CITIZEN SCIENCE DATA FACTORY

A Distributed Data Collection Platform for Citizen Science

Part 2: Technology Evaluation

Prepared by

azavea scistarter.com

The Citizen Science Data Factory report includes four parts:

- Part 1: Data Collection Platform Evaluation
- Part 2: Technology Evaluation
- Appendix A: Wireframe Designs
- Appendix B: Cloud Computing Performance Testing

It is available for download and distribution from:

- http://www.azavea.com/research/company-research/citizen-science/
- http://www.scistarter.com/research/

Funded by:

The report was funded in part by a grant from the Alfred P. Sloan Foundation. The opinions expressed herein are those of the authors and do not represent the opinion of the Foundation.

This report was researched and developed by Azavea and SciStarter, including contributions from:

- Darlene Cavalier, SciStarter
- Robert Cheetham, Azavea
- Rob Emanuele, Azavea
- Jeff Frankl, Azavea
- Mary Johnson, Azavea
- Tamara Manik-Perlman, Azavea
- Josh Marcus, Azavea
- Michael Tedeschi, Azavea

This report also benefited from input provided by both the project representatives interviewed and input from the Citizen Science Community Forum.

Cover Photography Attribution:

- Bird Watching by Daniel Schwen
- Water Sampling by Albert Herring

Azavea is a B Corporation that creates civic geospatial software and data analytics. Our mission is to apply geospatial technology for civic and social impact and to advance the state-of-the-art through research.

SciStarter brings together citizen scientists; thousands of potential projects offered by researchers and organizations; and the tools, resources, and services that enable people to find, pursue and enjoy these activities.

Published by:

Azavea
340 N 12th St, Suite 402
Philadelphia, PA 19107
Telephone: +1 215 925 2600
Web: www.azavea.com
Email: info@azavea.com

SciStarter
306 Delancey St
Philadelphia, PA 19106
Telephone: +1 267 253 1310
Web: www.scistarter.org
Email: darlene@scistarter.com

Table of Contents

Introduction

The tremendous growth of mobile, sensor and internet technologies promises new transformations in what is possible, both in everyday lives and in citizen science. But, as a recent CMU paper[1] evaluating mobile platforms for citizen science summarized:

> The proliferation of mobile computing technologies is making our urban environments rich in terms of sensing, providing diverse channels to scientists for data collection, and creating tremendous opportunities for everyday people to engage in scientific projects2. As such, mobile devices are well suited for spontaneous data collection by everyday people. However, under its guise of simplicity lies technical expertise and complex infrastructure. As a result, building such applications is a large investment that may limit smaller organizations.

As we consider the power of the mobile phones we carry with us every day, we can imagine plausible but surprising new applications. The Funf platform developed at MIT springs from the insight that our phones are already packed with sensors, and the data from those sensors could be shared and analyzed. An explosion of new Kickstarter projects strive to turn our cellphones into gateways for other devices and sensors to communicate, making it possible for anyone to use their phone as a "tricorder" from science fiction, collecting scientific observations from sensors to be automatically transmitted back to a central repository.

Sensor and internet technologies are an exciting and quickly evolving arena as well. A tremendous migration is underway on the internet changing from its original addressing schema ("IP v4") to a schema ("IP v6"), which has been an ambitious and costly undertaking. Why change the way that computers are addressed (similar to your own address or phone number) on the internet? The original IPv4 schema provided for 4.3 billion unique addresses, one for each server that needed to communicate on the public internet. But while the original adopters of the IPv4 schema wouldn't have considered that 4.3 billion addresses might not be enough, we now live in a world where a single person might have dozens of devices each of which can communicate on the internet. From this evolving reality comes the idea of the Internet of Things, a vision of a world in the not-too-distant future in which most or at least many of the devices in our home will want to communicate over a network, to allow us to learn about them or control them over the network. The first uses of this are home automation — for example, the ability to monitor or change the temperature in your home from your laptop at work — but the possibilities are endless. Why is this relevant to citizen science? If the "Internet of Things" is on its way, then we should expect to see on the market a range of inexpensive sensors that can communicate their observations using well defined standard protocols to publicly available services. These sensors and protocols can be used to collect, organize, and classify data for citizen science and related, evolving technical methods and techniques. They can also be adapted for use as part of mobile-based citizen science observation efforts.

The question, then, for our technology evaluation is: What kind of technology platform would enable researchers to engage with the vast potential community of citizen scientists and sensor data? It is amazing what is currently technically possible, but there is a large leap between what is technically possible and what is feasible for most science research efforts that work with limited resources. Our goal in this paper is to evaluate the existing technology platforms to see their potential for supporting these kinds of citizen science efforts, and sketch the general technical outline of what a future platform could look like.

In evaluating these platforms and in determining a possible path forward, we will use the following guiding principles:

1. **Open source**—Using open source components will reduce the initial development costs as well as the ongoing carrying costs of the project.
2. **Leverage existing efforts**—Where possible, the platform should leverage existing efforts where there is significant overlap. This both lowers development costs and helps build a community around existing development projects and citizen science efforts.
3. **Scalability**—The system must scale to many thousands of projects with millions of data points.

1 Sensr: Evaluating A Flexible Framework for Authoring Mobile Data-Collection Tools for Citizen Science. http://www.cs.cmu.edu/~sk1/publications/sensr_cscw.pdf

2 Estrin, D. (2010). Participatory Sensing: citizen science, scientific citizens, computational thinking (as referenced in 1)

4. **Cost of implementation** — The initial cost of implementation must be kept to a minimum, with architecture choices that allow flexibility and collaboration.
5. **Carrying cost over time**—The ongoing maintenance cost must be kept to a minimum, which may require that new citizen science projects can be added without direct intervention by programmers.

Data Form Design and Data Collection Tools

In many ways, the general challenge of data collection is one that most web sites on the internet tackle every day. A user enters information into a form on a web page and or a mobile application, the information is transmitted over the internet to a server which places it in some kind of data store, and the application makes it possible for the data to be retrieved later. But one of the reasons the challenge for citizen science is greater is that instead of building a single application for a single purpose, we want to build a more general platform that allows a researcher to essentially create their own website and mobile data collection tools for their own project, without additional software development. Scientists need a platform that will enable them to design forms for people to complete, include instructions on the protocols, create lists of valid answers and other definitions related to the definition of the collection process. Generally speaking, defining a form involves being able to define a set of questions that volunteers can answer. They may answer these questions by themselves (e.g. "What is the nearest intersection to where you are?") or a device they possess may answer the question behind the scenes (e.g. the GPS in their phone may report their current location). In many cases, the project organizers will want to define workflows: for example, first ask if the user has seen something ("Are there animals in the pond?") before asking for specific details or photos.

Another challenge is the many different kinds of data the scientist may want to collect. The physical location of the observations is likely to be critical, and collecting that position from a phone's GPS may be the most reliable way to do it. Photos, video, or audio may be important for validation or for other purposes. Qualitative answers may be much less useful analytically than requiring the user to select from a list of answers.

Beyond the specific need for this platform to allow citizen scientists to answer questions and submit data, there are a range of broader requirements that are worth considering. Part of the strength of a platform such as this would be the creation and cultivation of a community of committed citizen scientists interested in discovering and participating in new research projects. Otherwise, the process of distribution and outreach will need to be replicated over and over for each project, which is unlikely to be successful. If this platform is to leverage the wide availability of mobile phones, we will need a platform that has mobile clients for data entry for the major cellphone operating systems.

Frameworks for Building a Hosted Service

Of the technology platforms we evaluated, there are two general categories that define development and implementation: custom development and more generalized form-building software. Many of the projects are the result of custom software development with specialized tools. A custom software development approach makes sense when a project wants to build a small number of specialized and sophisticated citizen science surveys and where the project can employ software developers for ongoing development and maintenance. However, for a broad scale project designed to lower the cost and availability of citizen science this approach is less ideal. It does not enable the scientists to create the questionnaires and forms themselves, and it requires resource-intensive collaboration with software developers for each project.

An alternate approach would be a single software application that is designed to work "out of the box" for data collection. Ideally, this application would enable a scientist or researcher to easily define the data points (and questions) that they want to ask, the workflow and conditions under which different questions are asked, and then include the necessary software for users to answer these questions on the web or on mobile devices, as well as tools to monitor the results, find likely data collection errors or issues, and generate reports about the data collection. The following tools are aimed at one or more of these objectives. Though none of them fully satisfy all of these requirements, they have made significant progress, have existing communities of users and could be useful platforms to extend further.

Open Data Kit

Open Data Kit, a project that started at google.org and is now primarily developed at the University of Washington, is "an open-source suite of tools that helps organizations author, field, and manage mobile data collection solutions." [3] It is a member of the OpenRosa consortium, which is committed to open source and "standards based" tools for data collection. Most of the participating organizations developed tools that rely on the XForm standard from the World Wide Web Consortium (W3C), the major international standards body of the web. Open Data Kit is primarily focused on data collection on Android phones, and supports collecting GPS data, images, audio, and video, and automated upload to a server when a device has connectivity. It is currently deployed around the world, especially for mapping and data collection for social science and health research in locations that require support for disconnected data editing.

Perhaps Open Data Kit's greatest strength is that it makes it possible for someone who is not a programmer to create a sophisticated data collection form. Open Data Kit is composed a set of tools, including *Build*, a web based application that lets the user create a form by dragging a type of a question into a flow chart and then entering the properties of the question into a web form. When the web form has been completed, the form can be published directly to an *Aggregrate* instance that has been installed on the web. There are a variety of question types that can be added to an Open Data Kit form. Text questions are open-ended answers, like name. Numeric questions are any question answered by a number (like age), and date questions can capture a date (like the current date). Multiple choice questions can be specified, allowing either a single answer ("choose one") or multiple answers. Beyond those basic question types that you might expect on a paper form, there are some additional question types:

1. **Location (capturing where the user is, usually by GPS)**
2. **Media (audio, video, or photo)**
3. **Barcode** — A user can be given a printed sheet of possible answers and then can select one answer by scanning the barcode, which can be done with a phone's camera. For example, a user could be provided a print out with photos of birds with barcodes underneath.

The Build application also exposes some of Open Data Kit's power in a less user-friendly manner. Questions can be given conditions under which the questions are asked, grouped together, and separated into different branches that are followed under certain conditions. But how to specify these is not made clear in the Build application itself. While not necessary for all data collection efforts, these types of features become necessary in most data collection efforts that involve more than just a few questions.

Behind the scenes, ODK uses the XForm format to store the forms. This is an XML format, which is a format designed to be easy for the computer to parse but can still be edited by hand by someone familiar with the format. The XForm format is an official specification of a standards body called the W3C, and there are other data collection applications that also use the XForm format such as XFolite and JavaRosa but are based on older Java mobile platforms (like J2ME) that are no longer supported on modern smartphone platforms. However, Open Data Kit is part of the OpenRosa consortium of organizations that are committed to building an open infrastructure and open standards for mobile data collection — ideally commitment to these common standards will allow a citizen science platform to benefit from ongoing development and progress from various interoperable standards.

There is a range of more powerful form designers that require more technical knowledge (but not programming knowledge) such as Vellum, Kobo, and PurcForms. For example, in Kobo and PurcForms, the form designer creates an excel spreadsheet with the questions and configuration. While this does notrequire programming skills (unless one is trying to use more advanced features), it does require a good familiarity with spreadsheets and requires general technical aptitude and a willingness to learn the specific formats and structures of the form designer. It is not too difficult to use — for example, probably most scientists could learn how to build forms in a day or two — but it not a user-friendly process either. The Build tool is perhaps the closest to being user-friendly, but it is not currently at the standard of a user-friendly web application that anyone can immediately learn

3 http://opendatakit.org/about/

and use. With some additional development, however, the rough edges of Build could be hidden (advanced features turned off and hidden) or smoothed out with additional functionality and user components added.

Zooniverse / Scribe

As discussed in the citizen science case studies, Zooniverse is a successful platform for what the organizers call "data rich" citizen science. They have created a general, open source application called Scribe for transcribing (annotating with additional useful information) image files with additional information that researchers can use. Their platform has allowed almost 900,000 volunteers to observe and classify image, audio and text files, creating over 300 million observations. People are still much better at a wide range of classification and pattern matching than the most sophisticated machine learning technologies. The Zooniverse team developed their own custom transcription platform using the Ruby on Rails development system, which they later released to the open source community as software anyone can use and extend. There are other transcription software projects available (WikiSource, FromThePage, Scripto, Bentham Transcription Desk, and PyBossa) but a project with large numbers of users has very particular technical and usability requirements. The Zooniverse platform had to be extremely intuitive and inviting, as well as scale up and respond quickly to tremendous usage. A software developer, who familiarizes themselves with their codebase and knows the Ruby on Rails platform, can use their code base to create new workflows and new transcription application. It provides programming interfaces that simplify the creation of the workflows and manage the data collection process.

Each of the Zooniverse sites is visually attractive, has a great deal of custom design, and often features custom visualizations. This attention to visual details is part of Zooniverse's success, but the implication of this approach is a focus on creating a small number of high value and high volume citizen science projects. This has been a very successful approach for them, but requires that each project be able to bring significant financial or development resources to the project. As our research goal is to explore the potential for a platform that increases access to citizen science, this approach does not currently fulfill our requirements. However, it is an excellent model of what a potential platform might provide in terms of visual and user interface design. Another concern is that the current Zooniverse platform is not 100% open source, despite the fact that they have released components from it. In December 2013, the Zooniverse project announced that it had received $ 1.8 million Google Global Impact Award that will enable Zooniverse to re-build its platform to support a more flexible system and the creation of projects by researchers. Zooniverse has a great deal of experience creating visually attractive, engaging, and successful classification and transcription projects, and we look forward to seeing the new platform as it develops.

PyBossa

PyBossa is, similar to Zooniverse's Scribe, an open source platform that supports several types of classification tasks including: image transcription, image pattern recognition, sound pattern recognition, video transcription. Their own description of the project is as for a "open source platform for crowd-sourcing online (volunteer) assistance to perform tasks that require human cognition, knowledge or intelligence (e.g. image classification, transcription, information location etc)." Beyond image transcription and classification for citizen science, organizing this kind of volunteer analysis on existing data can be very broadly useful. For example, during environmental crises such as the 2010 earthquake in Haiti, volunteer effort was utilized to translate and categorize SMS messages. PyBossa, developed in conjunction with the Citizen Cyberscience Centre, is designed as a framework for developing data collection applications, not as an application or hosted service itself. It is, however, fully open source and has a very inviting website with good documentation.

Hosted Data Collection as a Service

Software as a Service (Saas) is a term that refers to the business model in which software is distributed as access to hosted service on the internet, often with an ongoing subscription fee. This can be an alternative to "shrink-wrapped" software that one would buy and install on one's own computer, or it can be a service that allows a client to have their own web application without hiring a programmer to create themselves — such as a hosting company that will allow you to set up a blog on their server for a monthly fee. Usually the process of signing up for such a service is automatic, as the software has been developed to allow someone to create a new site without manual steps. This sort of infrastructure would be ideal for a citizen science platform: a researcher could simply sign up, define the sorts of data collection they are interested in, and immediately have a hosted project and website and be connected to interested volunteers. This possibility led us to investigate whether Open Data Kit was available as a hosted service.

FormHub

There is a site, called FormHub, that will fulfill part of this need. The FormHub service enables the user to define a data collection questionnaire (as a spreadsheet) and will serve as the Collect server as part of Open Data Kit. It also features user accounts, the ability to download, browse, and visualize your data in a variety of formats. However, as discussed above, the current state of the project is such that simply FormHub or Open Data Kit as it stands would not be sufficient for a broadly accessible citizen science platform. But it is an open question whether creating a questionnaire through the creation of a spreadsheet is an easy enough process – it could be that with proper training and help resources it is a viable option. Like Open Data Kit, the code that powers FormHub is also available in the open source repository site, GitHub, and could be a critical component of a newly developed platform.[4]

EpiCollect

EpiCollect[5] is a functional, hosted data collection tool along similar lines of Open Data Kit created at Imperial College London. It has a number of excellent features, including Android and iPhone clients, as well as a web based form creation tool. Unfortunately, its overall functionality is significantly less than Open Data Kit – for example, not including conditional logic in form creation – and they have only released the Android client code as open source. The design and functionality of EpiCollect is not as advanced as all of the Open Data Kit components themselves – for example, the Build component of ODK has arguably better design and usability – but it does succeed at being simple and complete with the features it has. Its code is made public on Google Code and is set up to run on Google's App Engine platform. On top of the core functionality of data collection and synchronization with a central server, EpiCollect has functionality to view, chart and filter the collected data on Google Maps or Google Earth. It also will allow the data to be viewed on a map via mobile device using Google Maps. It would be candidate for extension for a citizen science platform if it was fully open source; however, while it claims to be an open project, this seems to be limited to releasing an Android mobile client and allowing the user to redirect data results to their own server.

PurcForms

Another notable application is PurcForms[6], which is a "browser based xforms form designer and runtime engine"[7]. It provides a user interface for creating XForms and a context for using them. It is powerful but not designed to be user-friendly. It is more appropriate for a very technical, sophisticated developer or user than as a component in a hosted service. It is, however, an open source technology that could be leveraged for some citizen science projects.

Sensr.org

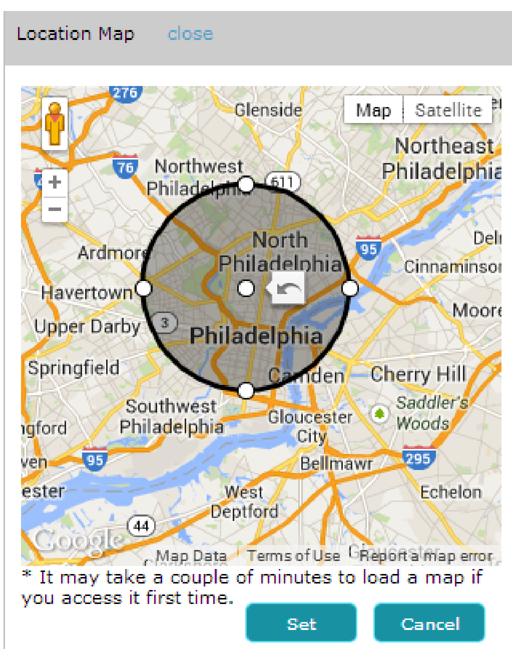
Sensr.org is the result of a student project at Carnegie Mellon University that produced an excellent paper with observations about mobile citizen science data collection efforts.

The paper identifies a number of critical requirements for a successful mobile-based citizen science platform. The Sensr.org project implements some of these features. First, it provides an easy-to-use hosted service — a user can simply log on and create a project. Second, it addresses the challenge of distribution — how do volunteers and researchers find each other — by providing a centralized web site that encourages participation along with

4 https://github.com/modilabs/formhub
5 http://www.epicollect.net
6 http://purcforms.appspot.com/FormDesigner.html
7 https://code.google.com/p/purcforms/

mobile client applications for iPhone and Android platforms that can access a variety of citizen science projects. Third, the developer argues that such a platform needs a simple to use, drag and drop interface to create the data collection form. In general, the concept is excellent — a drag and drop questionnaire creation tool is just what is needed and we agree with many of the insights from their paper.

When we initially tested the site in early 2013, the site and service did not seem to be "production ready" as we encountered a number of glitches when we attempted to log in. There were error messages, the first page ("My account") would not load, and a new account had someone else's data collection effort already included in it. It is also not an open source project that can be extended. However, when we tested again, later in 2013, the application was operating more effectively, and it was clear that it continues to be used. There were some interesting features. The application is aimed at supporting the creation of forms for gathering data on iOS devices. The form elements support photographs, text, and simple selectors as well as integration of a logo and header image. Data collections support map and table views, a mobile app simulation, discussion forum and data downloads. Of particular note are a form design process that supports the ability to specify a geographic are by picking a location a radius and a drag-and-drop form design process that is fairly intuitive. This platform does not have the breadth of functionality supported by Open Data Kit or other data collection form-building tools, but it a good assembly



of the full data collection, visualization and analysis workflow.

CrowdCrafting.org

Crowdcrafting.org is a new citizen science platform that is powered by PyBossa and has many excellent features for creating a citizen science community. It provides a common place where anyone can create a new project, import tasks for human categorization or transcription, and a place for interested volunteers to discover projects with which to engage. It has limitations – building a project requires some knowledge of Javascript and HTML, and it is focused primarily on leveraging human cognition for data classification and transcription of video, images, PDFs, rather than being a general data collection platform. However, it is thoughtfully designed and has a great deal of potential.

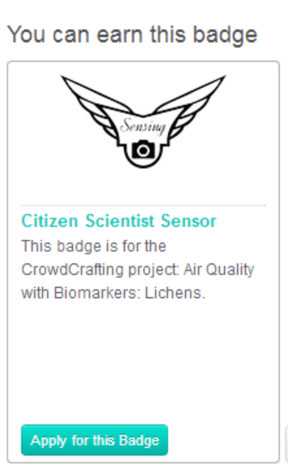
crowdcrafting

Further, several new features have recently been added that are very exciting, including support for an API, a badging system for recognizing contributors, and good integration of user statistics tracking. The features and design demonstrate a strong technical foundation and many opportunities for future growth. As mentioned above, the PyBossa platform is available under an open source license, is under active development, and the Crowdcrafting.org project is keeping pace with the full range of features that it provides. Crowdcrafting does not currently support many features of an open form-based data collection platform, but we believe platforms like Crowdcrafting.org, based on open source building blocks, are the critical next step for citizen science platforms.

Sensor Data Aggregation Platforms

There is both great promise and significant uncertainty in the quickly developing word of ubiquitous sensors, or the “Internet of Things” as it is often called. There has been a tectonic shift in consumer technology from personal computers to internet connected mobile devices, and that shift presents new opportunities for citizen science that have yet to be realized.

For a citizen science framework, the broad availability of consumer cellphones and smartphones with embedded sensors and internet connectivity make data collection theoretically possible at a vast scale that was inconceivable a few decades ago. One example is the Funf application developed at the MIT Media Lab, which is an application for Android phones that simply reports all of the data collected by the phone to a single database. Contemporary smartphones have a dizzying array of sensors: GPS, compass, accelerometer, microphone, high resolution camera, and even a gyroscope to determine the orientation of the phone in space, as well as the hardware necessary to communicate on and monitor the cellular and wireless internet spectrum. Combined with the promise of manual data collection — people answering questions about what they see or interviewing others — the phone becomes a powerful and ubiquitous tool. But the promise that truly expands the possible power of the platform is the possibility of connecting an array of external sensors to our smartphones, making it possible for millions of people to collect a wide variety of data about the world around them. One powerful example is the community of volunteers in Japan who used the Cosm service in 2011 after the nuclear accidents to create a data feed from Geiger counters monitoring radiation across the country.



However, in 2013, we are still just at the beginning of this quickly accelerating trend. We have identified some critical technology gaps that will need to be filled for a citizen science project to be successful at leveraging sensor driven data collection beyond the sensors already present in a cell phone. There are strong indications that these gaps will be filled in the near term, including initial products brought to market or funded by Kickstarter, and so the next 6-24 months will be a critical period to track these trends.

Technology Gap: Sensor Data Bridge

While inexpensive sensors are available, there is not a broadly commercially available hardware “bridge” that will transfer data from a range of sensors to a mobile device (using WiFi or Bluetooth) where it could be passed forward to a central server. This would need to be some sort of physical device that connects to a sensor (or has built in sensors). On the mobile device (a smartphone), there will need to be an application that can communicate with the data bridge and pass it along to a server. This is a critical component of any architecture in which inexpensive sensors are used with mobile devices for observation collection, unless the sensors are designed to connect directly with cellphones themselves.

While there is not a broadly available product that fills this need, there are many projects that are attempting to fill this gap and create reasons for optimism.

Technology Gap: Standard Protocols and Formats

As we explored the variety of projects coming to market or under research, the range of methods, protocols and approaches and the splintering of efforts – and the general chaos in this space at the moment – underlined for us the need for common protocols and formats. If there were agreed upon mechanisms for storing and transmitting sensor data, then the risks introduced by quickly evolving hardware platforms would be greatly diminished. As the situation currently stands, most sensor data bridge hardware is designed to communicate with specific software clients or data storage servers. We expect that as these platforms evolve standardized r protocols will evolve, but that has yet to happen. One method that comes close to a standard is the communication of a sensor result by a simple HTTP POST web request. If such a web request was standardized a little further, many different data aggregation services could record ongoing data – and it is very easy to add that capability to an existing system. That said, other systems are serving a more focused solution and are designed to communicate to a client application on a mobile device without being designed to forward that data to a central location. There are open standards for web-enabling sensors and sensor observations that have been developed by the Open Geospatial Consortium (OGC), including:

- SensorML, for modelling sensor systems and processing
- Observations and Measurements, for packaging observations
- Sensor Observation Service, web interface for accessing observations and alerts
- Web Notification Service, supporting asynchronous web notification (e.g. mail or SMS alerts)

Unfortunately, these standards are relatively verbose, XML-based standards that are more suitable (and fashionable) for large enterprise applications than small hobbyist projects or rapidly developed consumer products. Much simpler protocols (such as a basic HTTP POST) are more likely to be deployed in the consumer market and especially in the maker community who often prototypes these devices.

Kickstarter, Maker Culture, and Arduino

Several technology trends have accelerated development in the area around sensor data bridges. There is a growing community of technologists – some professional, some hobbyists – who are interested in small-scale engineering projects with a focus on invention and prototyping. Bringing together attitudes from punk DIY (do it yourself) culture and internet boom entrepreneurial spirit, the maker community has valued sharing information, whether it is open source code or "open source" hardware schematics that are freely shared. This trend has encouraged the development of new prototyping platforms, such as the Arduino microcontroller, which greatly simplifies and lowers the cost of prototyping an electronic device.

The interest in innovative electronic prototyping and the increased ease of prototyping has contributed to the explosive success of electronic device prototypes on Kickstarter. Kickstarter is a web site in which anyone can "back" – contribute funds – a creative project. Sometimes there is an explicit reward or product that the contributor will receive. For example, you can "back" a musician trying to raise funds for an album and, in return, receive the album once it is made, in effect buying the album before it is made. Open source hardware has been extremely successful on Kickstarter, possibly both because of interest in devices that have not yet been made and a general sense that open source projects are supporting a greater good or at least the advancement of technology that many are interested in seeing move forward.[8]

Sample Kickstarter Sensor Projects

Pebble

The Pebble E-Paper Watch, a smart watch which can communicate with a mobile device via Bluetooth, was the most funded project in Kickstarter history at the time (with over $10 million pledged), and has a number of embedded sensors, including an accelerometer, a compass, and a light meter. It is a sensor data bridge in the sense discussed above, as it can use the Bluetooth protocol to share information with a mobile phone and has a client. It does not, however, currently have clients designed to send that data to a central server. It was not designed to be extended with additional sensors, although it remains to be seen what additional functionality will be added.

Twine

Twine is a sensor data bridge (using WiFi) designed to be extended with additional external sensors. Twine's Kickstarter campaign by Supermechanical, a firm founded by designer-engineers from the MIT Media Lab, raised over $500,000 and has now shipped devices to its supporters. It is a 2.7" green square which includes an orientation sensor (the direction in which the square is currently facing) and a temperature sensor. The Twine that we tested also came with a magnetic sensor (for sensing whether a door was open or closed) and a light sensor. While the Twine is not

8 http://postscapes.com/internet-of-things-and-kickstarter

designed to be used with a mobile device, it could communicate to a central server over cellphone tethering (where the cellphone creates an ad-hoc WiFi network to share). But with that sort of use, the data would not have the GPS information from the phone. While Twine is designed to send its data back to Twine's own central data server, the Twine can be configured to send simple messages (HTTP POST) to a configurable internet address.

We were able to test a Twine device, and while we found that it might become an excellent device for enthusiasts in the future, it is still in an early testing phase, with many significant issues, at least anecdotally. The device we received did not work as expected, and the process of trying to understand why it was not working was a confusing process that revealed a number of bugs or issues yet to be resolved by the developers. The message boards on the Supermechanical web site also included many messages by unsatisfied supporters and customers. Supermechanical did acknowledge shipping a product still in alpha or beta testing (meaning there were still issues) but did so because they had promised a much earlier shipping date than they were able to accommodate.

Apart from these issues, there are many things that show great promise about Twine. It is extremely simple and pleasing to interact with, and it is easy to extend with additional sensors. Twine's interface has a beautiful design and, at first, the way you interact with the device is surprising and delightful. To change your Twine into setup mode, you just flip the twine over on its back. On its back, it tells you to go to twinesetup.com. Once there, there is a very clear step by step instruction on how to setup the device — and on step two, it asks you to connect to a new wireless network created by the Twine device itself. We hit snags at this step, but the overall process would next have had Twine sending its data to Supermechanical's data aggregation service, which provides a website that reports on the current status of your device. On their site, you can create rules and triggers – for example, if the temperature sensed goes beyond a certain temperature, you can have a text message or email message sent.

As currently set up, it would require custom software development to use Twine in a citizen science context – its current configuration process does not make it easy for an external service to access all of the sensor data. A trigger can involve an arbitrary web request and data can be downloaded, but this is an example of where standard protocols and sensor data methods would make integration of this technology easier. It is also important to remember that Twine is a data bridge designed primarily for home or stationary use, not for use with a mobile device.

Public Lab

Public Lab (or PLOTS) is a nonprofit organization creating a set of DIY and low cost sensors, software tools for processing and sharing sensor inputs, an open community of contributors, and a network of chapters around the world. They have had several successful efforts aimed at lowering the cost of environmental sensors and empowering the public to engage in environmental monitoring. Their hardware and DIY sensor kits include: kite and balloon mapping kits, a spectrometer attachment for smart phones, and a set of filters that can turn a digital camera into a near-infrared, multi-spectral camera. Public Lab has also created software projects that assemble the results of community data collection efforts. MapKnitter is an online web application for assembling aerial photo snapshots (captured with balloon mapping kits) into a seamless map that can be shared with others. This supports real-time aerial photography on a tiny budget. The Spectral Workbench is a database of spectra captured using their desktop or smart phone spectrometer attachments. The result is a database of materials that can be matched against spectra that have been captured by other members of the Public Lab community. Infragram is an online application for processing "infrablue" filtered images and processing them for plant health analysis. In addition to these projects, Public Lab has several ongoing research projects aimed at developing indoor air quality mapping tools, thermal photography, home endocrine disruptor testing, portable energy generator, stereo cameras and a potentiostat. Founded in 2010, the organization has demonstrated substantial success designing, developing and distributing sensors that leverage the computing power in phones and low cost digital cameras as well as building software for processing the data.

Mobile Sensor Platforms

While it is beyond the scope of this paper to do a full survey of sensor data bridges, there are a number of funded Kickstarter campaigns focused on the need for a mobile sensor platform or bridge. NODE[9] is a funded modular device that passes sensor data to Android and iPhone devices, which includes barometric pressure, elevation, wind speed, temperature, humidity sensors, as well as positional sensors (gyroscope, compass/magnetometer, and accelerometer) like in a smartphone. They plan to release gas sensor modules in the future. Sensordrone[10] is a similar product which received over $170 thousand dollars in Kickstarter funding and has already shipped an initial version. It advertises a client application for Android phones that can send data to Cosm, a data aggregation service (see Cosm section below). Both NODE and Sensordrone are candidates to become a standard sensor data bridge for mobile devices.

Sensor Data Bridges: Mobile Applications

FUNF

There are a variety of projects that focus simply on the sensors already included in a mobile phone. Funf[11] calls itself an "extensible sensing and data processing framework" and was developed at the MIT Media Lab. While the core concept is to provide an open source, reusable set of tools for collecting and uploading sensor data, the current implementation is primarily an Android application that collects data from the sensors included on an Android phone and make it possible to access or upload that data. An Android phone includes positional sensors such as a gyroscope, a compass (magnetometer), and an accelerometer, as well as a GPS. It also includes information about WiFi and cell tower connections. Funf also includes other data available on the phone, such as call logs. An excellent feature of Funf is its ability to create new Android applications for a particular use.

OpenRTMS

OpenRTMS[12] is an open source Android application that will collect data from an Arduino-based sensor with integrated Bluetooth communication through a module called Bluetooth Bee. While it aspires to be a more general open platform, it is currently an application that collects data from individual sensor integrated with an Arduino (or related) prototyping kit. It is still in the very early stages of development.

9 http://www.kickstarter.com/projects/108684420/node-a-modular-handheld-powerhouse-of-sensors

10 http://www.sensorcon.com/sensordrone/

11 http://funf.org/

12 http://openrtms.org

Sensor Data Aggregation

Once we have the capability for a sensor to communicate its measurements to the internet by means of a data bridge, the critical question becomes where to send and store that data. Once stored, we want to access, search, and analyze that data. This is the role of a sensor data aggregation service.

Proprietary Data Aggregation Services

There is a category of services coming to market that are essentially web accessible data warehouses for sensor data. The way they work is relatively simple. The service sets up what is called a "REST" or an "HTTP" API, which is a programming interface that enables one software program to talk to another using the same mechanisms that a web browser uses to request a web page. The assumption is that devices can either communicate on the internet themselves or can communicate their sensor readings to a bridge that will pass the data along or communicate it when a connection is available. Anecdotally, it appears that these services are primarily used, at this stage, for proprietary devices or applications or hobbyist projects — none of the services we examined listed more than a device or two accessible to a consumer. But there is great excitement around the "internet of things" and clearly an expectation these devices will become available soon.

Cosm

Cosm is perhaps the most well-known service in this category. Cosm began as Pachube (pronounced Patch bay) as a UK startup, and was purchased in 2011 by LogMeIn. It provides a service that allows data to be pushed from a sensor connected to the internet to Cosm, or Cosm will request data from a sensor every 15 minutes. Cosm provides a variety of data feeds and visualizations that can be included as part of a software or product developer's web-

Examples of Currently Available Sensor Bridge Hardware

Current Cost NetSmart meter[1]	http://currentcost.com	*Monitor energy usage; connects to Cosm*
Twine	http://supermechanical.com	*See above*
NODE	http://www.variabletech.com/products/	*See above*
Sensordrone	http://www.sensorcon.com/sensordrone/	*See above*
PASPORT AirLink 2	http://www.pasco.com/prodCatalog/PS/PS-2010_pasport-airlink-2/index.cfm	*Bluetooth bridge to connect proprietary sensors to SPARKvue analysis program*
Ninja Blocks	http://ninjablocks.com	*Open source, Kickstarter funding; sensor data bridges to the internet with mobile client applications for visualization and control*
knut	http://www.ampiric.com	*WiFi sensor bridge with embedded sensors; data collection and visualization application*
Beesper	http://www.beesper.com	*Wireless mesh sensor data bridge platform*
Good Night Lamp	http://goodnightlamp.com	*Internet connected lamps that automatically turn on associated lamps in other locations*

site. They also allow historical data to be analyzed or referenced, and real time alerts or triggers can be configured to respond to sensor data. Cosm has not released its software platform under an open source license and its use as a citizen science platform would require relying on the more limited free services it provides or hiring Cosm to provide services.

Carriots

Another example of a platform in this category is Carriots, which is short for "Carry the Internet Of Things forward". It enables a user to register up to ten devices which can send data to their platform and then script behaviors. However, they do require programming skills (in a programming language called Groovy) and the details on their site are vague. They list only a few devices, like Arduino and RaspberryPi, but they (like others) are assuming that these devices will become available to consumers soon. Essentially, this is a service for service developers — it saves the application developer from having to create a web service that accepts data from a sensor, stores and reports on that data. It is very much Cosm in that regard. It would serve an application developer who can program in Groovy and wants to sell a consumer home automation system, but given its functionality and its early stage of development (still beta) it is not clear that it would be a good partner for a citizen science platform.

Open Source Data Aggregation Servers

There are a number of options for open source data aggregation servers, although none of them as of the time of this writing are ideal solutions. When evaluating an open source platform, it is important to evaluate the community and current development velocity of a project. In the best case, it is being actively developed and used by a community of users. In the worst case, it was developed by a single developer or two and hasn't been updated (or obviously used) in years.

Because of constantly shifting technological requirements and a stream of software updates to key components, most software for the web needs some level of constant maintenance, and a healthy open source community is the best assurance that a project will have the necessary attention and updates it requires. A healthy community also increases the likelihood that support will be available if questions or issues arise. None of the three projects we evaluated had a substantial community around them, which may mean that custom development or partnership with a proprietary service will be necessary for a citizen science platform.

ThingSpeak

ThingSpeak is an open source project from the company ioBridge, which makes and sells sensor and control devices that can communicate on the internet. The ioBridge IO-204 Monitor and Control Module is a key product: a data and communication bridge with an ethernet jack (for hard-wired ethernet cables) to send sensor data to the internet and receive commands. Focused largely on the home automation market, ioBridge also sells a suite of sensors that can be connected to their data bridge. IoBridge provides a hosted data aggregation and altert system. When you purchase an ioBridge device, you can log in to their hosted web application and view, access, and analyze your data, create widgets for other web-sites, or add email alerts that should be sent out when your sensors are triggered.

IoBridge has released part of the software behind their hosted data aggregation on the social open-source coding site GitHub.[13] It is listed as an open source "Internet of Things" application and API "to store and retrieve data from things" over the internet. It has a simple HTTP (web-based) interface for other applications to request data. The application description reads: "With ThingSpeak, you can create sensor logging applications, location tracking applications, and a social network of things with status updates."

The application is based on the common Ruby on Rails web development framework, and uses the open source database system MySQL. There is a separate "fork" (an alternate version) of ThingSpeak which adds some features for scalability, storing data in a memory-based storage system called Memcached and providing the ability to "rate limit", which means to only allow sensors to update at a particular frequency.

ThingSpeak is easy to set up and install, and even provides some basic visualizations. It includes an API that enables for some basic numeric data processing, such as time scaling and averaging. They also share some related projects through GitHub, including an example of how to program an Arduino prototyping board to communicate with the ThingSpeak service.

While ThingSpeak does not have all of the functionality that would be desired in a data aggregation platform, it is certainly an appealing choice to extend. The application is relatively well documented and simple to understand if you are familiar with Ruby on Rails applications. The biggest reservation we would have about recommending it is that there does not seem to be much active development on it: there have only been a handful of updates in the past few years. There does not seem to be much community around it, as well. It's primarily designed for users of ioBridge products, and not as a general purpose platform. It also does not include any data collection capabilities through mobile devices.

13 https://github.com/iobridge/ThingSpeak

Nimbits

Nimbits is an open source sensor data aggregation system designed to function on the Google App Engine infrastructure.[14] Its code is available on GitHub[15] and documentation is available on the main product site. Like ThinkSpeak, Nimbits offers a hosted service that is powered by their open source product. The site advertises the "zero lock in" advantage of using a hosted platform that is based on open source software: in theory, you could set up the service yourself.

Nimbits is written in Java and uses a variety of Java technologies to implement its backend services. It is relatively full-featured, with a variety of methods for sending data over the web and a variety of export and alert functionality. It includes some functionality for graphing and charting your data through the Google Chart API. It provides an HTTP API for adding data points. The data points are specific to Nimbits, but the API exposes a significant amount of functionality.

From the Nimbits' Github site, we can see it is primarily the work of a single person without a developer community around the project. There is some activity (perhaps 10-20 messages a month) on the Nimbits support mailing list, which demonstrates that it has users, but it is not the focus of much community energy in terms of new contributions. Like ThingSpeak, it is a single company's technological solution that has been open-sourced but has yet to evolve into a community project. Its user interface has not received much design and is not particularly user-friendly, but it is certainly functional. If Java was the preferred language of choice

14 http://www.nimbits.com

15 https://github.com/bsautner/com.nimbits

of a development effort, Nimbits would be a reasonable choice to extend.

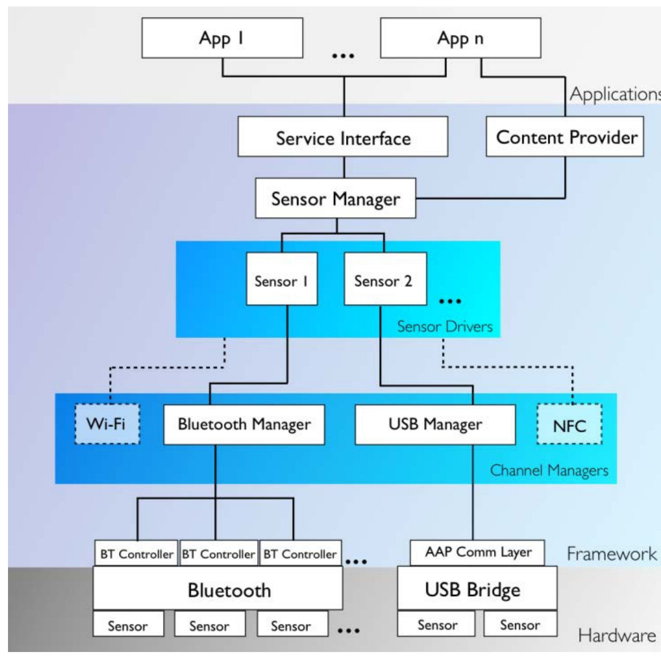
Open Data Kit Sensor

Open Data Kit Sensors is an academic project described in a 2012 paper that attempts to "simplify the interface between external sensors and consumer Android devices." [16] It describes a software architecture that abstracts over communication methods (Bluetooth, USB) and sensors to create reusable components for implementing sensor drivers and applications that can utilize the sensors. It is essentially the software architecture for a sensor data bridge as described above, with the Android phone serving as the intermediate hardware used in the other sensor data bridges described above. It simplifies the addition of new sensors to the data bridge and creates an overarching abstraction for development and communication.

The research described in the paper informs sensor bridge development more than application level development. However, it does indicate the feasibility of extending Open Data Kit directly to capture sensor data which could provide a single data aggregation platform for citizen science efforts. It would require new, custom development of Open Data Kit to add this support for sensor data aggregation on the web, but the existing codebase and user interface controls would certainly be an asset to this development. No development was done on using iOS (iPhones, iPads) for sensor data collection, which would also be additional work.

Evaluation of Visualization Frameworks

Data visualization and geographic data visualization are critical tools for data analysis, and a hosted platform would be able to provide a set of tools for quick analysis. For example, the platform could allow R scripts (described below) or geographic analysis to be executed against collected data by researchers, allowing near real time analysis as data is collected. But, as many researchers will have their own analytical tools of choice, it is likely that many will simply want excellent export capabilities. But visualizations have a different and very important role in citizen science: engagement and education. Data visualization gives volunteers a sense of the work that they are engaged in, a sense of their role in the overall research project, and – when the visualizations are visually impactful – a sense of awe surrounding the subject matter. A strong example is the collaboration between CartoDB and Zooniverse, in which the 1 million WWI Royal Navy locations transcribed by citizen scientists are animated on a map.[17]



Open Data Kit Sensor Architecture

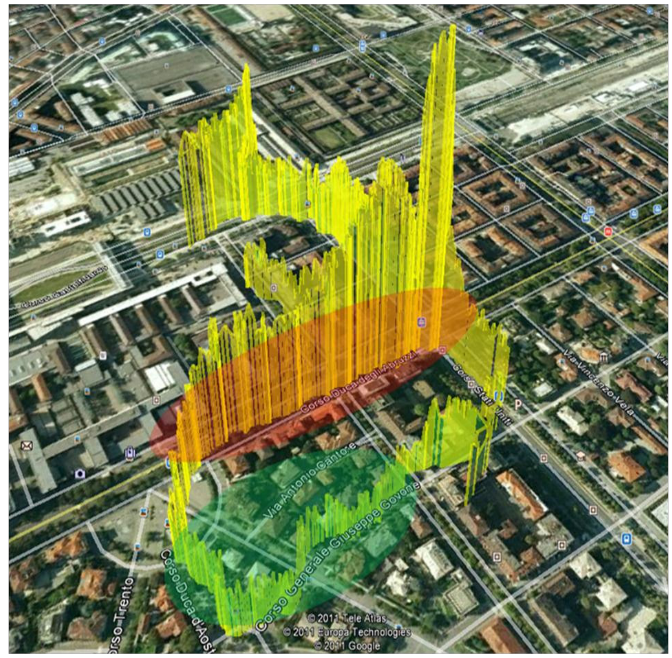
A data visualization can give a citizen scientist a sense of the observations and contribution they are making to a given project. For example, we have included a visualization from the EveryAware SensorBox "electronic nose", a mobile sensor array equipped with commercially available electrochemical and metal oxide gas sensors, indicators of pollution caused by traffic. The geographic visualization provides an easy means of initial analysis.

Creating and supporting a community of engaged volunteers is a critical aspect of a citizen science platform. This means involving, educating, and learning from the citizen scientists must be a key feature, and data visualization allows them to engage directly with the knowledge they are collecting and sharing.

16 http://homes.cs.washington.edu/~wrb/ODKSensorsMobiSys2012.pdf

17 http://cartodb.github.com/oldweather_wwi/

In a collaboration between CartoDB and Zooniverse, this animated map shows 1 million WWI Royal Navy locations transcribed by citizen scientists.



Visualization of SensorBox air quality measurements, showing a clear difference between the pedestrian area (green area) and the busy road (red area).

We have evaluated a number of visualization toolkits that could potentially be applied to a future citizen science data collection platform.

ArcGIS

Esri is one of the world's leading providers of geographic information systems, with over 40 percent of the market share of the GIS market[18]. Their ArcGIS suite has an exceptionally wide and powerful set of tools for geographic information analysis and display, and can be used to dynamically create and provide visualization and applications for the web. While a critical tool for many professional GIS analysts and departments, it is a proprietary software suite, so it does not meet one the requirements laid out for the citizen science platform. The new ArcGIS Online does provide new web-based cloud computation functionality and elements of the platform could potentially be integrated into a citizen science platform.

Pros: Powerful, enterprise quality geoprocessing engine
Cons: Proprietary, scalability

18 http://apb.directionsmag.com/entry/esri-has-40-of-gis-market-share/215188)

GeoTrellis

GeoTrellis is a geospatial analysis engine designed to power high performance and high volume data visualizations on the web. It provides a range of geospatial data visualization and analysis operations that can be integrated into web based applications. Much of the focus of GeoTrellis is bringing geospatial analysis that traditionally has been done in desktop applications to dynamic web applications. It provides a software development kit for building geospatial operations in the Scala programming language. However, it is a complex framework, does not currently have a user interface and would require integration to be useful for a citizen science data collection platform. Nonetheless, it would provide a mechanism for processing the large data sets that we expect would be submitted to a generalized citizen science data collection platform.

Pros: Open source, high performance, GIS capabilities
Cons: Software toolkit, not application

GeoServer

GeoServer is an open-source map rendering server that can be used to publish maps on the web. It has been designed for interoperability and implements a wide range of Open Geospatial Consortium (OGC) standards for sharing geospatial information on the web. It read a wide variety of data sources, and has a web administration interface for adding and styling layers. Some of the OGC standards, such as the Styled Layer Description, can be difficult to work with from a user or developer perspective, but GeoServer is an extremely useful and versatile tool for serving up geographic data on the web. It is also not the only open source map server, and credible alternatives exist in the form of projects like Mapnik and MapServer.

Pros: Powerful, open source map rendering engine
Cons: XML heavy standards can be difficult to work with

D3

D3 is a powerful visualization library for the web, created by Mike Bostock, Graphics Editor at the New York Times. It is a JavaScript library designed to operate on dynamic data, and has a wide variety of innovative and appealing visualizations, including a range of charts, graphics, animations, and a recently added suite of geospatial visualization tools. It is an excellent toolkit and used by some of the projects described in this paper, such as PyBossa, use the d3.js visualization library for the display of results.

Pros: Can create beautiful, dynamic visualizations in an infographic style
Cons: Development learning curve; all data must be sent to browser, limiting the size of the data sets

Processing

Processing is an open source programming environment for creating animations and other illustrations. It is written in Java which makes it more appropriate for generating static images on the server side (or desktop applications), but there is a Processing.JS project that allows "sketches" (as Processing programs are called) to be run in the browser.

Pros: Large community of digital artists and example code
Cons: JavaScript implementation has performance issues; underlying Java platform has limitations

R Project

R is a programming language for statistical computing and graphics. It is used widely for statistical programming, and can create charts and graphs. It is known for its powerful set of statistical, mathematical and computing techniques. However, it is less known for the beauty of its visualizations. R would make a great add on for researchers who are familiar with R and want to run R scripts on their data, but would make

Pros: Powerful statistical analysis toolkit
Cons: Functional but not beautiful visualizations

Conclusion

At the nexus of the maturing wave of mobile technologies and the still immature movement of the "Internet of Things" lies a very exciting but as of yet unrealized opportunity for a more general citizen science platform. If the proper infrastructure was created and current trends were to continue, the wide availability of powerful mobile phones and the availability of inexpensive sensors that could communicate across them would become a vast resource for researchers who need observations of the world that a new community of citizen scientists could collect and share. To be sure, there are some significant unknowns. While the first consumer hardware platforms for sensor communication are coming to market with great enthusiasm from a community of early adopters, product developers, and hobbyist engineers, the hardware and software platforms that will create the "Internet of Things" have not yet been built — or in the few cases where components exists, they have not yet matured or become clear standards. However, at the very least, it is clear that the products already on the market have the potential to develop, mature, and build communities that rely upon them. If the "Internet of Things" is, indeed, poised for a rapid leap forward, a platform that begins development soon would be poised to participate in the transformation and leverage the excitement around the new technology to create new communities and expand existing communities of citizen scientists.

Our assessment of current technology platforms confirmed that there is not an existing platform that makes it possible for researchers and citizen scientists to collect observations through answered questions and sensor measurements, despite significant interest from researchers and interested participants. While there is not an existing platform that can be directly used, we believe that a number of existing projects and technologies could be leveraged to create the open source platform that has been envisioned.

Open Data Kit, along with the XForm and OpenRosa standards, appears to be a good candidate for a data collection platform to extend. An alternative approach would leverage the PyBossa project and its growing CrowdCrafting.org service by extending the software components and platform beyond its micro-tasking and data classification origins to encompass a more general set of data collection tools. It is unclear what form the new, more general Zooniverse platform will take, but this also remains a contender. Finally, the National Geographic Fieldscope project also has a more general data collection platform under development. Open Data Kit, PyBossa/Crowdcrafting, Zooniverse and Fieldscope all have existing communities of users. However, in each case, considerable additional work would need to be done. Open Data Kit would need to develop a community platform for project discovery and additional development work on the existing Android and iOS mobile applications. The FormHub codebase could also potentially serve as part of the necessary development work to turn Open Data Kit into a hosted "software as a service" as described above. The PyBossa and Crowdcrafting projects have a very strong start and appear to be both thoughtfully designed and based on learning from other citizen science projects.

In terms of a sensor data aggregation platform, it is too early to make a clear assessment as to how standards and services will progress in the next 12-24 months. However, the options for development are relatively clear. Given that the availability of sensors will be a key requirement, this platform would either need to sell its own sensors and compatible sensor bridge with its own data aggregation platform — which is ambitious — or take a more flexible approach, in which a sensor data synchronization system will interface with other, external sensor aggregation platforms and associate that data with data collected through mobile applications. If a custom service needed to be developed, then we would recommend extending Open Data Kit or PyBossa platforms to accept sensor data (if budget allowed) or extending the ThingSpeak platform (especially because it is compatible with sensors currently on the market), and, ideally, a citizen science data collection platform would engage with an existing open source project as opposed to attempting to start a new sensor aggregation project.

There are also significant opportunities for stronger collaboration between existing online data platforms and sensor development efforts. A collaboration between CrowdCrafting, Public Lab, Cosm, and SciStarter, for example, would be a very exciting partnership. Crowdcrafting has developed a platform (based on the open source PyBossa software) for collecting data from humans. Public Labs has successfully invented several low cast hardware sensors for a broad range of applications. Cosm is able to accept high volume streams of sensor data. SciStarter has a large community of potential contributors.

Returning to the question with which we opened – what kind of technology platform would enable researchers to engage with the vast potential community of citizen scientist and sensor data? We are advocating for a platform that has the following attributes:

- Open source
- Scalable from small projects to large
- New project don't require a programmer
- Supports human data collection as well as sensor data feeds
- Supports data visualization
- Has a modest ongoing cost

Building this new citizen science platform will face many challenges, and we believe there are significant missing gaps. As mentioned above, we do not believe any current platform captures all of the necessary elements, but there are several potential contenders, including: Open Data Kit, Zooniverse, Crowdcrafting and Fieldscope. There is also room for new services to step in, as well.

For projects seeking an online data collection platform today, we are unable to recommend a single platform, but we would suggest asking the following questions as you consider the various options:

- What is the timing of your data collection efforts? The technology is moving rapidly and new capabilities will be available six months from now that are not possible today.
- What technology skills does your team have? Do you have access to programmers that could potentially adapt a toolkit like PyBossa or do you need this to be simple enough that a relatively non-technical person could set up the project?
- How long will the data collection process last? If it will be indefinite, how much can you afford in terms of ongoing costs?
- How important is privacy of the data? Will children be involved in collecting data or will you require other special privacy measures?
- Do you need data visualization or statistical analysis tools as part of the platform or will you do this yourself?
- Would you benefit from access to an existing pool of potential data contributors or will you recruit your own contributors?
- Will participants need specific sensors to gather data? Will participants need a smart phone or other device to collect data in the field or can they enter the data from a laptop or PC?
- Will your contributors need a specific set of skills in order to participate? Will you help them acquire those skills or do they need to have them in advance?

The citizen science data collection ecosystem is evolving and growing rapidly. It is an exciting time, and we look forward to seeing the new discoveries and innovations that will be created in the coming years.

Acknowledgements

This is Part 2 of a four part report. The full report includes the following:

- Part 1: Data Collection Platform Evaluation
- Part 2: Technology Evaluation
- Appendix A: Wireframe Designs
- Appendix B: Cloud Computing Performance Testing

This report was funded in part by a grant from the Alfred P. Sloan Foundation.

We respectfully acknowledge the individuals and projects that contributed to its development of this study; we could not have completed this document without their invaluable expertise and input. We are particularly grateful for the valuable insights provided by both the project representatives we interviewed and the input from the Citizen Science Community Forum.

These recommendations are meant to both celebrate their accomplishments and acknowledge the need for continued growth and improvement that will benefit everyone. To that end, we welcome additional comments from the citizen science community such that we may further refine this vision and move it forward toward reality.