A Science Products Inventory for Citizen-Science Planning and Evaluation

ANDREA WIGGINS, RICK BONNEY, GRETCHEN LEBUHN, JULIA K. PARRISH, AND JAKE F. WELTZIN

Citizen science involves a range of practices involving public participation in scientific knowledge production, but outcomes evaluation is complicated by the diversity of the goals and forms of citizen science. Publications and citations are not adequate metrics to describe citizen-science productivity. We address this gap by contributing a science products inventory (SPI) tool, iteratively developed through an expert panel and case studies, intended to support general-purpose planning and evaluation of citizen-science projects with respect to science productivity. The SPI includes a collection of items for tracking the production of science outputs and data practices, which are described and illustrated with examples. Several opportunities for further development of the initial inventory are highlighted, as well as potential for using the inventory as a tool to guide project management, funding, and research on citizen science.

Keywords: citizen science, evaluation, science products inventory, altmetrics, science policy

The term *citizen science* is used to describe a wide variety of projects involving nonprofessionals in producing scientific knowledge (Bonney et al. 2009). Citizen science is often considered a research strategy or methodology that involves actively engaging members of the public in one or more core steps of scientific inquiry. Citizen science has proven useful for answering questions about plant and animal distributions across landscapes, which can result in advances to basic research and conservation actions (Cooper et al. 2014, McKinley et al. 2015, Ries and Oberhauser 2015, Sullivan et al. 2016). It is also valuable for research that requires the human processing and interpretation of large amounts of data (Swanson et al. 2016), such as the projects hosted on the Zooniverse platform, which focus on image classification and transcription. When projects are carefully designed and well managed, they can produce science not achievable through other means (Bonney et al. 2009).

Defining metrics for assessing the full range of citizen-science outputs will help describe the breadth and value of this emerging field. Such metrics can also help project designers understand the relationships among project products and data practices, as well as set informed expectations. Although research is demonstrating learning outcomes from citizen science (Masters et al. 2016) and evaluating participant outcomes (e.g., Phillips et al. 2014), metrics for

evaluating science productivity and conservation outcomes are too simplistic or limited in applicability.

Science productivity in citizen science is not readily assessable through traditional counts of papers and citations (Dickinson et al. 2012, Wiggins and Crowston 2015). Current reliance on *bibliometrics*, or analyses that use publication and citation data to evaluate scholarship, "are based on assumptions, whether implicit or explicit, about how and why the authors of one work cite other works" (Borgman and Furner 2002, p. 57). Such analyses also assume that scholarly publications are a primary goal, which is not always true in citizen science. *Scientometrics* considers other quantitative measures of science activities and science policy impacts (Hood and Wilson 2001), focused on institutional rankings, faculty productivity and tenure standards, and productivity assessments, but still relies primarily on publication and citation counts. The Internet created new opportunities for evaluating impact: Article commenting, Wikipedia mentions, blogging, online video, and open data repositories provide new measures, known as *altmetrics*, to complement citation statistics (Priem et al. 2012). In citizen science, altmetrics may reflect goals for education, policy, and conservation, as well as evidence of scientific impact.

Citizen-science project productivity has been evaluated through measures such as publication rate, completeness

Table 1. Expert panel members.

Name	Affiliation (at time of panel involvement)	Project affiliation and/or expertise area
Rick Bonney	Cornell Lab of Ornithology	Evaluation
Anne Bowser	University of Maryland	Participatory design
Eric Graham	University of California, Los Angeles	App development, What's Invasive!
Sandra Henderson	NEON	Education, Project BudBurst
Megan Hines	Wildlife Data Integration Network	Data management
Gretchen LeBuhn	San Francisco State University	The Great Sunflower Project
Kelly Lotts	University of Idaho	Butterflies and Moths of North America
William Michener	University of New Mexico	Data management
Abe Miller-Rushing	National Park Service	Project design and management
Greg Newman	Colorado State University	IT development, CitSci.org
Karen Oberhauser	University of Minnesota	Monarch Larva Monitoring Project
Julia K. Parrish	University of Washington	Project leadership, COASST
Alyssa Rosemartin	USA National Phenology Network	Data management, Nature's Notebook
Eric Russell	National Geographic Society	IT development, FieldScope
Jennifer Shirk	Cornell Lab of Ornithology	Project development
Arfon Smith	Zooniverse	Community management, Galaxy Zoo
Robert D. Stevenson	University of Massachusetts Boston	Project design, PRCWA
Julian Turner	Colorado State University	IT management, Community Collaborative Rain, Hail, and Snow Network
Jake Weltzin	US Geological Survey	Project leadership, Nature's Notebook
Andrea Wiggins	DataONE	Project and IT design
Bruce Wilson	Oakridge National Labs	Data management

of analysis, resource savings, and effort distribution, but without evaluating "alternate" products such as data sets and secondary publications (Cox et al. 2015). Theobald and colleagues (2015) found that only about 12% of citizen-science projects surveyed showed evidence of turning data into peer-reviewed publications but did not evaluate other types of science outputs. Other potentially relevant indicators of productivity include data-set generation (Lagoze 2014), conservation actions (Sullivan at al. 2016), environmental-justice outcomes (Haklay 2013), education and community outcomes (Jordan et al. 2012), and policy impacts (McKinley et al. 2015). Although some of these measures may be generalizable, many cannot be used outside of a specific project or technology platform.

The lack of standard assessments for science productivity underlies a barrier to the acceptance of citizen science as a research strategy: uncertainty among peer reviewers about how to evaluate the scientific merits of citizen-science projects, particularly for funding decisions (Shirk et al. 2012, Bonney et al. 2014). Carefully considered criteria by which to evaluate the the science outputs of citizen-science projects will advance the research community's ability to review citizen-science proposals and can support the work of federal agencies to adopt and develop citizen-science activities (Holdren 2015). To meet these needs, we developed a comprehensive inventory of outputs based on observed characteristics of established successes, the science products inventory (SPI).

Methods

To generate an initial inventory of science productivity indicators, we adopted a variation of the *Delphi method* (Dalkey and Helmer 1963, Linstone and Turoff 2002). This method involves iterative, structured idea generation and ranking by a panel of experts until convergence is achieved. Because the Delphi method relies on anonymity not feasible for our project, we combined elements of the Delphi method with the nominal group technique processes for idea generation, voting, and ranking (Potter et al. 2004).

Our panel included the members of the DataONE Public Participation in Scientific Research Working Group and invited guests (table 1). The members of the panel were selected to maximize the diversity of perspectives, which required recruiting individuals with extensive experience in citizen science. The group's joint expertise involved direct experience with more than two dozen projects ranging from local to global in scale, developing and delivering citizen-science data products and conducting research and evaluation on citizen science. We also drew from across sectors—academia, nonprofits, and federal agencies—and across ranks, from PhD students to full professors and seasoned federal scientists. Of the 20 participating individuals,

14 had graduate training in ecology and related disciplines, 2 in information science and technology, 1 in astrophysics, and 1 in computer science. In addition, of the 20 participants, 18 were professional scientists, 2 were professional software developers supporting citizen-science projects, and all were considered practitioners because of the nature of their involvement in citizen science as project staff and project participants.

Our hybrid method was enacted through two intensive 3-day workshops held 6 months apart, with facilitated brainstorming, categorization, and prioritization of evidence and context for describing science productivity in citizen science. Every item in the inventory was vetted through in-depth discussions, with only two items added after pilot testing.

During the first meeting, brainstorming, clustering, and consensus processes generated initial lists of science outputs, descriptive characteristics of projects, and potential measures. The primary activities for this meeting included an agenda-setting overview, open discussion of the goals and desired outcomes for the process, structured brainstorming exercises to identify science products and potential metrics for evaluating them (concept generation, categorization, and voting), initial testing of the metrics for a few projects, discussion of strategy for data collection, and refinement of sampling criteria.

The resulting metrics for measuring science products were aggregated into an inventory spreadsheet for testing. The participants "piloted" the inventory by independently completing inventories for projects with which they were familiar while recording feedback for improvement. Based on the pilot feedback, the second meeting further refined item definitions, finalized the recommended metrics for each item, and selected and initiated documentation of case studies.

The panel completed eight case studies that represent a variety of participation models, scientific contexts, and types of success (box 1; table 2). Case studies were selected for diversity in scientific discipline (several topics within ecology, astronomy, and precipitation), geographic scale (local to global and aspatial), participation scale (40 to 325,000 contributors), and project goals including basic science and decision support for resource management and disaster prevention.

Results

The SPI includes multiple potential outputs that indicate scientific progress or products (table 3) and contextual details on data practices (table 4). The following sections describe these outputs in more detail, with examples from our case studies. The supplemental material includes templates of the SPI, as well as examples of the SPI for two representative case studies.

Science products. Science products include varied forms of dissemination for multiple audiences and purposes (table 3), via a variety of formats and venues, such as publications, videos, and social media, which describe or discuss the project's research design, progress, and results. Science products are subdivided into four categories: written, data, management and policy, and communication.

Written. Citizen-science projects present their research to multiple audiences through a variety of formal and informal products designed for multiple stakeholders and purposes (table 3). We included four written product types, most of which involve peer review: scholarly publications, dissertations and theses, reports, and competitive grant awards.

Scholarly peer-reviewed publications are typically assessed with a count of papers selected for inclusion on the basis of disciplinary conventions. Although simple to specify, this accounting can become complicated by the surprisingly broad variety of uses of citizen-science data (Lagoze 2014). Full publication counts are often underrepresented by indexes of journal databases, so staff for several case-study projects kept manual records informed by automated citation alerts and correspondence with external data users. Although several case-study project leaders wished to subdivide "scholarly publications" into subcategories such as papers *about* a citizen-science project and papers *based on the data* from a citizen-science project, they found it unfeasible to implement these subcategories retrospectively. Dissertations and theses are more easily categorized, and although typically "unpublished," they may lead to subsequent scholarly publications and promote scientific inquiry more broadly. Formal and informal reports produced from citizen-science data can include white papers, technical reports, environmental assessments, species status reports, and policy advisory memos. Reports can be strikingly similar to scholarly papers focused on supporting management, conservation, and policy goals, and many are routinely peer reviewed. Aside from relevant reports curated by project organizers, however, these types of publications can be difficult to discover and track systematically because of nonstandardized or inconsistent use of keywords (e.g., Cooper et al. 2014). Finally, initial implementation of the inventory with science teams at a government research facility identified competitive grant awards as a strong indicator of project success and primary criterion for evaluation. This item can be measured through number of awards received or monetary value and could be further subdivided by award type or funder.

An example of a project with all four types of written products is eBird. The project team regularly produces peer-reviewed scholarly publications (e.g., Sullivan et al. 2016) and also keeps account of publications by others resulting from access to eBird data. To date, eBird has identified more than 150 scholarly papers that either studied the project or used its data, and it has served as a case study for dissertations studying citizen science. The data have been used to study such topics as bird biology, natural-resource management, and machine learning. The data are also applied extensively in "gray literature," including technical reports

> ## Box 1. Summary descriptions of case-study projects.
>
> **eBird** has collected more than 400 million observations of bird abundance and distribution from around the world since 2005, representing a collective investment of nearly 30 million hours in the field. The data have been used for numerous conservation and management applications, many public-interest publications, hundreds of public talks and presentations, and scholarly publications across a diverse range of disciplines.
>
> **Galaxy Zoo** enlisted volunteers from 2007 to 2008 in classifying the morphology of galaxies in more than a million photographs from the Hubble telescope. At the project's conclusion, all images had been classified multiple times and shared through the Sloan Digital Sky Survey, with several unexpected discoveries. Galaxy Zoo data appear in scholarly papers and the Zooniverse software platform has supported dozens of additional projects.
>
> The **Coastal Observation and Seabird Survey Team (COASST)** participants collect standardized, effort-controlled data on deceased birds and marine debris. Since 1999, monthly data from more than 450 beaches along the Pacific coastline contribute to a unique data set with extensive details documenting bird carcasses for 186 species. COASST data have been used in scientific papers, news media, and dozens of reports for regulatory action and decision support.
>
> **Monarch Larva Monitoring Project (MLMP)** volunteers have collected data on monarch butterflies' egg and larvae distribution and abundance at more than 1000 sites across North America since 1996 and raised over 15,000 larvae to examine survival and parasitism rates. The data were valuable to a recent petition to list monarchs as a threatened species under the US Endangered Species Act, scholarly publications, and a field guide to milkweeds.
>
> The **Great Sunflower Project (GSP)** launched in 2008 to collect information about pollinator service for the United States on a continental level and to evaluate and improve pollinator habitat, collecting a unique data set on pollinator presence and absence at around 8500 sites. GSP data have been used in scholarly publications across multiple fields and more than 30 talks and presentations to scientific audiences.
>
> The **Community Collaborative Rain, Hail and Snow (CoCoRaHS) Network** started in 1998 to collect precipitation data for better weather forecasting and disaster preparedness, with 12 active observation protocols including daily precipitation, significant weather, and hail. CoCoRaHS data for the United States, Canada, and Caribbean islands are used extensively in weather forecasting and reporting and feature in several scholarly publications.
>
> **Nature's Notebook** has accumulated records of the life cycles (phenology) of plants and animals in over 18,000 US locations since 2009. Operated by the USA National Phenology Network and supported by the US Geological Survey, the project also provides multitaxon national scale phenology protocols and a software platform that supports other groups. Nature's Notebook data have contributed to scholarly publications and reports for decision support and natural-resource management.
>
> The **Parker River Clean Water Association (PRCWA)** ran its tidal restrictions study from 1996 to 1997 to generate decision-support data for saltmarsh restoration in Massachusetts' Great Marsh, identifying significant tidal restrictions at half of the sites surveyed. The data were used to prioritize over a dozen restoration projects by state agencies and other authorities. Results were presented in community meetings and a technical report, with methods published as a stand-alone guide to volunteer-based assessments of tidal restrictions.

for decision support and popular media such as magazines. Finally, the project is supported in part through competitive grant awards, cumulatively in the millions of US dollars.

Data. The second category of science outputs includes raw data and value-added data products created and distributed for use by others (table 3). Theobald and colleagues (2015) observed that some of the most successful projects—at least in terms of peer-reviewed publications—are those that make their data readily available. Indicators that project data are being supplied for external users include whether the project has application programming interfaces for automated data exchange (APIs), data packages, and metadata. The availability of data visualizations can also provide an indicator of data production as a type of preliminary result (Snyder 2017). Finally, demand for project data measured in number of data requests or data volume transferred is a good indicator of data distribution. These metrics require

more substantial infrastructure. Tracking data transfer volumes (bandwidth consumption) requires having data access options already in place; such infrastructure is typically developed *after* demand for data access is established via increasing frequency of data requests that become burdensome to manage through other means. The volume of data requests is often tracked through email messages or Web form submissions when access is mediated by humans or else via server logs for self-serve database access.

The Community Collaborative Rain, Hail, and Snow Network (CoCoRaHS) is experienced in measuring data product usage. The project provides multiple APIs plus reports, descriptive metadata, and staff support. Estimates of recent demand for data averaged more than 9 gigabytes of data served daily to satisfy a total of around 14,000 requests per month. CoCoRaHS's staff can break down these statistics into more specific activities and uses on the basis of close relationships with data consumers, such

Table 2. Case-study projects and features as of 2016.

Project (see box 1 for additional info)	Sponsoring organizations	Science focus	Years active	Total volunteers	Paid staff	Peer-reviewed papers	Data points
Coastal Observation and Seabird Survey Team (COASST)	University of Washington	Seabird mortality; marine debris	1999–current	1K	6 FT and PT, 20–25 interns	20+	100K
Community Collaborative Rain, Hail, and Snow Network (CoCoRaHS)	Colorado Climate Center	Precipitation	1998–current	37.5K	6 FT	30+	32M
eBird	Cornell Lab of Ornithology	Bird abundance and distribution	2005–current	325K	20 FT and PT	150+	400M
Great Sunflower Project	San Francisco State University	Pollinator service	2008–current	120K	1 PT	10	125K
Galaxy Zoo	Oxford University and Johns Hopkins University	Galaxy morphology	2007–2008	100K	3 FT and PT	50+	40M
Monarch Larva Monitoring Project (MLMP)	University of Minnesota, Chicago Botanic Garden	Monarch butterfly abundance and distribution	1996–current	700	Varies, PT	18	24K
Nature's Notebook	USA National Phenology Network, US Geological Survey	Plant and animal life cycles	2009–current	7.6K	12 FT and PT, interns	24	8M
Tidal Restrictions Survey (PRCWA)	Parker River Clean Water Association	Saltmarsh tidal restrictions	1996–1997	40	1 PT	0	1.4K

Abbreviations: FT, full-time employees; PT, part-time employees; K, thousands; M, millions.

as the National Weather Service. The drawback of making CoCoRaHS data so readily accessible was that staff had difficulty tracking research publications. By contrast, access to comprehensive eBird data packages and downloads requires a Web form submission, which allows project staff to monitor publications by users. Because CoCoRaHS's goals are to generate data for decision support and emergency preparedness, data request volume is likely a stronger indicator of achieving their targets than the number of peer-reviewed publications.

Management and policy. A third category, management and policy products, identifies direct actions, decision-support products, and policy impacts from citizen-science projects (table 3). These items are among the least straightforward to evaluate because awareness of them is often limited and they can take many forms. The case-study projects were able to report on only the management and policy impacts with which they were directly involved. Decisions, policies, and actions can lead to conservation outcomes that are typically evaluated separately but also follow from the science. The inventory focuses on use of project data as an input to policy and management decisions, regardless of the subsequent outcomes. Management and policy impacts are likely best measured through internal tracking or other forms of direct monitoring by the parties involved in translating science to decisions and policy, who are best informed about what counts as a meaningful management or policy outcome. Three types of management and policy outputs identified as indicators of science productivity are regulatory action (e.g., enforcement or investigation by an authority), decision

support (e.g., land management or conservation actions), and forecasting or models (often used for management and decision support).

The Parker River Clean Water Association (PRCWA) was a small, short-term project that achieved substantive management impacts. Data collected over 1 year influenced the prioritization of major natural-resource management investments in multiple restoration projects. In addition, the monitoring methodology had further impact when several other coastal monitoring projects adopted it to support management decisions and restoration actions.

Communication. Public discourse and science communication products offer evidence of scientific productivity. Communications specifically targeted at public audiences require additional capacity and effort beyond the traditional research team and can help advance project goals (table 3). These indicators include written communications (e.g., blogs, newsletters, and social media); multimedia content (e.g., videos); and discursive events (e.g., public talks and presentations). Although science communication is easily consigned to an "outreach" category, it is critical for volunteer recruitment and retention, in which ongoing communication can provide evidence that science is progressing prior to the availability of formal products such as reports (Snyder 2017). Several types of science communications, configured as binary presence or absence indicators, can be extended with counts to document annual project activity levels.

The Coastal Observation and Seabird Survey Team (COASST) demonstrates strong science communications with regular e-newsletters of project progress and skills

Table 3. Science products.

Category	Product	Definition
Written	Dissertations, theses (#)	Number of theses and dissertations using data from or reporting on the project
Written	Scholarly publications (#)	Number of published peer-reviewed science papers that report on the project or apply its data
Written	Reports (#)	Number of formal reports reporting results, such as white papers, technical, and other reports
Written	Grants awarded (#, $)	Existence (or total monetary value) of competitive funding awards from private or public funders
Data	APIs (Y/N)	Existence of technologies for automated data exchange between computers
Data	Data packages (#)	Number of curated exports of data and related documentation, usually as a downloadable zip file
Data	Metadata (Y/N)	Existence of documentation describing data structure, formats, and contents
Data	Visualizations (Y/N)	Existence of visual representations of data, such as graphs, maps, and animations
Data	Specimens/samples (#)	Number of material data points in the form of physical specimens or samples
Data	Requests (# requests, transfer volume)	Number of individuals or technical systems requesting data, or volume of transferred data
Management and Policy	Regulatory action (Y/N)	Existence of legal rulings or regulation enforcement based on project data and findings
Management and Policy	Decision support (Y/N)	Existence of decisions based on project data and findings (e.g., for policy or management)
Management and Policy	Forecasting/models (Y/N)	Existence of models based on project data that simulate or predict complex phenomena
Communication	Blogs (Y/N)	Existence of online informal written communications about project processes and findings
Communication	Newsletters (Y/N)	Existence of structured publications for project stakeholders, produced in hard copy or digitally
Communication	Videos (Y/N)	Existence of publicly available digital videos on project content, activities, and findings
Communication	Presentations (Y/N)	Existence (or number) of oral presentations at conferences or public events
Communication	Website (Y/N)	Existence of dedicated website for the project

practice; Web-based interactive data visualizations highlighting trends in time, space, taxonomy, and conservation; an interim results blog with frequent use of data visualizations and graphic representations; and a Facebook page and Twitter feed driving participants to these products. Galaxy Zoo similarly established blogs and social media as standard communication tools for the globally distributed contributors to Zooniverse projects.

Data practices. Data management and sharing practices maximize the value of volunteers' contributions. To achieve this impact, data must be available, discoverable, and well documented. As for science products, data practices and tools can take both simple and complex forms, ranging from basic downloads of plain text files to direct database access with accompanying schema documents (table 4). The key components identified by our expert panel mirrored the FAIR Data Principles: findable, accessible, interoperable, and reusable (Wilkinson et al. 2016).

Findable. *Findability*, or discoverability, is evaluated with binary items (*yes* or *no*) because of the wide array of specific configurations that may be appropriate for each individual project. Metrics to assess findability include the availability of data directly from the project, typically via its website, and the availability of project data via research data repositories or registries.

Accessible. Once found, data must be relatively straightforward to access, worth the trouble of doing so, and delivered in a usable format. The inventory includes multiple ways to assess data availability, with the expectation that not all metrics will be applicable to every project. These items include the availability of data file downloads and database querying tools. In addition, documentation is important for accessibility and is evaluated through the presence or absence of explicit data licensing, descriptive metadata (i.e., a data dictionary with specifics of the data), and API documentation where applicable. For example, eBird custom query downloads are delivered with terms of use, recommended citations, metadata descriptors, and extensive documentation (Sullivan et al. 2014).

Interoperable. Accessibility and interoperability are closely linked, because interoperability supports accessibility. The inventory includes one item for interoperability, identifying whether metadata employ appropriate structural data standards, such as EML or FGDC, for machine-enabled data discovery.

Table 4. Data practices.

Category	Practice	Definitions
Findable	Data available from project website (Y/N)	Availability of data from the project's own website in a downloadable or queryable format
Findable	Data available from repositories or registries (Y/N)	Availability of data in a research data repository or via a data clearinghouse or registry
Accessible	Downloadable data file(s) available (Y/N)	Existence of download data files via project website, repository, or third party
Accessible	Tools for data exploration (Y/N)	Existence of tools for visualizing, summarizing, or querying project data via an app or website
Accessible	Data licensing specified (Y/N)	Existence of text specifying terms and conditions for data use
Accessible	Metadata available (Y/N)	Existence of documents with descriptive metadata such as known problems and data cleaning tips
Accessible	API documentation (Y/N)	Existence of documentation to support users of an API, where applicable
Interoperable	Data recorded in standard formats for discipline (Y/N)	Application of disciplinary standards for structural metadata and data formatting
Reusable	Uniqueness of data (describe)	Description of the unique contributions and features of the project's data
Reusable	Time scale of data (# yrs)	Number of years of records in the data set; may include historical data
Reusable	Spatial scale of data (describe)	Description of the geographic range for project data, such as continent, country, state, city, or watershed
Reusable	How much data (# data points, describe)	Description of data volume in terms relevant to the data collected, such as number of data points
Reusable	Errors documented (Y/N)	Existence of documentation for known errors in the data set
Reusable	Quality assurance or quality control documented (Y/N)	Existence of documentation for quality-assurance and quality-control processes
Reusable	Changes documented (Y/N)	Existence of documentation for data edited after initial receipt
Reusable	Questionable data flagged (Y/N)	Existence of documentation for data considered questionable or problematic
Reusable	Software or platform development (Y/N)	Existence of software or hosted technologies (platforms) that support external projects

Reusable. *Reuse*, or use of data by a third party, is contingent on data management practices plus the inherent value of the data and evidence of its quality and rigor. Mainstream scientific communities often require precise metadata on the location, date and time, and effort-control of data, in addition to information on verifiability (Burgess et al. 2017). The inventory assesses the "unique" qualities of the data descriptively (i.e., as free text). Potential uses of the data are frequently determined by their spatial and temporal extent (Theobald et al. 2015), included as categorical items. The total number of data points can also influence potential uses, although the unit of observation or measurement is highly variable (table 2).

Reuse depends on careful documentation, summarized with binary items for the presence of documentation on known errors, quality-assurance or quality-control processes, questionable data points, and data provenance or audit trails for changes to data after initial ingestion. For example, data on the MLMP project website are provided unedited, so downloads of basic monarch density data come with a warning that there may be errors in the data, and more detailed data sets are provided with recommended cleaning criteria.

Reuse also applies to data infrastructures, such as the provision of freely available reusable software (e.g., the Zooniverse code base, available through GitHub) and low- or no-cost hosting platforms (e.g., Nature's Notebook). Software and related technical infrastructure, platforms, and services are a critical element in many citizen-science projects, so reuse of existing systems can contribute substantially to the field. For projects offering technical infrastructure, the number of known groups adopting the platform or using the code base may be a good measure of impact. Among our case studies, both Zooniverse and Nature's Notebook offered no-cost platforms. eBird has informally been used to similar effect, with a fee-based service for customized portals.

Discussion

The SPI is a tool for documenting science products and data practices as indicators of the science contributions of individual citizen-science projects. It has applications for objective project evaluations both internally (by project leaders) and externally (e.g., by external evaluators).

Projects interested in planning and self-evaluation can apply the SPI internally to assess actual and potential science productivity or to adjust activities or resource allocations.

This requires identifying the applicable inventory items, which may involve adapting some measures for project-specific needs, eliminating others that are inapplicable, and collecting data from a variety of sources, including project staff and external records. Additional considerations include potential for weighting specific inventory items to reflect project priorities.

The SPI could be similarly applied by funding or hosting organizations to guide assessments of current or potential future projects and inform resource allocations. The inventory was designed with this likely application in mind. Although the inventory can provide initial guidelines for project evaluation and comparison, it is open ended and descriptive enough to require informed judgment. Similarly, it is flexible and customizable, such as via weighting of items aligned with funder objectives or organizational goals.

The SPI also lays a foundation for meaningful research on the relationships between data practices and science products and can be used in combination with other citizen-science evaluation tools for more comprehensive project assessment. By advancing the use of inventory-based evaluation tools, such as this one, the field can support evidence-based decision-making for science funding, targeted development of resources to strengthen citizen-science projects, and transparency in support of many objectives, such as more equitable and informed peer review of citizen-science products.

These potential applications of the SPI also raise a point of concern for practitioners and reviewers: Any such inventory must be used judiciously to prevent unintended apples-and-oranges comparisons between different projects, and, as in other settings, productivity measures should be interpreted with care, because their value varies by context and purpose. Organizational features may affect the resources devoted to the project and the expected project outputs; for example, well-established projects such as eBird and Galaxy Zoo are more likely to have more completed science products (table 3) and more sophisticated data practices (table 4) than recently established projects and those that were never intended to produce scholarly publications (e.g., PRCWA). The parameters that make projects more or less directly comparable are a matter of judgment and should consider the purpose of comparison and availability of supporting descriptive information about projects.

Conclusions

As an evaluation framework for science productivity in citizen-science projects, the SPI presented in this paper is based primarily on expert consensus and exemplars. It should be applied with care and consideration because not all citizen-science projects are necessarily designed to a singular purpose such as science products. This initial version of the SPI provides a structured and relatively objective inventory for evaluating the science products of citizen science to address multiple needs across practical, policy, and research contexts.

We note that many of the items in this inventory are equally applicable to evaluation of science productivity more broadly.

This study is limited by its methods, which relied on an expert panel to develop and prioritize items and used case studies to identify and illustrate science productivity. In addition, it reflects the priorities of scientists rather than of the participants, who may place different value on these and other science products. To address these limitations, future work could further develop and refine this tool and evaluate its utility for reflecting citizen-science participants' interests. For example, a number of inventory items would benefit from establishing categorical values through empirical study. Alternate items may also emerge through usage and to assess different types of products, resources, or practices. Several items should be revised as technologies evolve, and additional work is needed to identify generalizable indicators of excellence in citizen-science projects more broadly. The SPI offers an initial framework for guiding both citizen-science project management and research on the science of citizen science.

Supplemental material
Supplementary data are available at *BIOSCI* online.

References cited

Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips TB, Rosenberg KV, Shirk J. 2009. Citizen science: A developing tool for expanding science knowledge and scientific literacy. BioScience 59: 977–984.

Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, MillerRushing AJ, Parrish JK. 2014. Next steps for citizen science. Science 343: 1436–1437.

Borgman CL, Furner J. 2002. Scholarly communication and bibliometrics. Annual Review in Information Science and Technology 36: 2–72.

Burgess H, DeBey LB, Froehlich H, Schmidt N, Theobald EJ, Ettinger AK, HilleRisLambers J, Tewksbury J, Parrish JK 2017. The science of citizen science: Exploring barriers to use as a primary research tool. Biological Conservation 208: 113–120.

Cooper CB, Shirk JL, Zuckerberg B. 2014. The invisible prevalence of citizen science in global research: Migratory birds and climate change. PLOS ONE 9 (art. e106508).

Cox J, et al. 2015. Defining and measuring success in online citizen science: A case study of Zooniverse projects. Computing in Science and Engineering 17: 28–41.

Dalkey N, Helmer O. 1963. An experimental application of the Delphi method to the use of experts. Management Science 9: 458–467.

Dickinson JL, et al. 2012. The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment 10: 291–297.

Haklay M. 2013. Citizen science and volunteered geographic information: Overview and typology of participation. Pages 105–122 in Sui DZ, Elwood S, Goodchild M, eds. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice. Springer.

Holdren J. 2015. Memorandum to the Heads of Executive Departments and Agencies: Addressing Societal and Scientific Challenges through Citizen Science and Crowdsourcing. White House Office of Science and Technology Policy.

Hood W, Wilson C. 2001. The literature of bibliometrics, scientometrics, and informetrics. Scientometrics 52: 291–314.

Jordan RC, Ballard HL, Phillips TB. 2012. Key issues and new approaches for evaluating citizen-science learning outcomes. Frontiers in Ecology and the Environment 10: 307–309.

Lagoze C. 2014. eBird: curating citizen science data for use by diverse communities. International Journal of Digital Curation 91: 71–82.

Linstone HA, Turoff M, eds. 1975. The Delphi Method: Techniques and Applications. Addison-Wesley.

Masters K, Oh EY, Cox J, Simmons B, Lintott C, Graham G, Greenhill A, Holmes K. 2016. Science learning via participation in online citizen science. Journal of Science Communication 15 (art. A07).

McKinley DC, et al. 2015. Investing in Citizen Science Can Improve Natural Resource Management and Environmental Protection. Issues in Ecology, vol. 19. Ecological Society of America.

Phillips TB, Ferguson M, Minarchek M, Porticella N, Bonney R. 2014. User's Guide for Evaluating Learning Outcomes in Citizen Science. Cornell Lab of Ornithology.

Potter M, Gordon S, Hamer P. 2004. The nominal group technique: A useful consensus methodology in physiotherapy research. New Zealand Journal of Physiotherapy 32: 126–130.

Priem J, Piwowar HA, Hemminger BM. 2012. Altmetrics in the wild: Using social media to explore scholarly impact. arXiv 2012 (art. 1203.4745).

Ries L, Oberhauser K. 2015. A citizen army for science: Quantifying the contributions of citizen scientists to our understanding of monarch butterfly biology. BioScience 65: 419–430.

Shirk JL, et al. 2012. Public participation in scientific research: A framework for deliberate design. Ecology and Society 17 (art. 29).

Snyder J. 2017. Vernacular visualization practices in a citizen science project. Pages 2097–2111 in Lee CP, Poltrock S, Barkhuus L, Borges M, Kellogg W, eds. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. Association for Computing Machinery.

Sullivan BL, et al. 2014. The eBird enterprise: An integrated approach to development and application of citizen science. Biological Conservation 169: 31–40.

Sullivan BL, et al. 2016. Using open access observational data for conservation action: A case study for birds. Biological Conservation 208: 15–28.

Swanson A, Kosmala M, Lintott C, Packer C. 2016. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conservation Biology 30: 520–531. doi:10.1111/cobi.12695

Theobald EJ, et al. 2015. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation 181: 236–244.

Wiggins A, Crowston K. 2012. Goals and tasks: Two typologies of citizen science projects. Pages 3426–3435 in Institute of Electrical and Electronics Engineers (IEEE) Computer Society Conference Publications Operations Committee, eds. Proceedings of the 45th Hawaii International Conference on Systems Sciences (HICSS). IEEE.

———. 2015. Surveying the citizen science landscape. First Monday 20 (art. 5520). (5 March 2018; *http://dx.doi.org/10.5210/fm.v20i1.5520*)

Wilkinson MD, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (art. 160018).

Andrea Wiggins (wiggins@unomaha.edu) is affiliated with the College of Information Science and Technology at the University of Nebraska at Omaha. Rick Bonney (reb5@cornell.edu) is affiliated with the Cornell Lab of Ornithology at Cornell University, in Ithaca, New York. Gretchen LeBuhn (lebuhn@sfsu.edu) is affiliated with the Department of Biology at San Francisco State University, in California. Julia K. Parrish (jparrish@uw.edu) is affiliated with the School of Aquatic and Fishery Sciences at the University of Washington, in Seattle. Jake F. Weltzin (jweltzin@usgs.gov) is affiliated with the US Geological Survey, in Tucson, Arizona.