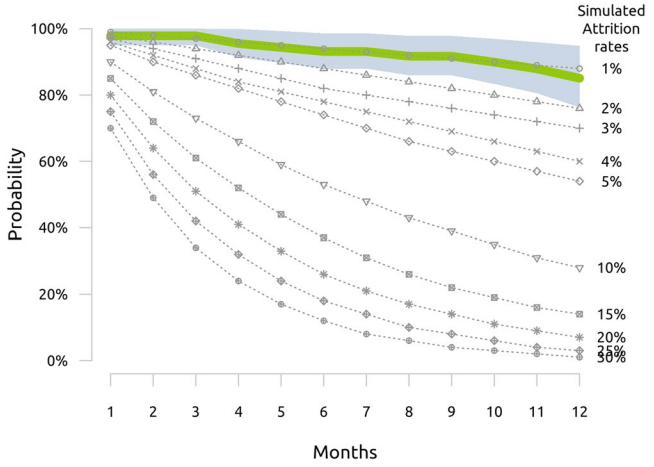

Figure 1. Probability of continued participation by institutions in 12-month crowdsourced data collection study. Blue shading indicates the 95% confidence interval for the probability at each time point (survival). Gray lines show survival based on simulated data with rates of attrition found in the literature on survey panels.

scales are transformed into numeric scores on the same standard scale with a mean of 0 and a standard deviation of 1 (Weisstein, n.d.). The Y axis shows the probability of a response at each level on the standard scale and data for all WZAM studies are available for download at https://bit.ly/2EgNLNN). As stated above, participating institutions completed about three quarters of the modules, on average.

As apparent in Figure 1, the empirical survival curve based on the participating zoos and aquariums (green line), tracked the simulated survival curve for a compounding attrition rate of 1%. The lower bound of the 95% confidence interval for the survival probability at each time point (blue shading) overlapped with the simulated survival curve for a compounding attrition rate of 2%. All other survival curves fell outside the 95% confidence interval for the empirical survival curve. This means that our effort at sector-wide crowdsourcing of data performed better at maintaining longitudinal participation with a compounding attrition rate of 3%. While the cumulative loss of participating zoos and aquariums was more than double the cumulative loss one would expect from a 3% compounding attrition rate (NA% vs. 30%), this difference is mainly attributable to the attrition on months 7, 8, 10, 11, which were the only months where attrition significantly exceeded 3% (6%–15%; all Exact Binomial Tests, $p < .05$).

The best way to test Hypothesis 1 would be an experiment where we manipulated the conditions for continued participation. None of the factors identified in the present study—budget, staffing, regions, and others regularly measured by sector leaders and associations—accounted for continued participation. So an experiment would require that we first identify a factor that varies across the sector (but is outside the typical metrics used by the sector) and can serve as either a quasi-experimental condition or an experimentally manipulated condition. Until then, the present results offer preliminary evidence in support for Hypothesis 1: our sector-wide crowdsourcing approach

sustained participation roughly equivalent to that reported by reputable and reliable national survey organizations.

Study 2: testing reliability & replicability of crowdsourced data

The survey modules included two validated scales that enabled the research team to compare the crowdsourced data to that of other studies using the same scales (Hypothesis 2). Study 2 A involved the Wildlife Value Orientation scales and Study 2B involved the STEM Learning Ecology scale.

Study 2A

Methods

As part of our exploration of the public's view of the role of zoos and aquariums, we used the Wildlife Value Orientation scales to assess how visitors think about wildlife after visiting a zoo or aquarium. In general, the Wildlife scales are used to segment respondents into Wildlife values groups, and those groupings have been shown to predict attitudes about wildlife policy and beliefs about non-human animals (Manfredo et al., 2018). This module used the version of the Wildlife Value Orientation scales available in Manfredo et al. (2018). It includes two sub-scales. One subscale assesses the dominance of wildlife orientation, including items on the "Appropriate Use" of wildlife and subscale on hunting. The second subscale assesses the mutualism with wildlife orientation, including items on social affiliation with wildlife and caring for wildlife. The first step in using scales to segment respondents is to check whether they are self-consistent (reliable) in how they respond to conceptually related items in the scale (Kaufman & Rousseeuw, 1990). To extract consistent groupings of respondents across different datasets requires that each dataset exhibit similar levels of individual reliability. To assess the validity of the crowdsourced data, we compared the subscale reliabilities against those obtained by Manfredo et al. (2018).

Participants

Overall, 358 visitors to 47 zoos and aquariums attempted the Wildlife Value Orientation (WVO) survey module; 330 participants completed at least 80% of the survey. We did not collect demographic information about participants. Analyses are based on these 330 participants, representing 46 institutions. As a rule of thumb, the reliability analysis we report below requires at least 10 participants per the number of items in each subscale of the WVO—Appropriate Use (6 items), Hunting (4 items), Social Affiliation (4 items), Caring (5 items). 330 participants far exceeds that minimum number of participants.

Analysis & results

To compare the subscale reliabilities for the crowdsourced data collection study against those obtained by Manfredo et al. (2018), we calculated the 95% confidence interval for Cronbach's α statistic on each subscale, by repeatedly resampling subscale item

responses (commonly known as boot-strapping). We then observed whether the Cronbach's α statistics from Manfredo et al. (2018) fell within the confidence interval (Table 3). This result would mean that the reliability of Manfredo et al. (2018) did not differ statistically from the reliability range for the crowdsourced data.

As apparent in Table 2, the crowdsourced data yielded subscale reliabilities comparable to Manfredo et al. (2018); the confidence intervals subsumed the value of Cronbach's α on three of the four subscales. While the reliability of Appropriate use beliefs in Manfredo et al. (2018) exceeded the upper limit of Cronbach's α for the crowdsourced data, the difference was small and the mean value of Cronbach's α for the crowdsourced data was acceptable (based on a rule-of-thumb threshold for Cronbach's $\alpha \geq 0.7$).

Study 2B

Methods

The STEM Learning Ecology survey module explored how visitors perceive zoos and aquariums within the STEM learning ecology—the constellation of settings and modes in which people engage with STEM learning throughout their lives. The module asked about the modes in which people engage in STEM learning at Z/As, and the STEM topics that they engage with in these experiences. This module was designed based on the results of the national survey study (Gupta et al., 2020). In order to reduce the cognitive burden on visitors responding to the survey, the crowdsourced survey queried a fixed set of six settings—Science Center, Natural History Museum, Botanical Garden, Back/Front Yard, as well as Aquarium and Zoo—instead of using the 23 settings used in the national survey. Otherwise, the module asked about all four STEM disciplines, 10 modes of engagement, and all fourteen topics queried in the national survey. The modes of engagement were: Conversations, Digital Media, Exploring Plants, Hands On, Learning Facts, Observing Animals, Presentations, Public Program, Reading Signs, and Using Senses. The 14 topics were: Animal Behavior, Architectural Design, Climate Change, Conservation, Construction, Ecosystems, Food Nutrition, Geography, Medicine, Reproduction, Species Names, Statistics, Sustainability, and Water Quality. To assess whether crowdsourced data replicates national results, we compared the distributions of responses to the crowdsourced survey and national survey.

Participants

Overall, 755 visitors to 52 zoos and aquariums attempted and completed the STEM Learning Ecology survey module.

Table 3. Comparisons of subscale reliabilities (95% confidence interval for Cronbach's α statistic) from crowdsourced data to reliabilities observed in Manfredo et al. (2018).

	Manfredo et al. (2018)	Observed lower	Observed mean	Observed upper
Appropriate use beliefs	0.78	0.66	0.72	0.76
Hunting beliefs	0.80	0.41	0.60	0.81
Social affiliation beliefs	0.82	0.75	0.81	0.85
Caring beliefs	0.80	0.80	0.84	0.86

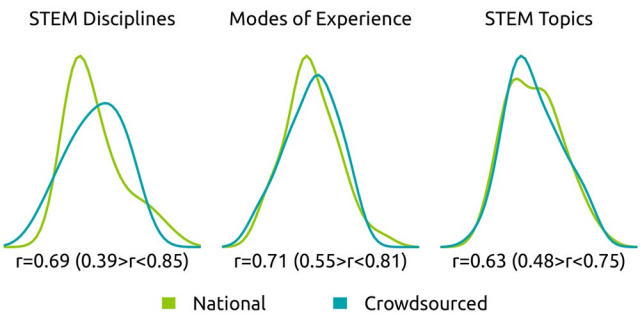

Figure 2. Comparing the (scaled) probability distributions from national survey data and crowdsourced data on encountering STEM disciplines (left), modes of experience (center), and encountering STEM topics (right) in informal STEM learning environments. Each plot includes the pointwise estimate of the correlation coefficient and its 95% confidence interval.

Analyses & results

To compare the distributions of responses to the crowdsourced survey and national survey, we calculated and tested the correlation coefficients for each of the three sets of summary responses: 1) The mean frequencies with which participants reported encountering the four STEM disciplines at each of the six setting listed above (24 STEM × Setting values), 2) the mean frequencies with which they reported using each of the 10 modes of engagement (60 Mode × Setting values), and 3) the mean frequencies with which they reported learning about each of the 14 STEM topics (84 Topic × Setting values). Figure 2 shows the probability distributions of responses on a common scale[1] for easy comparison. One can interpret the graphs by comparing the shapes of the lines: do they overlap, do they stretch across a similar range, do the peaks appear near the same location?

As apparent from the shape of the curves in Figure 2, the data collected through crowdsourcing yield distributions that were highly similar to those collected through a national panel survey (all r>.6). To test whether the correlations were reliable (i.e., differed significantly from zero), we calculated the 95% confidence interval of the correlation coefficient (r) for each set of mean frequency values, by repeatedly resampling subsets of values. We then observed whether $r = 0$ fell outside of the confidence interval, which would mean that the sets of frequencies reliably followed correlated distribution patterns. In all cases the confidence intervals excluded $r = 0$, we calculated the 95% confidence interval of the correlation coefficient (r) for each set of mean frequency values, by repeatedly resampling subsets of values. We then observed whether $r = 0$ fell outside of the confidence interval, which would mean that the sets of frequencies reliably followed correlated distribution patterns. In all cases the confidence intervals excluded $r = 0$: for STEM disciplines, .39>r<.85; for modes of experience, .55>r<.81; for STEM topics, .48>r<.75. The power of these tests to detect correlations of this size, given the number of values compared, was near certain probability, ranging from .97 to .99 (cf., Cohen, 1988).

Discussion

At first glance, it seems that Hypothesis 1 (that a large proportion of institutions would be likely to complete the yearlong study) was not supported by the data, as

only about one third of institutions completed the full study. However, there is a silver lining. First, institutions were typically able to complete 8 or 9 of the 12 modules. Second, the longer that institutions participated over the course of the year-long study, there was only a slight decrease in the probability of continued participation. Based on this information, we believe that the study reduced the burden on institutional participants. It is unlikely that we would have been able to collect this magnitude of data with a single research team traveling to multiple institutions, or with a smaller number of institutions collecting a larger proportion of data.

The data supports Hypothesis 2 (that we could replicate nationally representative survey studies with a crowdsourcing method focused on sector-wide data), as two tests on validated scales showed our data was comparable to data from national survey studies that focused on individual participants. This finding demonstrates that crowdsourced data collection can produce accurate and reliable data sets.

Opportunities

Our results show that crowdsourcing data collection through institutional members of an association can produce a data set that captures an accurate picture of a sector with reliable and replicable data. In this case, we gained a rich understanding of the zoo and aquarium sector, which, to our knowledge, had not previously set out to do crowdsourced data collection at this scale.

Our experience of conducting this study has indicated there are opportunities for researchers, professional associations, and advocacy organizations. While our project ran for 12 months, we recognize that many studies could be done on a much smaller scale. Shorter studies using this approach will likely be able to attain high response rates and reliable data, since institutions in our study had high rates of participation in the first few months.

We also believe that a crowdsourced data collection approach could be successful in a variety of sectors, including those that are highly centralized around a governing body as well as those that are more decentralized. In particular, this approach might work well in the following sectors: out of school learning, social services, arts, and community services or services.

Professional associations in any sector can be an important component in this research approach. Partnering with a professional association behooved our research because it elevated the profile of the study. We believe this partnership also increased the degree of trust in our project team and signaled the importance of participating in the project as a way of gathering information that could benefit participating institutions.

Another important advantage of this method is the opportunity to involve a wide range of institutions. This is especially important when it comes to the size of an institution. Our study included institutions with annual operating budgets ranging from $200,000 to over $26 million, which indicates that participation was feasible for institutions of every size in this sector. With traditional research methods and administrative costs, large-scale research collaborations are typically out of reach for small institutions that do not have dedicated research or grant writing staff. But with crowdsourced data collection methods, institutions of any size have the potential to participate

in and benefit from research. We wish to underscore this particular advantage: we have observed that there is a general lack of research initiatives that involve small institutions in the museum field. This dynamic presents a representation problem in museum studies—if small institutions are not well represented in research, how can this research be considered representative? By extension, if resources in the museum field are allocated based on research, then these resources likely do not meet the needs of small institutions. Research based on crowdsourced data collection could be one way for small museums to participate in and benefit from research.

Another opportunity presented by this research approach is its potential to increase institutions' capacity for research. We believe that most institutions that don't have the capacity or skills to develop their own research can participate in crowdsourced data collection. We estimate that the costs of participation were reasonable in our study: a representative from each institution completed an online training in protecting human research subjects. We estimate that they also spent a half hour to two hours each month on the study, depending on the modules for that month. On their part, staff who did not have experience with research had the opportunity to orchestrate and implement research at their organization. While our research was not designed to conduct training beyond data collection, we speculate that future research initiatives could invest more heavily in training representatives from participating organizations to assist with developing the instrument modules and interpreting the results.

Finally, the crowdsourced data collection approach can enable researchers and leaders to study an entire sector in ways that aid advocacy. Data that conveys a sector's value to the public is essential for making the case for new resources for that field. Most notably, these resources can be in the form of funding dollars from government and philanthropic sources. Crowdsourced, sector-level data can also move a field toward social and professional change. We anticipate and are hopeful that institutions that participate in crowdsourced data studies will have a sense of ownership over the results and implications of the research.

Limitations & considerations

We also observed limitations that may constrain other researchers if they choose to use this method. The scope of our study was ambitious: we worked with 95 institutions spread across North America for 12 months on assembling a single data set for this study. It required a team of researchers to be involved in refining survey modules based on previous months' data collection rates, conducting initial analysis on in-progress data to share with participants, communicating consistently and clearly, supporting coordinators at institutions who were new to research involving human participants, and cleaning a data set with a wide variety of nuances. This undertaking required a detail-oriented research coordinator who could manage all of these variables from month to month and maintain a positive personal relationship with the participating institutions. Ultimately, this method was time intensive for researchers, and required patience while data slowly accumulated over time. While this method was expensive to conduct, it demonstrates that associations have the capacity to gather rich and robust public data through their membership. We recommend that researchers weigh the benefit of a study of this size and length.

There may be limitations for institutions that participate in this crowdsourced data collection as well. Staff turnover among institutional representatives created challenges in continuity and clarity around the project at several institutions. We also observed that no schedule will work well for every institution and scheduling created challenges with capacity at several institutions. For instance, most of the modules involving visitors were distributed during the summer months, which are highly demanding for staff at some institutions. In some of these cases, staff did not have the capacity to participate in the module for that month. Meanwhile, other institutions' fluctuations in visitation did not affect the modules.

Another consideration for participating institutions is that this approach may not produce extensive institution-level data of use to institutions themselves. Instead, the technique prioritizes sector-level data. This factor may reduce institutions' interest in participating. However, it can be said that what benefits the sector also benefits the institution, as long as expectations are set for the process. To obtain institution-level data, we advocate for other research methods, such as multi-site surveys and case studies.

Conclusion

The method of crowdsourcing data collection to assemble, maintain, and update a sector-wide data set appears to be an unconventional yet effective approach for social science research in the museum field and other sectors. Our study produced an expansive data set that accurately portrays the zoo and aquarium sector, and components of the public's interactions with these institutions. This methodological approach could be successfully applied elsewhere, in other types of professional sectors, including those that serve local communities, nurture the arts, provide out of school learning, and offer critical social services to their communities. We hope that future research on this technique can surface its potential value and limitations, especially in the following areas: inclusion of institutions that lack resources to mount their own research, the effect on increasing the research skills of staff, and use of the research results after a crowdsourced study is completed. Finally, we conclude that while a crowdsourced data collection approach may be useful, it should not replace other methods the museum field and other sectors already use. Rather, we see this approach as another tool to add to the box.

Note

1. The X axis shows the value of responses on the so-called standard or z scale. The raw scores on different scales are transformed into numeric scores on the same standard scale with a mean of 0 and a standard deviation of 1 (Weisstein, n.d.). The Y axis shows the probability of a response at each level on the standard scale.

Acknowledgments

The research was conducted as part of the National Science Foundation-funded STEM Matters: Investigating the Confluence of Visitor and Institutional Agendas initiative (DRK-1612729 &

DRK-1612699) led by Knology, Oregon State University's STEM Research Center, and COSI's Center for Research & Evaluation in collaboration with the Association of Zoos and Aquariums. The authors are solely responsible for the content of this paper. First and foremost, we thank the personnel from the 95 zoos and aquariums whose collegial participation in data collection made this study possible. We also thank team members Joanna Brucker, Jena Barchas-Lichtenstein, Rebecca Norlander, Darcey Glasser, and Konstantinos Voiklis for their support with data collection, analysis, and review. We also extend our thanks to Dr. Tara Teel from Colorado State University for providing information about the Wildlife Value Orientation scale.

ORCID

John Voiklis http://orcid.org/0000-0002-1590-9028
Kate Flinner http://orcid.org/0000-0003-3936-2108
Shaun Field http://orcid.org/0000-0002-3730-5006
Rupanwita Gupta http://orcid.org/0000-0002-7276-4312
John Fraser http://orcid.org/0000-0001-8383-0699
Joseph de la Torre Dwyer http://orcid.org/0000-0002-2717-9077
Shelley Rank http://orcid.org/0000-0003-2097-1165

References

AZA – Association of Zoos & Aquariums. (2018). 2018 Benchmark Reports.

Bureau of Economic Analysis. (2020). Statistical Areas. https://apps.bea.gov/regional/docs/msalist.cfm?mlist=2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's mechanical turk as a tool for experimental behavioral research. *PloS one*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Flinner, K., Field, S., Fraser, J., Voiklis, J., Attaway, E., Thomas, U. G. (2020). ACM Trends: Data-Driven Outcomes for Children's Museums. Retrieved January 4, 2022, from https://knology.org/article/acm-trends-data-driven-outcomes-for-children-s-museums

Fraser, J., Bicknell, J., Sickler, J., & Taylor, A. (2009). What information do zoo & aquarium visitors want on animal identification labels? *Journal of Interpretation Research*, *14*(2), 7–19.

Geiger, N., Gasper, K., Swim, J. K., & Fraser, J. (2019). Untangling the components of hope: Increasing pathways (not agency) explains the success of an intervention that increases educators' climate change discussions. *Journal of Environmental Psychology*, *66*, 1–8.

Geiger, N., Swim, J. K., & Fraser, J. (2017a). Catalyzing public engagement with climate change through informal science centers. Creating a climate for change: Interventions, efficacy and public discussion about climate change. *Journal of Environmental Psychology*, *51*, 104–116.

Geiger, N., Swim, J. K., Fraser, J., & Flinner, K. (2017b). Catalyzing public engagement with climate change through informal science centers. *Science Communication*, *39*(2), 221–249. https://doi.org/10.1177/1075547017697980

Gupta, R., Fraser, J., Rank, S. J., Brucker, J. L., & Flinner, K. (2019). Multi-site case studies about zoo and aquarium visitors. *Perceptions of the STEM Learning Ecology, Visitor Studies*, *22*(2), 127–146. https://doi.org/10.1080/10645578.2019.1661737

Gupta, R., Voiklis, J., Rank, S. J., Dwyer, J. D. L. T., Fraser, J., Flinner, K., & Nock, K. (2020). Public perceptions of the STEM learning ecology–perspectives from a national sample in the US. *International Journal of Science Education, Part B*, *10*(1), 1–15. https://doi.org/10.1080/21548455.2020.1719291

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., & Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829–842. https://doi.org/10.3758/s13428-015-0642-8

Guthrie, J., Burritt, J., & Evans, E. (2011). The relationship between academic accounting research and professional practice. In E. Evans, R. Burritt, & J. Guthrie (Eds.), *Bridging the gap between academic accounting research and professional practice* (Vol. 2, pp. 9–20). The Institute of Chartered Accountants in Australia.

Helmsley-Brown, J., & Sharp, C. (2003). How do teachers use research findings to improve their professional practice. *Oxford Review of Education*, *29*(4), 449–471.

Howe, J. (2006 June 1). The rise of crowdsourcing. *Wired*, *14*(6), 1–4. https://www.wired.com/2006/06/crowds/

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.

Kisiel, J., & Ancelet, J. (2009). Uncovering visitor conceptions of fossils and the fossil record. *Visitor Studies*, *12*(2), 133–151.

Manfredo, M. J., Sullivan, L., Don Carlos, A. W., Dietsch, A. M., Teel, T. L., Bright, A. D., & Bruskotter, J. (2018). *America's wildlife values: The social context of wildlife management in the U.S. national report from the research project entitled "America's Wildlife Values"*. Colorado State University, Department of Human Dimensions of Natural Resources.

Moss, A., Jensen, E., & Gusset, M. (2017). Probing the link between biodiversity-related knowledge and self-reported proconservation behavior in a global survey of zoo visitors. *Conservation Letters*, *10*(1), 33–40.

National Weather Service. (n.d). Cooperative Observer Program (COOP). National Weather Service, National Oceanic and Atmospheric Administration, US Department of Commerce. Retrieved August 9, 2021, from. https://www.weather.gov/coop/overview

Nielsen, A. (2019). *Practical time series analysis: Prediction with statistics and machine learning*. O'Reilly Media, Inc.

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*(9), 1352–1375. https://doi.org/10.1080/00140130802170387

Peterman, K., & Gathings, M. J. (2019). Using a community-created multisite evaluation to promote evaluation use across a sector. *Evaluation and Program Planning*, *74*, 54–60. https://doi.org/10.1016/j.evalprogplan.2019.02.014

Pew Research Center. (2019a November). *U.S. Public Views on Climate and Energy*. Washington, D.C. https://www.pewresearch.org/science/wp-content/uploads/sites/16/2019/11/Climate-Energy-REPORT-11-22-19-FINAL-for-web-1.pdf

Pew Research Center. (2019b December). *Most Americans Say the Economy Is Helping the Rich, Hurting the Poor and Middle Class*. https://www.pewsocialtrends.org/wp-content/uploads/sites/3/2019/12/SDT.12.11.19_economic.recovery.REPORT_forrelease.pdf

Pew Research Center. (2020 April). *About Half of Lower-Income Americans Report Household Job or Wage Loss Due to COVID-19*. https://www.pewsocialtrends.org/wp-content/uploads/sites/3/2020/04/PSDT_04.21.20_covidfinance_FULL.REPORT.pdf

Pollard, M., & Baird, M. D. (2017). *The RAND American life panel: Technical description*. RAND Corporation. RR-1651. https://www.rand.org/content/dam/rand/pubs/researchreports/RR1600/RR1651/RANDRR1651.pdf

Smith, T. W., & Son, J. (2010). An Analysis of Panel Attrition and Panel Change on the 2006-2008 General Social Survey Panel. GSS Methodological Report No. 118. National Opinion Research Center, University of Chicago.

Wavell, C., Baxter, G., Johnson, I., Williams, D. (2002). *Impact Evaluation of Museums, Archives and Libraries: Available Evidence Project. Resource: The Council for Museums, Archives and Libraries*. https://www.informalscience.org/impact-evaluation-museums-archives-and-libraries-available-evidence-project

Weisstein, E. W. (n.d). Z-Score. Wolfram Research, Inc. Retrieved August 14, 2021, from https://mathworld.wolfram.com/z-Score.html

Yung, G., & Liu, Y. (2020). Sample size and power for the weighted log-rank test and Kaplan-Meier based tests with allowance for non-proportional hazards. *Biometrics*, *76*, 939–950. https://doi.org/10.1111/biom.13196

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493–504. https://doi.org/10.1037/pspa0000056

About the authors

John Voiklis is a cognitive and social psychologist who uses data science techniques to understand the "social life" of human cognition. Specifically, he studies how people think about their social world that includes categories, stereotypes, attitudes – and how they act in that social world through choices, behaviors, and relationships. johnv@knology.org

Kate Flinner is a researcher and writer who studies how people interact with cultural institutions and community organizations. kate.flinner@gmail.com

Shaun Field is a researcher and project manager at Knology. He managed the instrument development, data collection, and analysis of this study. His background is in Climate science and social impact. shuanf@knology.org

Rupanwita Gupta is a conservation psychologist and evaluator who studies the role of equity and inclusion in the environmental movement and STEM learning in formal and informal settings. Her research also focuses on climate resilience partnerships between community organizations and cultural institutions. rupu.gupta@gmail.com

John Fraser is a psychologist and architect serving as President & CEO of Knology and a Past-President of the Society for Environmental, Population, & Conservation Psychology, Division 34 of the American Psychological Association. His research focuses on how collective identification with groups influences self-efficacy and willingness to engage in positive social change. johnf@knology.org

Joseph de la Torre Dwyer is a political scientist at Knology whose research focuses on equality of opportunity, especially with respect to labor market policies. josephd@knology.org

Shelley Rank is a Research and Evaluation Associate at the Wildlife Conservation Society (WCS). Prior to joining WCS she worked at Knology as a researcher and project manager. Her current research focuses on visitor engagement in zoos and aquariums and the relationship between people, animals, and nature. srank@wcs.org

Kathryn Nock is an educator, writer, and researcher interested in the formation of public opinions on social and environmental issues related to climate change. kathryn.m.nock@gmail.com