

ICT Services for open and citizen science

Mikołaj Morzy

Received: 13 October 2013 / Revised: 22 March 2014 /
Accepted: 15 July 2014 / Published online: 23 August 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract Ideas of open access, open data and open science are transforming the world of scientific inquiry as we speak. Every day thousands of ordinary citizens are engaging in data collection and data processing, giving rise to the new field of citizen science. Never before has the technology enabled scientists to reach out to such vast numbers of collaborators and show their work to the public. From pattern recognition in Hubble space telescope images of distant galaxies to field observations of migration patterns of birds in the rural areas of United States, the possibilities are countless. Certainly this new trend poses important problems and challenges, but it is also obvious that wide acceptance of citizen science can lead not only to great scientific results, but to the popularization of scientific method among the public. In the paper we examine the current state of citizen science, we outline some of the most interesting and difficult challenges in leading scientific projects on such scale, and we present typologies of citizen science projects. We also provide a survey of ICT tools available for citizen science projects.

Keywords Open science · Citizen science · Web 2.0

1 Introduction

The traditional model of scientific research involves only professionals on all stages of the process. From the formulation of the research question, through study design and data collection phases, to data analysis and interpretation, all steps are limited to science professionals and the general public may be passively involved only in the dissemination of results. However, this model is highly ineffective and costly. All stages of the process may be greatly enhanced by widening of the group of people involved. One may easily imagine the

Research supported by the National Science Centre grant 2011/03/B/ST6/01563

M. Morzy (✉)
Institute of Computing Science, Poznan University of Technology Poznan, Poland
e-mail: Mikolaj.Morzy@put.poznan.pl

benefits of distributing the data collection process among the network of independent volunteers who may be able to collect amounts of data of orders of magnitude larger. Amateurs may be involved in data analysis (in particular, when the analysis requires tasks that do not yield well to automation), but their feedback may be useful even during data interpretation and conclusion inference phases.

The history of volunteer participation in scientific projects dates back to the XIX century, but only recently, due to the advances in information technology field, the broad employment of amateurs in the scientific process has become a viable option to many fields of research. There are two major scientific movements fueled by the social network revolution that we are witnessing today. The first movement aims at making the scientific process transparent and accessible to public, both amateurs and trained professionals. This movement is often referred to as the *open science* and it advocates steps such as publishing results in open access venues, making datasets publicly available for easy replication of results, practicing open notebook science (i.e., publishing not only final results, but making all internal steps of the project publicly visible), and putting much effort into communication of science to the public. However, open science does not explicitly promote the involvement of amateurs in the process. This idea is the cornerstone of the second movement, called the *citizen science*. Following [18] we define citizen science as the form of collaboration involving active engagement of members of the public in scientific projects which address real world problems. We distinguish citizen science from other modes of public involvement in scientific process namely by the active engagement. Projects which require passive engagement, for instance, sharing resources to increase the computational power (e.g., SETI@Home), do not comprise citizen science. Also, we exclude from consideration projects in which members of the public are asked to participate in a study, e.g. by filling surveys and providing personal data, but do not engage actively in the scientific process. In this paper we focus mainly on citizen science, but many of our remarks can be easily applied to open science as well.

As previous studies show [15], citizen science can lead to massive scale experiments resulting in high quality data, interesting insights, valid scientific finding and innovation, provided that projects are carefully planned and designed. The availability of modern information and communication technology (ICT) tools and services allows broad audiences to engage in scientific process, usually by providing collected data using agreed upon protocols, or by performing tasks, such as classification, recognition or computation, that require uniquely human capabilities. We want to analyze citizen science from the point of view of modern services and tools offered by ICT, and in particular, we want to establish the minimum set of technical requirements necessary to support a wide range citizen science project. Our attention is focused on Web 2.0 stack of technologies and how these services can be accommodated for citizen science projects.

The structure of the paper is the following. In Section 2 we present a brief history of citizen science and we discuss some closely related paradigms. Section 3 presents the typology of citizen science models and projects. We introduce main challenges and problems facing citizen science projects in Section 4 and we describe relevant ICT services, tools and applications in Section 5. We conclude the paper with a summary in Section 6.

2 History of citizen science

Using volunteer participation in conducting scientific research has been a popular method of inquiry. Historically, amateurs have been employed mostly in the fields of biology and

ecology to assist scientists in large scale research, although other fields of science also benefited from mass participation of amateurs. Often cited as one of the first successful examples of citizen science is the 1874 project of measuring the transit of the planet Venus [14]. The project has been funded by British government and supported by the Admiralty, whose vast personnel has been used to gather and collect data from multiple positions on the globe simultaneously. Another example of a large scale project is the Christmas Bird Count [1], an annual event initiated in 1900 by a single person, Frank Chapman of the American Museum of Natural History, and continued until today under the auspices of the National Audubon Society. Due to the large number of skilled amateurs and willing volunteers, the field of ornithology has enjoyed many successful projects involving the public. Notable examples include the Breeding Bird Survey [5] started in 1966 and nest record counting, initiated by the Cornell Lab of Ornithology and continued until today as one of the most prominent online citizen science projects Neighborhood NestWatch [9]. Today, the Cornell Lab of Ornithology is supervising over 600 citizen science projects. Many projects revolve around organismal monitoring, e.g., the Invasive Tracers program¹ uses citizens to count the number of invasive crab species in coastal areas of the USA, the Great Sunflower Project² aims at monitoring the diversity of bee and other pollinator species, and the MykoWeb project³ collects observations on the instances of fungi species. Some projects, such as the Citizen Weather Observer Program⁴, focus on environmental monitoring rather than observing living organisms. Finally, there is also a large set of non-ecological projects, ranging from astronomy (GalaxyZoo⁵), through microbiology (Foldit⁶), to machine learning (GWAP⁷). A comprehensive list of past and current citizen science projects is available at Wikipedia⁸.

Citizen science is deeply rooted in online collaboration communities and can be easily mistaken with other forms of collaboration. For instance, citizen science shares many features with peer production and crowd-sourcing. These two latter names refer to a broad set of models that rely on calls for contributions from an open, large, and undefined network of people [4, 10]. The models of peer production and crowd-sourcing, as well as citizen science models, use open data and open participation (although sometimes an initial training of participants is required to guarantee a certain level of service quality). Most of projects are mediated using information and communication technologies and assume long-term commitment from participants. However, there are notable differences between citizen science models and typical online communities analyzed in the literature [8]. Contrary to most online collaboration projects, citizen science projects do not exhibit self-organization nature, but are strictly hierarchical. The structure of tasks performed within citizen science projects is very similar to peer production, but the hierarchical structure of the entire project introduces very different contexts of authority, management, and responsibilities. Probably, the most similar area to citizen science are networks of practice [7], although motivations are quite different.

¹<http://massbay.mit.edu/exoticspecies/crabs>

²<http://www.greatsunflower.org/en/about-project>

³<http://www.mykoweb.com/articles/CitizenScience.html>

⁴<http://www.wxqa.com>

⁵<http://www.galaxyzoo.org>

⁶<http://foldit.com/portal>

⁷<http://www.gwap.com/gwap>

⁸http://en.wikipedia.org/wiki/List_of_citizen_science_projects

3 Citizen science models

Before presenting the typology of citizen science models and projects, it is worthwhile to consider the general process of scientific research. Given a set of data, various techniques such as data mining, statistical model testing, or visual analysis can be employed to detect patterns and trends in available data. Based on this step a research hypothesis is formulated, which can be used to make predictions. Existing data is then used to statistically validate and test predictions. The results of statistical tests fuel the inference which modifies, validates or rejects the tested hypothesis. All steps of the scientific process, namely: question formulation, study design, data collection, data analysis, inference, and dissemination of results is performed by trained professionals.

The first model of citizen science that we want to discuss is the so-called *science shop* model [11, 16], also known as the *scientific consulting research* model. This model has been developed in the Netherlands in the late 70s and has been successfully implemented in several projects. According to the science shop model, the problem is formulated by community groups of amateurs, and an established institution of science (e.g. a university) serves as the consultant which verifies and modifies the scientific process, analyzes the collected data, but often it is the university which is responsible for data collection and study design. Usually, projects executed as science shops are limited to small geographical locations and aim at solving local community problems, research outcomes of these projects are conventional, and additional emphasis is put on scientific education of the public.

The second model is the *participatory action research model* [12, 17]. In this model volunteer amateurs work alongside professionals on all steps of the science process. The distinctive characteristics of this model include the relevance of research subject to problems and needs of the local community, promotion of the collaboration between scientists and local communities, and strong emphasis on educational aspects of science associated with the increase of local community awareness. Prior to the emergence of social networks and online collaboration tools most citizen science projects could be assigned to this category. However, modern citizen science projects operate on much larger scales than it was possible for participatory action research, and the iterative and direct collaboration with scientists is either limited, or virtualized. In the typical scenario amateurs participating in citizen science projects are used mainly during the data collection step, and sometimes they contribute to data analysis and interpretation. Contrary to the participatory action research model, in citizen science research model geographical scope of the project is broad, and research goals are usually set much higher.

Finally, a new model of *adaptive citizen science research* has been introduced in [6]. This model advocates shifting management responsibilities to individual participants, with an additional iterative feedback loop. In addition, amateur participants cooperate with scientists in analyzing and interpreting results, which in turn may influence the research hypothesis being tested. This model is best suited for conservation projects in residential landscapes, where the cooperation of the local public is of utmost importance. Since urban areas quickly overtake large portions of the habitat, the development of a framework for conducting large scale studies or answering particular local problems is urgently needed. Adaptive citizen science research model claims to provide necessary tools for such framework.

Although citizen science projects may vastly differ in scope, domain, methodology, and requirements, it is possible to construct a typology of such projects. Besides providing additional insight into the mechanics of citizen science projects, the typology allows to extract

most important features of these projects, which in turn informs of technological requirements of these projects. Over the years several typologies have been proposed and below we present two proposals. The first proposal has been introduced in the report of the Center for the Advancement of Informal Science Education [2]. This typology differentiates citizen science projects based on the criterion of public participation [19]. The authors consider the following steps of a scientific research: defining the question, gathering of information, development of hypothesis, designing a study, collecting data, analyzing samples, analyzing data, interpreting data, drawing conclusions, disseminating results, discussing results, and asking new questions. Depending on the degree to which amateurs are involved at each step, the authors divide citizen science projects into three classes (Figure 1).

- *contributory*: amateurs are involved primarily in data collection, with sporadic participation in data analysis and result dissemination,
- *collaborative*: amateurs are used to collect and analyze the data, but their feedback may be used during the study design step as well, in addition general public is partially involved in interpreting data, drawing conclusions and disseminating results,
- *co-created*: projects belonging to this class are characterized by the full involvement of public at all steps of the scientific research.

The above classification, however useful and providing useful hints about the distinctive features of citizen science projects, narrows the analysis by focusing exclusively on the degree of public participation in the process. There are several other factors that need to be taken into account when classifying these projects: organizational characteristics of the project, technical aspects of the project, management of the project, or infrastructure employed by project participants. Thus, a much more thorough survey of existing citizen

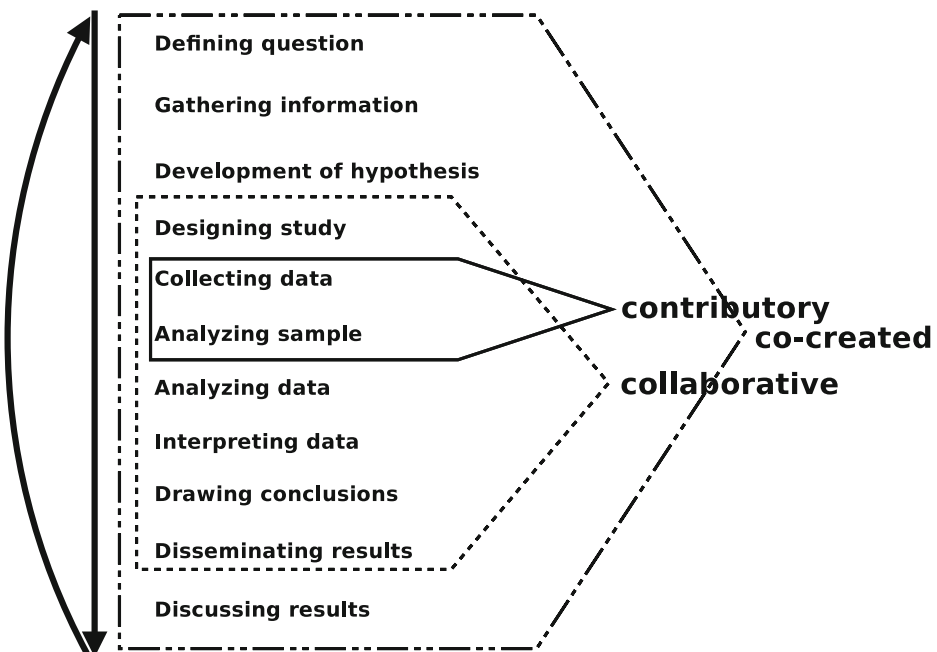


Figure 1 Typology 1 of citizen science projects

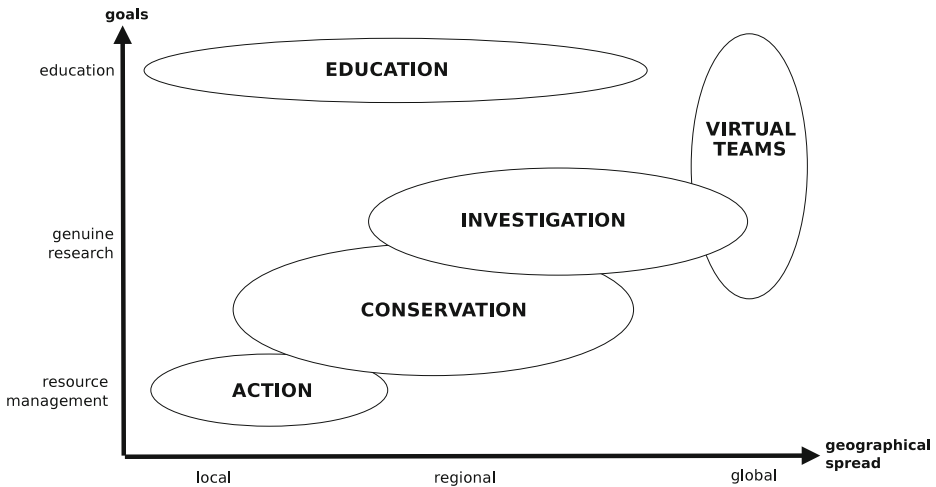


Figure 2 Typology 2 of citizen science projects

science projects has been performed, taking into consideration over 80 different facets of projects. Next, a qualitative clustering of data points has been performed and its results were reported in [18]. The authors identified five major classes of projects (Figure 2). To simplify the picture and outline similarities between these types of projects we have chosen to display these classes using only two dimensions: the degree of geographical reach of the project and the main goal of the project. However, it should be stressed that the original typology was constructed by clustering more than 80 variables representing different facets of projects.

- *Action*: These projects aim at addressing local problems and concerns using scientific method, with participatory action research model in their core. A notable difference is the fact that action projects usually use bottom-up organization and management. Projects have long-term aims and also assume long-term commitment and participation of local members. Also, projects are planned by amateurs rather than scientists, who serve as consultants or help in tasks requiring more rigorous research procedures. Projects have very concrete goals rooted in local physical world and their scientific value is mediocre. Strong localization of projects means that these projects do not scale well and suffer from the lack of sustainability (mostly due to the fact that it is difficult to provide continuous financial support for these initiatives). From our point of view the most interesting are technical issues. Action projects make little use of advanced ICT services, often their use of technology is limited to the creation of simple websites with announcements, calendars of events and general descriptions. Data collection does not introduce technological challenges as the localized nature of action projects allows for easy data sharing during project meetings, and even the most rudimentary technologies can be used for this purpose. A good example of action project is “ReClam the Bay”⁹, a project whose aim is to repopulate the Barnegat Bay in New Jersey with clams and oysters.

⁹<http://www.reclamthebay.org>

- *Conservation*: This class of citizen science projects encompasses a wide variety of projects whose primary goal is natural resource management, protection, and conservation in the area of ecology. These projects share several characteristics with action projects, for instance these are highly localized projects with volunteers contributing mostly to data collection step. Differently from action projects, conservation projects put more stress on education value of participating in science endeavors, these projects operate at slightly larger scales (regional, state) and in cooperation with state and federal agencies. This tight cooperation with established authorities simplifies management and sustainability. Also, the scientific quality of projects is usually higher with universities and individual researchers involved at several steps of the project. Since the results of conservation projects are often used to inform regional decision-making, the formal requirements are more strict. An interesting feature of conservation projects is their bimodal characteristics with respect to technology issues. Approximately half of examined projects used very simple tools, similar to the infrastructure employed by action projects. However, the other half of examined conservation projects employed sophisticated infrastructure and most modern tools, for instance, harvesting earthquake related data from Twitter feeds or using geo-tagging of pictures of invasive plant species. An example of a conservation project is “Spotting the Weedy Invasives”¹⁰ which aims at providing close monitoring of invasive species along hiking trails in the states of New Jersey and New York.
- *Investigation*: This is probably the class of projects most closely related to the traditional understanding of citizen science. These projects require distributed data collection from physical environments, but the educational aspects of the process are considered to be fairly important for these projects. The range of projects varies from regional to international, and so does the number of participants. Investigation projects scale well and it is not uncommon to see projects with thousands of amateur contributors. The primary domain covered by investigation projects are biology and ecology, but there are notable examples of projects in the field of meteorology and climatology. Scientific issues are at the core of investigation projects, with rigid protocols for data collection and reporting. Often, the fact that the amateur contributors are geographically distributed over large areas is an important part of project design and planning. With respect to the technology utilized in surveyed projects, investigation projects are as diversified as conservation projects. Some projects use only the simplest tools, e.g. an online survey, a web form to send data, and basic spreadsheets to manage the collected data. There are also investigation projects that use custom multi-device platforms, with sophisticated back-end systems for data storage and processing, and specialized applications for data collection. A good example of this class of projects is the “Lost Ladybug Project”¹¹, which uses volunteers to track the distribution of ladybugs in North America.
- *Virtual*: Virtual projects are a special case of citizen science projects, where all interaction between participants and researchers is mediated using ICT tools, with no physical objects involved in the process. These projects resemble investigation projects, but in contrast to the later, they do not require any interaction with the physical environment. Virtual projects are quickly becoming popular in many areas of science, after initial

¹⁰www.rci.rutgers.edu/~trails

¹¹<http://lostladybug.org>

l success of projects initiated in astronomy (“GalaxyZoo”¹²), proteomics (“Foldit”¹³) and paleontology (“The Open Dinosaur Project”¹⁴). Virtual projects scale very well, but they also suffer from the problem of data quality assurance. The relative ease of joining the project results in the increase of the number of dubious submissions of data. The project design must take it into careful consideration to propose a rigid protocol of data collection and analysis, but at the same time, it should encourage the participation of committed amateurs. Often we observe gamification mechanisms being used to create incentives for long-term commitment and high quality of data. Virtual projects are exclusively designed and created by research facilities using a strict top-to-bottom approach. The sustainability of these projects is maintained using research funding from federal and state sources. Virtual projects are the most interesting from the point of view of computer science, as they present serious challenges for modern innovative infrastructure. The collaboration tools include: wikis and discussion forums for providing a space for discussions, gamification tools such as leaderboards, badges, halls of fame, and performance measuring tools to assess individual progress. The back-end services must be robust and efficient to serve thousands of requests for resources, or submissions of data analysis. In Section 5 we will discuss which modern ICT services may be of the greatest utility to virtual projects.

- *Education*: Finally, there is a subset of citizen science projects whose primary concern is education and public outreach. These projects do not intend to conduct a groundbreaking research, but rather focus on enabling the participants to learn and improve their scientific literacy. Little emphasis is put on the scientific validity of the collected data or discovered patterns, rather, projects try to improve the understanding of the scientific process, enhance critical thinking skills and reach out to the general audience. Interestingly, education projects tend to have most of data analysis and result interpretation passed to untrained volunteers. These projects also present a challenge for ICT, because many steps of these projects need to be virtualized or standardized. Also, since amateurs are expected to participate in almost all steps of the scientific process, several custom components are required, with a lot of variation depending on the demographics of the audience. For instance, education projects aimed at children need to have the environment designed in a game-like way and significantly simplified, whereas projects addressed to adults need more functionality at the cost of less eye candy.

4 Challenges and problems

Citizen science projects are among the most promising developments in the field of science. They improve public understanding of science, increase public awareness and appreciation for science, and often lead to truly interesting results. However, one must be careful when designing a citizen science project, because the involvement of a large body of untrained amateurs poses significant challenges, threats, and problems. These problems have been

¹²<http://www.galaxyzoo.org>

¹³<http://www.fold.it>

¹⁴<http://opendino.wordpress.com>

raised in the literature previously and recent experiences gained from finished projects allows to formulate some warnings about possible sources of errors.

4.1 Variability of participants' competency

Untrained amateurs enter citizen science projects usually motivated by their interest in a particular field of science. Unfortunately, enthusiasm alone cannot compensate the lack of skills required for the project. The mere fact that volunteers vary significantly in their training, accuracy, understanding of project's aims and ramifications, is the source of bias which may become expressed in the gathered data. One should point out that this problem cannot be easily solved using outlier detection techniques, because it is not related to signal noise in the data. Rather, it is a systematic bias introduced by the variability of participants' competency, demographics, commitment to the project, etc. Another phenomenon which affects many citizen science projects is the quality level drift, i.e., the gradual increase of quality over time. Participants learn how to follow project's protocols, improve their observational skills, become more familiar with the project's domain, etc. Citizen science projects must develop mechanisms to protect against observer bias by providing additional meta-data about observers, using data quality policies, introducing multiple levels of filters on data streams, and building machine learning models to automatically validate incoming data.

4.2 Sampling bias

It is very difficult to verify the strict adherence to data collection protocols, in particular, when data collection involves periodic data gathering in the same point. Amateurs may not be aware of possible sampling bias and regard projects with too rigid protocols as unfriendly. Often, when designing a citizen science project, researchers loosen the protocols to widen participation, but reserve a subset of crucial measurements for a core of trusted, possibly trained, participants. Standardization of sampling is often infeasible and the analysis of the collected data must take into consideration the amount of effort expended in the collection of data (number of participants, localization of observations, time span between observations). Sampling bias may result from two different factors: variation in sampling effort (time sacrificed to observations) and variation in sampling space (spatial heterogeneity of sampling). Some bias can be discovered and annotated using collaborative filtering methods, where similar observations (with respect to time and space) are compared and their coherence is measured. However, there are no clear guidelines on how to avoid sampling bias and each project must plan for this challenge individually.

4.3 Financing of citizen science

Citizen science is one of the most cost-effective means to perform scientific research. Partly this results from a significant input of free workforce contributed by amateurs, but citizen science projects are particularly well suited for crowd funding. For instance, Cornell's FeederWatch Project participants support the project by small fees and contribute not only their work, but additional \$3 000 000 annually. Citizen science gains momentum and becomes an important player in the world of large scientific projects. According to the "Green Paper on Citizen Science" published by SOCIENTIZE, an EU-funded consortium coordinating the debate about the role of Citizen Science in Europe, only within the 7th Framework Programme of the EU at least 17 citizen science projects have been supported. Year 2014 marks

the beginning of the Horizon 2020, a funding program for research and innovation forming the core of the Europe 2020 strategy. The official stance of the European Commission on citizen science is that:

With the aim of deepening the relationship between science and society and reinforcing public confidence in science, Horizon 2020 should favour an informed engagement of citizens and civil society on research and innovation matters by promoting science education, by making scientific knowledge more accessible, by developing responsible research and innovation agendas that meet citizens' and civil society's concerns and expectations and by facilitating their participation in Horizon 2020 activities.

It is important to note that the funding of citizen science projects is not different from the way traditional science projects are funded. Major scientific funding institutions, such as National Science Foundation in the USA or national science centers in the European Union are treating citizen science project proposals uniformly with other types of projects. There also exist specialized funds for supporting citizen science projects, e.g. The Takagi Fund for Citizen Science or OPT Citizen Science Foundation.

5 ICT Services for citizen science

The Web 2.0 revolution spreading over the Internet has dramatically changed the way data is gathered and processed by web applications. The static, authoritarian model of the Web has been abandoned in favor of dynamic, community-driven model of user-generated content. Social networks appear abundantly in all domains of human activity, presenting users with limitless volumes of data, information, and knowledge. Let us now turn our attention to new ICT technologies and discuss their usefulness to citizen science projects.

5.1 Web 2.0

The term Web 2.0 has been coined by Tim O'Reilly in 2004 during the first O'Reilly Media Web 2.0 conference. According to O'Reilly [13] "*Web 2.0 is the business revolution in the computer industry caused by the move to the Internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: build applications that harness network effects to get better the more people use them*". This basic definition has been further extended by O'Reilly with the following set of rules for Web 2.0 applications:

- *the perpetual beta*: software is not an artifact, it is a way to engage with users,
- *small pieces loosely joined*: data and services should be opened to others for re-use, similarly, external data and services should be re-used whenever possible,
- *software above the level of a single device*: applications are not bound to specific clients or servers, instead, they exist in the space between devices,
- *the law of conservation of attractive profits*: open APIs and standard protocols do not contradict the idea of competitive advantage,
- *data is the Intel inside*: source of future lock-in and competitive advantage is not software or hardware architecture, but user-generated data, namespaces, and proprietary formats.

Many people perceive Web 2.0 as a set of buzzwords, teenager Web applications, and a certain graphic design style. These people criticize the entire idea of Web 2.0 as being yet another dot-com bubble, similar to the one of March, 2000. Contrary to these claims, in this dissertation we will accept the perspective advocated by O'Reilly *et al.*, that Web 2.0 defines a new quality in software engineering and provides a new paradigm for application design. It is also quite evident from the above rules that the Web 2.0 ideology is closely aligned with most aspects of citizen science. Perpetual improvement of research infrastructure and applications, distribution of work among small, independent and autonomous groups or individuals, using a collection of devices coupled together to form the research toolbox, or the stress on the importance of data collection and sharing, all these features of Web 2.0 applications can be directly translated to principles of citizen science.

The first rule disqualifies software as an artifact, a by-product of software development, but defines software as means to communicate and engage with users. Indeed, a user contribution is crucial for the success of modern software. This rule coincides with the second rule that encourages opening of application programming interfaces (APIs) and services, as well as sharing the data. One can hardly challenge this point. The immense success of services such as Google Search, Google Maps, Facebook, Amazon, and many others, lies primarily in enabling access to these services via open APIs. Openness and accessibility of both services and data allow thousands of external contributors to develop value-added solutions on top of existing systems. In the context of citizen science projects, the first two rules of Web 2.0 define the way to engage participants. The third rule basically departs from deprecated software architectures, such as the client-server architecture, and advocates the adoption of multi-tier and service-oriented architectures, where software components act like autonomous agents communicating and exchanging data. This rule is mainly dictated by unprecedented advances in personal and mobile computing hardware. The pace of innovation in these devices makes software development window very tight and, effectively, dictates the need to develop software independently of a current hardware platform. Again, looking at long-term citizen science projects one must assume that the dominant platforms change every couple of years and the sustainability of many projects depends on the early availability of applications for new generations of hardware. The fourth rule is the axiom of open source community. The main idea is that the profit and competitive advantage lies not in proprietary closed code, but in innovation, adaptation, and ease of approval. This rule, however important for open-source movement, is not relevant to our focus of attention. Finally, the most important fifth rule states that the main reason for lock-in and the most valuable asset is neither software nor hardware, but the data itself. This conviction has been long popular among database engineers, but not widespread in computer science community. Software platforms change every five years, hardware platforms change every seven years, but the data remain forever. This claim may seem exaggerated, but it is counter intuitively correct. Most contemporary database systems do not consider deleting any data and many free services do not impose limits on the volume of data being stored. Storing and querying data has become easier with advances in relational database technology, but acquiring data and feedback from users is difficult. The main mantra of Web 2.0 proponents is to turn spectators into participants. This means creating incentives for users to participate and contribute. Many solutions from the world of Web 2.0 rely heavily on users input, e.g., folksonomies, blog networks, wikis, recommender engines. Projects that manage to attract the biggest crowd of devoted followers are most likely to gain competitive edge and win the race.

5.2 Service domains

But which ICT tools and services are the most useful for the design and implementation of citizen science project? To answer this question one has to consider the general model for developing a citizen science project from scratch, using experiences from previous models. This perspective is employed in [3], where authors divide the planning into a set of distinct steps:

- *choose a scientific question*: Having a simple project protocol and an inspiring question helps reach large audiences of participants. At this point researchers may try to formulate scientific question themselves, but may also use existing online platforms to start verifying the assumptions behind the project from the beginning. An example of a service which could be used at this stage is the online knowledge base StackOverflow¹⁵.
- *form a team*: Research team consists of scientists, educators, technologists, and volunteers. Such teams can be formed using simple online invitations, but several other possibilities are viable. For instance, some innovative science projects have been announced as Kickstarter¹⁶ campaigns seeking public support under the form of funding or working.
- *develop, test, and refine protocols, data forms, and educational support materials*: This step is usually performed internally by the institution overseeing the citizen science project. Simple group collaboration tools can be used to create necessary material, but researchers should consider using social networks for the dissemination of the material. Having a YouTube channel with simple informative videos related to the project, or maintaining a Twitter account for quick announcement of project development, certainly helps to extend the initial reach of the project.
- *recruit participants*: Recruitment is an important aspect of the project design. One may of course depend on simple e-mail of web form subscription and registration, but that might prove not enough. The social webs are full of narrow focused communities which may be utilized as hubs to broader public. Researchers should try to actively reach out to online communities, such as Reddit¹⁷, in search of volunteers. For small localized citizen science projects one may use more sophisticated tools, e.g., using event management services, such as Evenea¹⁸.
- *train participants*: For small projects training is performed during regular meetings, thus ICT tools are not required for this step. For large scale projects the only viable alternatives are vidcasts, online meetings (e.g., WebEx¹⁹, or online conference calls with desktop sharing (such as offered by Skype).
- *accept, edit, and display data*: For this step literally hundreds of alternatives exist. Data may be collected using web forms and file uploads, or send using e-mails. Several online office suites, such as Google Docs or Zoho, provide tools for data collection and storing. Finally, data may be displayed using a wide range of tools, online charts, Web Service endpoints, and many more.

¹⁵<http://www.stackoverflow.com>

¹⁶<http://www.kickstarter.com>

¹⁷<http://www.reddit.com>

¹⁸<http://www.evenea.com>

¹⁹<http://www.webex.com>

- *analyze and interpret data*: The task of data analysis and interpretation is usually assigned to trained science professionals, who employ their tools of choice, which in turn is affected by the particular choice of the scientific question. But even here wide collaboration is possible. For instance, some projects use the version control systems, such as Subversion²⁰ or Git²¹, to manage various stages of data analysis. Traditionally, these systems have been used to enable distributed collaboration on software code. Today many researchers use Git to manage their scientific notebook, files, intermediate analysis, and the like.
- *disseminate results*: Similarly to *accept, edit, and display data* step, dissemination of results can be performed in hundreds of ways. Beside traditional publication venues, such as conferences and journals, results of citizen science projects should be pushed to local communities via publications in popular newspapers, newsletters, and blogs. Wiki systems are particularly well-suited for dissemination of results, because they allow direct interaction with the published results and encourage the public to engage in discussions about the projects.
- *measure impacts*: Several measures can be used to assess the success of a particular project. Among others, researchers can measure the number of papers published in peer-reviewed venues, the number of citation of results, the number and sizes of grants received for citizen science, size and quality of citizen science databases, and frequency of media exposure of results. If the project uses blogs to publish partial and final results, trackbacks and pingbacks can be used as measures. Apart from scientific value of the project, researchers may want to measure the social impact of the project. This can be done by measuring duration of involvement by project participants, improved participant understanding of science content, and better participant attitudes toward science.

5.3 Implementation of a citizen science project

Following the list of steps outlined in Section 5.2 let us see how these steps can be implemented using an example of a real citizen science project: Cornell's Lab of Ornithology (CLO). CLO is in fact a collection of projects, some of them as old as twenty years of uninterrupted activity. This is a highly successful example of citizen science at work, with numerous high profile publications and thousands of citizens invited for active participation. The first step in the development of a new citizen science project is the choice of a very specific question. This question cannot be too general or vague, it must be answerable and understandable to the public. CLO projects do exactly that: the main question behind the FeederWatch Project is simply "*What are the patterns of irruption in winter finch populations?*", whereas project NestWatch tries to answer the following question: "*How do clutch sizes of eastern bluebirds vary with latitude?*". These are simple questions that do not require sophisticated knowledge from participants, who can easily estimate the size of the bluebird's clutch. Sometimes, more elaborate questions may be chosen if experimental protocol is simple enough for practical applications. For instance, in another of CLO's projects, the Seed Preference Test, participants were providing birds with different types of seeds and have been measuring birds' responses. The next step is to form the team, consisting of a

²⁰<http://subversion.apache.org>

²¹<http://git-scm.com/>

scientist (to ensure that the project maintains scientific integrity and correctness), an educator (responsible for reaching out to the general public), a technologist (for administering the required technological framework), and an evaluator (whose job is to ensure that the project has measurable outcomes and to constantly monitor the state of the project). After forming the team CLO projects develop, test, and refine protocols, data forms, and educational support materials. For instance, when developing the House Finch Disease Survey scientists performed multiple meetings with local school groups and bird watch groups to create correct protocols for data collection. As the result the researchers realized that citizens were far more likely to report positive sighting results than negative results. This has led to improved educational materials explaining in detail the importance of negative results. The next step is to recruit participants. Here, the possibilities are countless: press releases, newsletters, mass mailing campaigns, local advertisement, etc. As an example, CLO's eBird project has been recruiting young participants in middle schools by first working closely with biology teachers in over 100 schools. After recruitment the training of participants commences. Again, depending on the target demographics of participants different strategies are available. In CLO's Celebrate Urban Birds! project the main tool for training were workshops organized in science museums and youth centers. For the next step, accept, edit, and display data, the most convenient way is to set up a dedicated website or develop a dedicated mobile application. This approach has been taken by virtually all current citizen science projects. The following step of data analysis and interpretation is often performed at the university by highly skilled statisticians and data scientist, so it is not crucial from the point of view of a particular project's implementation. Finally, the dissemination of results may be done by project's website, a newsletter (this approach is taken by many CLO's projects, who publish results in CLO's quarterly newsletter *BirdScope*), or may be presented directly to policy makers and land managers, as is the case of the Avian Knowledge Network.

The evaluation of the project is a crucial and indispensable step in project's implementation. The evaluation can be twofold: firstly, we can measure the scientific contribution of the project. This can be measured quantitatively easily by looking at the number of papers accepted to peer-reviewed journals, papers presented during conferences, counting the number of citations, evaluating the sizes of grants awarded to citizen science projects, or even looking at the sizes and data quality metrics of citizen science databases. Secondly, we may try to measure the scientific literacy outcomes. This becomes more tricky, but several metrics are available: the number of participants, the average participation duration, understanding of the scientific process by participants, surveys on citizens' attitudes to science, increased popular interest in science as a career, etc. These measures may be taken by pre- and post-project surveys, email questionnaires, direct interviews, and many more.

6 Conclusions

As we have seen, citizen science is a promising and quickly developing area of scientific research. Many fascinating promises revolve around the model of citizen science, with improved public understanding of science, engagement of local communities in solving their problems, and performing large-scale studies that were previously infeasible to conduct. On the other hand, we have discussed challenges and threats to citizen science projects and we have identified key software components that can be used to help design successful citizen science projects. It turns out that many of the Web 2.0 tools and services are readily available for various steps of citizen science projects. Our discussion is not limited to virtual projects, which are entirely mediated by ICT tools. We believe that recent

developments in software architectures, collectively referred to as Web 2.0, enable the usage of citizen science paradigm in domains that were reluctant to rely on public involvement in science process.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Bock, C.E.: The christmas bird count and avian ecology (1981)
2. Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., Wilderman, C.C.: Public participation in scientific research: defining the field and assessing its potential for informal science education. A case inquiry group report. Online Submission (2009)
3. Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J.: Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience* **59**(11), 977–984 (2009)
4. Brabham, D.C.: Crowdsourcing as a model for problem solving an introduction and cases. *Convergence* **14**(1), 75–90 (2008)
5. Bystrak, D.: The north american breeding bird survey (1981)
6. Cooper, C.B., Dickinson, J., Phillips, T., Bonney, R.: Citizen science as a tool for conservation in residential ecosystems. *Ecol. Soc.* **12**(2), 11 (2007)
7. Downes, S.: Learning networks in practice (2007)
8. Ellis, D., Oldridge, R., Vasconcelos, A.: Community and virtual community. *Annu. Rev. Inf. Sci. Technol.* **38**(1), 145–186 (2004)
9. Evans, C., Abrams, E., Reitsma, R., Roux, K., Salmonsens, L., Marra, P.P.: The neighborhood nestwatch program: Participant outcomes of a citizen-science ecological research project. *Conserv. Biol.* **19**(3), 589–594 (2005)
10. Howe, J.: The rise of crowdsourcing. *Wired Mag.* **14**(6), 1–4 (2006)
11. Leydesdorff, L., Ward, J.: Science shops: a kaleidoscope of science–society collaborations in europe. *Publ. Underst. Sci.* **14**(4), 353–372 (2005)
12. McIntyre, A.: Participatory action research, vol. 52. Sage (2008)
13. O'Reilly, T.: What is web 2.0 O'Reilly (2009)
14. Ratcliff, J.: The transit of Venus enterprise in Victorian Britain, vol. 3. Pickering & Chatto Limited (2008)
15. Trumbull, D.J., Bonney, R., Bascom, D., Cabral, A.: Thinking scientifically during participation in a citizen-science project. *Sci. Educ.* **84**(2), 265–275 (2000)
16. Wachelder, J.: Democratizing science: various routes and visions of dutch science shops. *Sci. Technol. & Hum. Values* **28**(2), 244–273 (2003)
17. Walter, M.: Participatory action research. *Social Research Methods*, 2nd edn. Walter, M. (ed.), vol 21. Oxford University Press, Australia (2009)
18. Wiggins, A., Crowston, K.: From conservation to crowdsourcing: a typology of citizen science. In: 2011 44th Hawaii International Conference on System Sciences (HICSS). IEEE (2011)
19. Wilderman, C.C.: Models of community science: Design lessons from the field. In: Citizen Science Toolkit Conference. In: McEver, C., Bonney, R., Dickinson, J., Kelling, S., Rosenberg, K., Shirk, J.L. (eds.) Cornell Laboratory of Ornithology. Ithaca, NY (2007)