Chapter 25

Future Data and Knowledge Needs

Coordinating Lead Authors: Florence Daguitan (Tebtebba, Indigenous Peoples' International Centre for Policy Research and Education), Charles Mwangi (Global Learning and Observations to Benefit the Environment [GLOBE] Program), Michelle Tan (ADEC Innovations)

Lead Authors: Graeme Clark (University of New South Wales), Daniel Cooper (University of Oxford), James Donovan (ADEC Innovations), Sheryl Gutierrez (ADEC Innovations), Nina Kruglikova (University of Oxford), Pali Lehohla (Pan African Institute for Evidence), Joni Seager (Bentley University), William Sonntag (Group on Earth Observations)

GEO Fellow: Amit Patel (Planned Systems International, Inc.)

Executive summary

Citizen science is providing unprecedented opportunities for engaging the public in collecting and analysing vast amounts of environmental data (*well established*). The potential for massively dispersed teams of observers, coupled with new technologies such as smart sensors, mobile telephony, Internet and computing capabilities, is offering new approaches for research and engaging the public on environmental issues. As well as collecting large volumes of data, the advancement of new technologies has also enhanced the quality and veracity of the data collected. Key opportunities presented by citizen science include greater frequency of data from dispersed sources, the ability to address large knowledge and funding deficits, the ability to educate the public about environmental policy issues, and the use of local knowledge. {25.2.1}

Big data is one of the world's emerging valuable resources, shifting the landscape of environmental assessment at global, national and local scales (*well established*). Traditional processing techniques cannot handle the volume, velocity, variety and veracity of big data, demanding new algorithms, programming and statistical methods to derive information and draw evidence-based conclusions. There is enormous potential for advancing environmental knowledge if big data can be effectively harnessed and interrogated. {25.2.2}

Governments, organizations, academia and the private sector have initiatives seeking opportunities to tap the potential of big data for sustainability and development (*well established*). Current initiatives include the establishment of the United Nations Pulse Labs for pilot studies on big data, the formation of the United Nations Global Working Group on Big Data in monitoring the Sustainable Development Goals (SDGs), and the availability of repository sites and open data sources from multilateral organizations, research centres and government collaborations. Big data from web-based and geospatial mapping technologies, remote sensing and statistical visualization provide a basis for environmental assessment. {25.2.2}

Challenges for using big data in environmental assessments include its accessibility, quality, varying scale and context, and incomplete time series (*well established*). Despite efforts to generate globally acceptable and available big data, capacities are limited by resources and funding, especially in developing countries. Much real-time big data are controlled and held by the private sector, though many data products are made freely available for public good in a process known as data philanthropy. Recommendations for building a holistic system for big data include the establishment of leadership and data governance; collaborations among governments, institutions and the private sector; and institutionalizing legal frameworks with safeguards on information. {25.2.2}

Strengthening the ability to gather, interpret and use data for effective planning, policymaking, management and evaluation could provide countries with a comprehensive view of environmental impacts (*well established*). Governments and society need to adapt to the evolving data landscape, including the possible use of artificial intelligence to manage environmental concerns. Coping with the shift in the data landscape entails new information-technology skills and a holistic approach in utilizing emerging and existing data and knowledge tools. {25.3}

Traditional knowledge held by indigenous peoples and local communities is increasingly seen as a valuable resource for environmental assessment and sustainable development (*well established*). This revaluation is evidenced by the increase in discussions and studies on traditional knowledge, and its inclusion in global policy agreements. In order to address current and future challenges such as climate change, research suggests that the best approaches may be characterized by the coordination of modern science and technology with traditional knowledge. While cooperation between local and global communities and knowledge systems has proven to be successful for the health of individuals and the planet, certain challenges remain. {25.2.3}

25.1 Introduction

This chapter discusses emerging areas of environmental information and statistics, including citizen science, big data and traditional knowledge. It aims to summarize the gaps and opportunities for improving the environmental knowledge base.

The global landscape is changing, technology is advancing and more and more data are available. These new data sources will not override the need for traditional means of data collection but will provide additional opportunities for environmental monitoring and assessment. This chapter analyses these new and emerging means of data collection and presents a perspective for the future of environmental monitoring and assessment.

25.2 Emerging tools for environmental assessment

Citizen science, big data and traditional knowledge are not new sources of information; what is new is their regular and systematic use in environmental assessments. This section highlights some current experiences and the need to use these innovative sources of information to fill data gaps.

25.2.1 Citizen Science

Citizen science entails the engagement of volunteers in science and research. Volunteers are commonly involved in data collection, but can also be involved in initiating questions, designing projects, disseminating results and interpreting data (Blaney *et al*. 2016). Coupling Citizen Science with new emergent technologies is providing unprecedented opportunities for doing research and sensitization of the public on environmental issues (Newman *et al.* 2012, p. 298).

The possibility of tapping into a massive, dispersed team of observers in different regions of the world has created opportunities for collating and analyzing data at unprecedented spatial and temporal scales. Citizen Science projects have the potential to gather large amounts of scientific data but this is only helpful if data collected is utilized in one way or the other (Dickinson, Zuckerberg and Bonter 2010; Kim *et al.* 2011; Dickinson *et al.* 2012).

Citizen science has numerous benefits, the main one being the opportunity to collect data over wider spatial coverage and longer periods at lower cost. Additional benefits include the creation of jobs, increased scientific literacy, citizen engagement in local and environmental issues, cost effectiveness for governments and benefits to the environment being monitored. Citizen Science also allows the expertise of scientists to be brought to the public while at the same time exposing the scientists to the indigenous knowledge and expertise available within the local community (Conrad and Hilchey 2011; Blaney *et al.* 2016). Some of the key benefits of citizen science are highlighted in **Figure 21.1: Selected targets and their related clusters as examined in this chapter**.

The fields of astronomy and ornithology have led the charge for citizen science. In 1900, Frank Chapman, an ornithologist with the American Museum of Natural History initiated the Christmas Bird Count (CBC). This project has survived thanks to the enthusiasm of citizen scientists over the years and is currently being run by the National Audubon Society (Dickinson, Zuckerberg and Bonter 2010). Since then, there have been many citizen science projects over the years at local, regional and global scales, covering different areas of interest.

More recently, citizen science projects have included a wide variety of initiatives, ranging from building collaborative knowledge (e.g. Wikipedia, OpenStreetMap), volunteer computing (e.g. CitizenGrid, climateprediction.net), and pattern classification (e.g. Galaxy Zoo, eyewire), to the community collection of observations (e.g. bird counting, air sensor toolbox) (Mathieu *et al.* 2016).

Many environmental interests that transcend government boundaries, such as pollution and bird migration, have increased the engagement of citizen scientists to monitor these issues of concern. More innovative projects include the use of Google's reCAPTCHA, which has facilitated the digitization of books and millions of articles by turning words that cannot be read by computers into CAPTCHAs for people to solve (Conrad and Hilchey 2011; Google 2018).

There are two main approaches used in the organization of citizen science projects; top-down or bottom-up. These approaches are similar to the concepts in Chapter 10 on evaluation of policy effectiveness.

Figure 25.1: Some of the benefits of citizen science

Individual Citizen

- Learn observational and analytical skills
- Gain a better understanding of the natural world
- Job opportunities
- Capacity building

Governments

- Lower cost of data collection
- Wider spatial and temporal coverage of data
- Promote environmental stewardship

Communities

- Monitor the health of the environment
- Increased interaction of the community
- Promote environmental stewardship

Scientists and Researchers

- Large numbers of participants reduce workload
- Scientists are able to build connection with community
- Teach people how to research

The top-down approach is mostly driven by scientists who train volunteers on the procedures and the research to be undertaken. Based on this approach, the volunteers play limited roles mostly in data collection. The bottom-up approach is driven by the community. More often than not, this is driven by the need of the community to understand or gather evidence of a concern. The community can then approach scientists for support and guidance during the process (Roelfsema *et al.* 2016; Shirk *et al.* 2012).

The level of engagement, skills and knowledge needed by volunteers to participate in citizen science projects varies depending on the scope of the research. Some projects require basic data collection knowledge requiring minimal or no training of volunteers, while others require intensive training (Haklay 2013; Shirk *et al.* 2012). **Figure 21.4: Future projections of global average crop yield (top left), crop production (top right), agricultural area (bottom left), and forest and other natural land area (bottom right)**, illustrates the various levels of engagement of volunteers in citizen science projects.

Citizen scientists can help to uncover critical information about our environment which could possibly take scientists years to discover by themselves. An example is illustrated by the infographic shown in **Figure 25.3** where rivers need a citizen science movement for monitoring, and how the collected data and findings are used to maintain ecosystem services and human wellbeing (Pottinger 2012). The figure also demonstrates a step-by-step procedure for conducting citizen science. This data collection and analysis procedure can be replicated across the Drivers (Chapter 2) and the various environmental themes (Chapters 5 to 9).

Figure 25.2: Levels of citizen science by increasing depth of the participation

Level 4 — Extreme Citizen Science
- Collaborative science - problem definition, data collection and analysis

Level 3 — Participatory Science
- Participation in problem definition and data collection

Level 2 — Distributed Intelligence
- Citizens as basic interpreters
- Volunteered thinking

Level 1 — Crowdsourcing
- Citizens as sensors
- Volunteered computing

Source: Haklay (2013).

Trends in citizen science

The technology revolution has heralded multiple novel ways of collecting, archiving, analyzing, and transmitting data. The emergence of the internet-of-things (IoT), miniaturized smart sensors with geo-location functions, ease of accessing internet and data as well as the potential of cloud storage and computing has expanded the possibilities and opportunities for data collection and analysis. This rapid advancement in technology coupled with greater exposure and sensitization of the public, have led to an explosion in uptake of projects based on citizen science (Mathieu *et al.* 2016).

The availability of internet and geographic information system (GIS)-enabled web applications has enabled citizen scientists to collect large volumes of geographically-referenced data and submit them electronically to centralized databases. An example of such a system is the Global Learning for Observation to Benefit the Environment (GLOBE) program, which uses students to collect environmental data and archive it in the GLOBE program database (Dickinson *et al.* 2012; GLOBE 2018).

The expansion in the use of smartphones, the possibility of digital photo validation of observations, and the capability of creating simple online data-entry systems is revolutionizing the process of initiating citizen science projects while ensuring data accuracy at minimal cost. Currently, it is now possible to create mobile phone apps for collecting different types of datasets and automatically geo-locating the data, using the in-built GPS receiver chip on most mobile phones (Dickinson, Zuckerberg and Bonter 2010; Dickinson *et al.* 2012).

Scientists are now increasingly using citizen scientists to collect geo-referenced *in-situ* data which can be used to support the calibration and validation of Earth Observation satellite data products. Citizen scientists are also involved in the interpretation and digitization of Earth Observation (EO) data sets (Mathieu *et al.* 2016; See *et al.* 2016). Tomnod is such an example of using crowdsourcing and citizen scientists to identify objects and places in satellite images. Tomnod was used in trying to locate the missing Malaysian Airlines flight MH370 aircraft using satellite imagery. Approximately 2.3 million Internet users submitted 18 million tags for over 745,000 satellite images clearly illustrating the potential of citizen science (Mazumdar *et al.* 2017).

Another example of the use of citizen scientists to validate satellite data is the partnership between NASA's Global Precipitation Measurement (GPM) satellite mission and the GLOBE program. The GLOBE program is an environmental

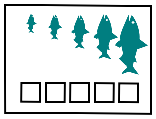
Figure 25.3: An example of citizen science that demonstrates how it is needed and can be replicated

Case Study: Why our rivers need a citizen science movement

Most of our decisions are based on incomplete or inadequate data/information - in the abscence of professional scientists to fill these gaps, citizen science can step in to help uncover information and findings. This case study explores opportunities where citizen scientists have filled this gap.

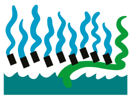
TOO MUCH WORK, TOO FEW SCIENTISTS

- There is too much research that needs to be done yet hardly enough scientists to undertake it all by themselves.
- Volunteers in the US state of Oregon are helping scientists survey 146 miles of streams by locating and counting salmon and native trout species and helping restore habitat.
- Hundreds of volunteers with The Nature Conservancy annually survey how much desert land is made wet by the San Pedro River; they cover more than 250 miles.



BECAUSE THEY'RE OUR RIVERS

- Volunteers for the Mystic River Watershed Association in the Eastern US, with support of scientists, take monthly samples at 15 locations along the river to monitor water quality.
- Advocacy based on their results has helped improve the river's cleanliness and has enabled the residents get involved in their natural environment through hands-on science.



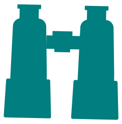
Just 1 in 10 rivers now reach the sea

8/10 people depend on river resources

Just 2/3 of Earth's major rivers are dammed

WHAT WE DON'T KNOW CAN HURT US

- Citizen science can be used to document basic information about a river system, as well as changes over time to its flow, sediment load, species and water quality.
- In China, the South-North Water Transfer Project was envisioned to have massive impacts on many key waterways, volunteers were recruited for a 4 year assessment of 10,000km of China's western rivers.



UNCOVERING RIVER MYSTERIES

- Citizen scientists can fill in gaps in crucial baseline knowledge about a river's species or general health.
- The Mekong River in Southeast Asia supports several species of giant fish and very little is known about them. More information is needed on where they spawn, what natural cues drive them to spawn, population estimates, and maps of their life-cycle territory.
- A step-by-step guide on Mekong River citizen study illustrates how this can be realized.

STEP-BY-STEP CITIZEN SCIENCE GUIDE: CASE STUDY OF FISHING VILLAGERS DOCUMENTING MEKONG'S RIVER'S NATURAL WEALTH

IDENTIFY THE QUESTIONS YOU WANT TO ANSWER

- In 1994, Thailand built Pak Mun Dam on the largest tributary of the Mekong, destroying local fisheries and harming river-based communities.
- Information on local fisheries was scant.
- In 2001, the Thai government opened the dam's floodgates 1-year study of its impacts to fisheries.

FORM A RESEARCH TEAM

- South East Asia Rivers Network (SEARIN) and Assembly of the Poor teamed up to monitor the changes caused by the dam.
- Their innovative citizens' science research method, called Thai Bahn (Thai Villager) research, relied on local fishers to gather information.

DEVELOP A PLAN OF ACTION

- Methods, areas of study, and research team members were all decided by the local villagers.
- SEARIN helped develop a plan of action, write up their findings and increase international awareness.

DOCUMENT YOUR FINDINGS

- The natural flows of the one-year trial period allowed people to resume traditional ways of life and eased resource conflicts among river communities.
- Local fish species not seen for eight years came back; researchers found a total of 156 fish species had returned to the Mun River.

ANALYZE YOUR DATA

- SEARIN helped create a report on the team's findings, in two languages.
- The report is considered one of the most thorough documentations of Mekong fisheries produced for that area.

SHARE YOUR FINDINGS, AND USE THEM FOR ACTION

- Thanks to this citizen science effort, the villagers succeeded in getting the Thai government to open the dam gates for four months each year to allow for fish migration.
- Subsequent governments have not implemeneted this agreement.
- The project has inspired many other citizen science projects to protect rivers in the region.

Source: International Rivers (2012).

25

educational program for primary and secondary schools, where students from schools across the world collect precipitation data using rain gauges as shown in **Figure 25.4**. The collected data, as well as data collected from other sources, is used by NASA to calibrate and validate GPM precipitation measurement data (United States National Aeronautics and Space Administration [NASA] 2018).

Figure 25.4: GLOBE Students in St. Scholastica Catholic School in Nairobi collecting and recording the amount of precipitation for the GPM Satellite Mission field campaign

Source: © GLOBE Program (Kenya).

Automated and autonomous equipment such as drones, remotely operated sensors, autonomous underwater vehicles (AUV's) and underwater gliders are predicted to play an increasing role in citizen science. These autonomous systems can be a primary source of data or complement data collected *in situ*, provide high resolution data nearly in real time, be deployed on a need basis and often enable access to remote or extreme locations such as observation of marine environments. In addition, they are low cost compared to satellites and are thus offering alternative and credible sources of EO data (Macauley and Brennan 2016; Garcia-Soto 2017).

Citizen science, as well as other data sources, contributes to Big Data collection and these huge volumes of data need processing. Numerous approaches have been explored to involve the huge numbers of citizen scientists to assist in analyzing these huge volumes of data, one of which is the development of game-like systems (gamification). Citizen participation in these games help to speed up the data analysis and allow science to advance more rapidly (Van Vliet and Moore 2016; Spitz *et al.* 2017; McCallum *et al.* 2018).

An example of gamification is Cropland Capture, a game version of the GeoWiki project, which engaged citizen scientists in global land cover research, helping researchers identify farmland around the world. The game managed to collect 4 million classifications from over 3,000 players identifying images with and without cropland present (See *et al.* 2013).

Table 25.1 shows some of the global and regional projects dedicated to citizen science.

The potential of citizen science should not be limited to engaging volunteers to collect and collate scientific data as illustrated in **Figure 25.5**. Citizen science can be used to sensitize and engage the community on issues related to their natural environment, to better understand them and allow them to take charge, and provide an avenue for showcasing the need

Table 25.1: A selection of citizen-science projects and websites

Programme	Region	Description	Website
UNEP Environment Live	Global	UN open access platform of global, regional and national environmental data	https://environmentlive.unep.org
SciStarter	Global	Aggregates information, video and blogs about citizen-science projects	www.scistarter.com
Data Observation Network for Earth	Global	Provides a framework to access data from multiple data sources (including citizen science data)	www.dataone.org
CitSci.org	Global	Provides tools for citizen scientists to guide them on the entire research process such as: process of initiating research projects, managing the process of data collection, and analysis	www.citsci.org
iSpot	Global	Website aimed at helping anyone identify anything in nature by connecting citizen scientists with experts in species identification	www.ispotnature.org
eBird	Global	Online database of bird observations with real-time data about bird distribution and abundance	www.ebird.org

Figure 25.5: Citizen scientists collecting environmental data

Source: © GLOBE Program (Kenya).

to maintain and conserve our ecosystems given the increasing pressures on the environment (Roelfsema *et al.* 2016).

Challenges of citizen science

Challenges in citizen science mostly revolve around three main issues: *organizational* issues, *data-collection* issues and *data-use* issues. At the organizational level, the challenges include the process of recruiting volunteers, motivating and providing incentives for their participation and ensuring sustainability of the initiative as well as funding. On data collection, the issues that arise include: data fragmentation, data representativeness, data quality (for example data intentionally flawed by the data collector) and/or lack of essential metadata. In data use, the challenges include: differences in protocols and standards, legal issues, data-privacy concerns and the question of allowing open access (Conrad and Hilchey 2011; Hochachka *et al.* 2012; Rotman *et al.* 2012; See *et al.* 2016)

Due to misunderstandings and lack of technical knowledge and skills to handle such data, concerns have emerged over the credibility, comparability, completeness of, and lack of metadata, as well as challenges in data access and sharing, and these have resulted in these data not being seriously considered by policy and decision makers. In most cases, perception of poor data quality, rather than the actual data quality and fitness for use, have influenced the value and use of citizen science data (University of the West of England, Science Communication Unit 2013; Storksdieck *et al.* 2016).

The key opportunities presented by citizen science, mainly include:

i. use of local knowledge;
ii. timely data from dispersed sources;
iii. capability to address large knowledge and funding deficits;
iv. ability to educate the public about environmental policy issues; and
v. enhance participatory democracy.

For citizen science to be widely accepted, there is a need for appropriate training and support for citizen science project coordinators and those that use the data that emerge from it. Careful design of citizen science projects and application of appropriate quality assurance methods, as illustrated in **Figure 25.3**, can ensure that the effort of citizen scientists is not wasted (University of the West of England, Science Communication Unit 2013; Storksdieck *et al.* 2016).

There are on-going initiatives, such as the Public Participation in Scientific Research (PPSR)-Core data model framework as illustrated in **Figure 25.6**, to establish data and metadata standards to facilitate international collaboration and improve data standardization, interoperability, integration, accessibility, and dissemination of citizen science data (Bowser *et al.* 2017). Citizen science has the potential to provide credible data to bridge the data gaps highlighted in Chapter 3 and to provide data to enable the monitoring of SDG environmental indicators.

25.2.2 Big data and data analytics

Big data can be defined as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze" (Manyika *et al.* 2011).

Data are one of the world's valuable resources, shifting the landscape of environmental assessment across global, national and local scales ("The world's most valuable resource is no longer oil, but data" 2017). From 1.8 zettabytes (1.8 trillion gigabytes) of data generated in 2011 (International Data Corporation [IDC] 2012), the total amount of data is expected to reach 40 zettabytes (40 trillion gigabytes) by 2020 (Dell EMC and IDC 2014). With this influx, traditional processing applications will be unable to cope with the quantity of data from multiple sources. Big data is characterized by the four Vs of large storage capacity (volume), speed at which data are generated and transmitted (velocity), the complexity of unstructured data types (variety), and the uncertainty of data sources (veracity) **(Figure 25.7)**. A fifth V (value) is achieved through the application of data analytics (International Business Machines [IBM] 2017).

The science of data analytics is needed to create patterns from intricate data sets and find correlations (e.g. chemical pollution and locations in aerial photographs) by using algorithms, programming, and mechanical and statistical methods to draw evidence-based conclusions and obtain information that is useful for decision-making purposes (Monnappa 2017). Examples of insights drawn from big data analytics include those from projects in the United Nations Global Pulse initiative, such as:

i. urban dynamics drawn from mobile data used to improve transportation in Sao Paulo and Abidjan;
ii. campaign developments based on a survey of perceptions of HIV on social media; and
iii. and a support-services location plan based on the spatial epidemiology of Dengue fever (Kirkpatrick 2016).

Current trends and initiatives in big data
United Nations member states, in partnership with the academic and research communities, non-governmental organizations and the private sector, are seeking out innovations and looking for opportunities to tap into the optimum potential of big data for sustainability and development.

Innovation for public good
The United Nations Global Pulse initiative was founded in 2009 to progressively establish a global network of Pulse Labs to collect digital data for decision-making purposes (United Nations 2018a). Pulse Labs continue to innovate machines and to conduct pilot studies on the scalability of the capture and analytics of big data for sustainable development – some examples are presented in **Table 25.2.**

Figure 25.6: The PPSR-Core data-model framework