Advancing sustainability together?

Citizen-generated data and the Sustainable Development Goals.

Danny Lämmerhirt
Jonathan Gray
Tommaso Venturini
Axel Meunier

Global Partnership for Sustainable Development Data
OPEN KNOWLEDGE INTERNATIONAL
Public Data Lab

Executive summary

Citizen-generated data (CGD) expands what gets measured, how, and for what purpose. Initiatives cover areas from cartography to government policies, public services or environmental research. As the collection and engagement with CGD rises in relevance and visibility, public institutions can learn from existing initiatives about what CGD initiatives do, how they enable different forms of sense-making and how this may further progress around the Sustainable Development Goals.

Our research examined different approaches to doing and organising CGD, as well as how governments already engage with these types of initiatives. It identified several concrete benefits for implementing and monitoring the SDGs, underlining the importance for public institutions to further support these initiatives.

Key points:

- Dealing with data is usually much more than 'just producing' data. CGD initiatives open up new types of relationships between individuals, civil society and public institutions. This includes local development and educational programmes, community outreach, and collaborative strategies for monitoring, auditing, planning and decision-making.
- Generating data takes many shapes, from collecting new data in the field, to compiling, annotating, and structuring existing data to enable new ways of seeing things through data.
- CGD initiatives can help gathering data in regions otherwise not reachable. Some CGD approaches may provide updated and detailed data at lower costs and faster than official data collections.
- Beyond filling data gaps, official measurements can be expanded, complemented, or cross-verified. This includes pattern and trend identification and the creation of baseline indicators for further

research. CGD can help governments detect anomalies, test the accuracy of existing monitoring processes, understand contextual factors, and initiate their own follow-up data collections.

- CGD can inform several actions to achieve the SDGs. Beyond education, community engagement and community-based problem solving, this includes baseline research, planning and strategy development, allocation and coordination of public and private programs, as well as improvement to public services.
- CGD must be 'good enough' for different (and varying) purposes. Governments already develop pragmatic ways to negotiate and assess the usefulness of data for a specific task. CGD may be particularly useful when agencies have a clear remit or responsibility to manage a problem.
- Data quality can be comparable to official data collections, provided tasks are *sufficiently easy to conduct,* tool quality is high enough, and sufficient training, resources and quality assurance are provided.

Authors

Danny Lämmerhirt (Open Knowledge International), Jonathan Gray (King's College London), Tommaso Venturini (Public Data Lab), Axel Meunier (Sciences Po, Medialab)

Governments can support by:

- Developing sectoral laws and policies to provide for citizens' rights to produce data and to incentivise departments, ministries and agencies (DMAs) to consider, test, acknowledge, and support CGD.
- Developing pragmatic ways to negotiate with citizens when data becomes relevant and when data quality is sufficient. This includes providing resources (trainings, educational material), and dedicated staff to serve as contact person. If data quality classification systems are developed, these should be made public to ensure transparency.
- Proactively communicating how people can contribute data. People should know their rights to contribute data, their rights enabled through data, and how their data can be used within public institutions.
- Publishing statements whenever government decides not use citizen-generated data, so that the public can follow up.
- Fostering open government, open data and right to information policies. These can support CGD initiatives reliant on government data.
- Considering CGD as an expansion of existing participatory and administrative structures. This allows to strengthen these structures and to avoid double costs.
- Including budgets for CGD initiatives alongside funding for institutional data collection. This may be important in environments of shrinking donor funding. Budgets may cover community outreach, coordination, recruiting and training staff, data validation, and other processes.
- Supporting intermediary organisations with established relationships to people. Community

networks, maker labs, higher education institutions, and NGOs can provide programs to train, equip, and educate people.

Civil society and intermediaries can support CGD by:

- Documenting and sharing tools, protocols, and other strategies with other initiatives. To increase relevance for other initiatives, be explicit about methodological steps taken and data created, instead of using jargon and exclusive terminology.
- Seeking collaboration and possibilities to share data on existing data infrastructure and with existing communities.
- Considering the depth of engagement with communities. Beyond consultations, this includes training people to collect and analyse data (data literacy), or to help people understand data's relation to public institutions, and their rights (institutional literacy).
- Being mindful when data collections cover sensitive topics or include personally identifiable information.
- Ensuring data protection principles are followed so as to minimise the risk of unauthorised access to sensitive data, or undue and illegitimate data use.
- Clarifying the terms of use under which data is planned to be shared.

Introduction

For centuries statisticians have worked together across borders to collect and analyse information about collective life and develop the standards, conventions and methods through which this information is produced and communicated.

Recent years have seen a proliferation in the availability of different types of data from different sources, including not just statistics, but also data from sensors, satellites, drones, online platforms, mobile phones and numerous other digital devices and infrastructures, as well as from individuals. The Global Partnership for Sustainable Development Data (GPSDD) is a growing network of over 300 members, including governments, the private sector, civil society, international organisations, and academic institutions aiming to drive better decisions and better lives for all by facilitating the production and use of better data from all these diverse sources.

Beginning in late 2017, a group of GPSDD members working on citizen-generated data (CGD) created a Task Team[1] to further the CGD agenda, particularly advocating for the inclusion of CGD in development programmes and by formal data users when working towards implementation and monitoring the Sustainable Development Goals (SDGs).

As part of this effort, the GPSDD CGD Task Team designed a research effort with two specific objectives:

1. To produce recommendations on what types of CGD initiatives are best suited to different purposes, actors and needs, and what this means for their potential to scale and be combined with other data sources for implementing and monitoring the SDGs
2. To produce guidance aimed primarily at official stakeholders at all levels of government and

1. http://www.data4sdgs.org/initiatives/citizen-generated-data-task-team

Mayur Nayi, Tanzania Bora Initiative from Data Zetu / CC BY SA

international organisations on how to navigate and engage with different types of CGD initiatives

This report and an accompanying guide follow from that effort. They provide an overview of CGD methods, highlighting some prominent ways of generating CGD, and how each method relates to implementation, monitoring and achievement of the SDGs. What CGD methodologies work well for what purposes? What is the role of citizens in generating data, and what can data 'generation' look like? How are participation and use of citizen data organised?

Starting by drawing out general differences across CGD methods, the report discusses how the participation of citizens makes a difference to data quality and breadth and concludes with a review of benefits from using CGD for the SDGs, providing suggestions for how governments, international organisations, and civil society can better engage with CGD initiatives.

What is citizen-generated data? After reviewing research, reports and websites which employ the term, we encountered many different ways in which it was used with different connotations in different fields such as citizen science, citizen journalism, crowdsourcing, witnessing, civic participation, participatory design, patient involvement and social media listening.

These ways of envisaging CGD partly result from different traditions which pre-date the digital technologies upon which they often draw – from self-taught naturalists filling notebooks with observations in the 1980s, to 'social audits' which grew to prominence in the 1980s and 1990s.

CGD can be implicitly or explicitly contrasted with other ways of doing things and knowing things, such as official governmental information, professional journalistic reporting or institutionalised scientific inquiry.

Mayur Nayi, Tanzania Bora Initiative from Data Zetu / CC BY SA

CGD may be produced in ways which differ from professional knowledge production. CGD projects may involve means, methods and resources different from those available to their professional counterparts. Citizen groups may not have access to multimillion dollar equipment or legal mandates to standardise reporting within or across states. Instead they may involve paper surveys, inexpensive devices and web-based interfaces in order to accomplish their goals. As such they may be produced in ways which depart from established conventions to obtain the quality, interoperability and verifiability of data and sometimes abide by 'good enough' standards for operational use, different from those of established official professional statistics.

However, the fact that CGD projects may not always and immediately conform with these established practices of knowledge production does not mean they are less valuable. The data produced by citizens may also differ from official statistics not because they are produced through different techniques and technical tools, but because they focus on different issues and pursue alternative objectives.

The involvement of citizens can be understood in different ways in relation to professionalised institutions of knowledge production, such as:

- Enabling citizens and organisations to foreground, describe and help addressing problems that are otherwise unnoticed in existing data collections.
- Introducing political dynamics around problems, which may open up spaces for collaboration, but also face resistance if the problem is not acknowledged or de-prioritised by government.
- Producing evidence that may otherwise be ignored because it's politically awkward (e.g. challenging the views of established institutions and helping them to evolve with society).
- Collecting information which would otherwise be considered prohibitively expensive (e.g. nationwide wildlife counts).
- Gaining new perspectives drawing on local insights and expertise (e.g. geographical knowledge in

remote communities).

- Checking plans against experiences that are not otherwise readily available (e.g. ground-truthing and verification).
- Enabling civic involvement and participation (e.g. involving communities in surfacing and responding to issues).

While these differences can make it difficult to integrate CGD into the established system of national and international statistics, they also constitute a major source of innovation. CGD can extend and enrich professional data collection and analysis as well as generate new types of data and questions.

Furthermore, there are many ways of doing CGD, including different ways of configuring relations between each of the three parts of the phrase, 'citizen', 'generated' and 'data':

- **Citizens** may be enrolled as sensors, auditors, monitors, reporters, community members, observers, co-investigators, analysts or platform users.
- **Generating** may involve identifying, tagging, transcribing, compiling, mapping, describing, evaluating, quantifying, photographing, recording, translating, narrating, deliberating, writing, sensing, conceptualising or noticing.
- **Data** may result from a wide variety of devices, methods and infrastructures, including scientific instruments, paper surveys, online platforms, mobile phones, maps, satellite imagery and documents.

CGD projects emerge from different kinds of 'assemblages' of such methods, devices, materials and actors, which involve citizens in different ways and call on them in different capacities. Such configurations operate within different kinds of social, economic, organisational and cultural setups – from the platforms of for-profit companies (e.g. Premise[2]), to toolkits from community-driven non-profits (e.g. Public Lab[3]), to initiatives from public institutions (e.g. Statistics Canada).

In this report we draw attention to different ways in which CGD initiatives could contribute to progress around the SDGs. Rather than assuming that such initiatives can only be measured against institutionally established norms, standards and purposes, we have sought to empirically study them on their own terms. Our analysis framework and our case studies illustrate differences between CGD initiatives and examples how governments already engage with CGD. It highlights ways how governments can see CGD's fitness for different governmental purposes. A mapping of CGD initiatives against SDGs complements the analysis. We hope to spark new ways of thinking about CGD, and how it may contribute to implementation, monitoring and achievement of the SDGs. Further details on our process to gather and analyse CGD examples are provided in the appendices.

The report is accompanied by a guide to help interested parties select and engage with different CGD projects. Built around a questionnaire, it takes organisations and people through the process of identifying problems and data needs, provides a review of existing CGD methods, and lists ways to organise participation as well as outlining the resources needed to set up a CGD initiative.

2. https://www.premise.com/

3. https://publiclab.org/

Making sense of citizen-generated data

An early definition of "citizen-generated data" by CIVICUS' DataShift programme, defines it as:

"[...] data that people or their organisations produce to directly monitor, demand or drive change on issues that affect them. It is actively given by citizens, providing direct representations of their perspectives and an alternative to datasets collected by governments or international institutions."[4]

In this definition, participatory methods, problem focus, benefits for citizens, and active data production are seen as defining elements of CGD. The concept of CGD co-exists with citizen science, community-based monitoring, participatory geography, crowdsourcing, and others. Interestingly, each concept can stress different aspects of similar concepts. Citizen science may distinguish itself by emphasising its scientific rigour, with discussions revolving around methods, academic institutions, academic publishing or other concerns. Social accountability - or community policing - might look at similar initiatives but emphasise the role of data to represent or enable their rights, empowering individuals and communities.

CGD's wide definition might accommodate several of these concepts, but exclude some projects due to its problem orientation (some initiatives might be conducted out of pleasure, or for engagement and educational purposes). Different actors associate the term with different meanings, definitions, practices and projects.[5] It appears that CGD has particular traction amongst some groups - such as development organisations, international NGOs, some government bodies such as the European Commission,

4 The DataShift (2015): What is Citizen-Generated Data And What Is The DataShift Doing To Promote It?, Available at: http://civicus.org/images/ER%20cgd_brief.pdf

5 We conducted a method called 'search as research (Rogers 2009) to repurpose Google search engine results as a reference of the 'language' different actors use when speaking about citizen-generated data. This yielded 156 query results and shows which actors use the term in which contexts. See Gray, J. (2018). "Citizen generated data" according to Google Search, 30th July 2018. figshare. https://doi.org/10.6084/m9.figshare.7376243.v2

universities like Cornell University or Technical University Delft, and tech companies like ESRI, whereas it may be less common among other actor groups, or not resonate with their motivations and self-identification.[6]

For the purpose of this report, we acknowledge these differences and use the term 'citizen-generated data' widely in order to take a broad look at what these different initiatives do, and how they can be understood. We may surface questions such as: What does data generation look like? What is the role of citizens in generating data and how much agency do they have? Who are the citizens participating in CGD? How is CGD embedded in broader data infrastructures?

What does data generation look like?

CGD can include many methodical steps. Figure 1 provides an overview of some of the tasks that can be part of CGD initiatives. The tasks are derived from our case studies. Neither do we suggest a linear value chain model, as CGD initiatives can start with any of the tasks outlined. Nor do we suggest that the tasks listed are discrete. For instance, tasks can have the purpose to enrich or to analyse, depending on the question at hand and data involved. Instead, we wish to emphasise that CGD initiatives involve different actions, which can prompt different questions for their design.

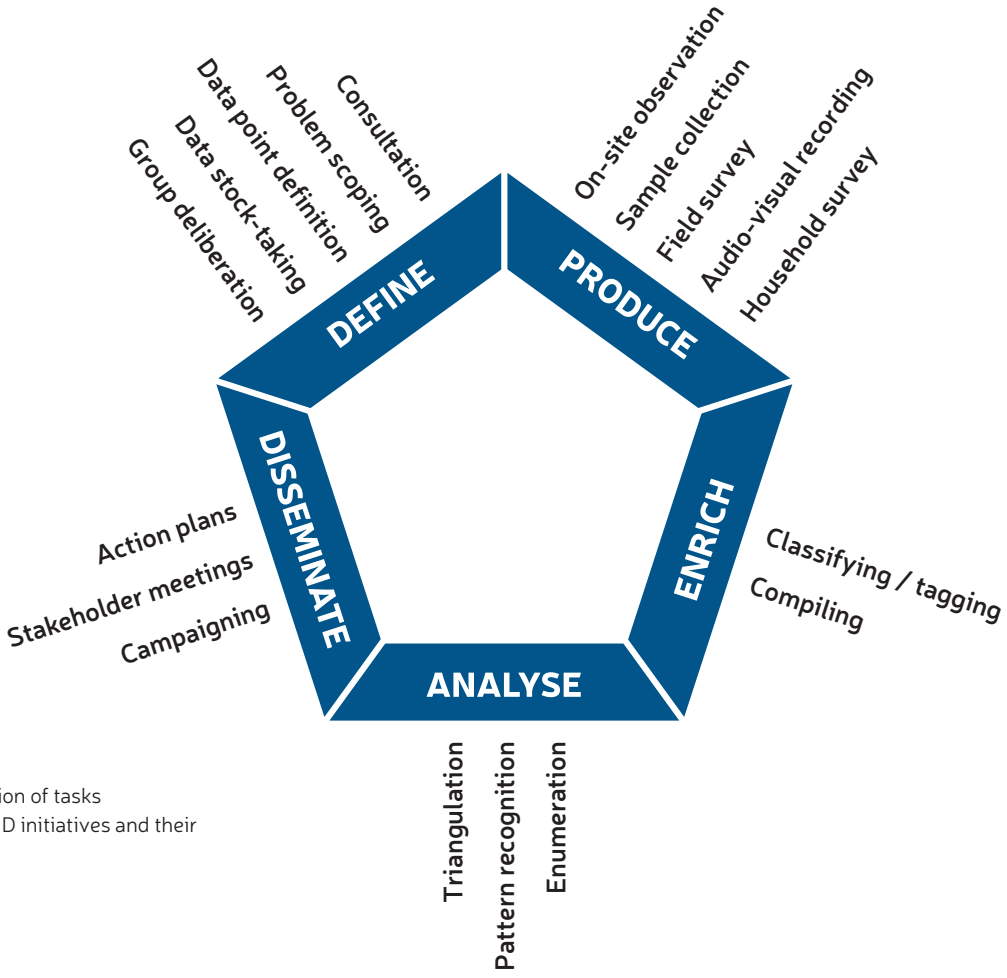


Figure 1: Illustration of tasks underpinning CGD initiatives and their workflows

6 While we were unable to study how actors self-identify through the names they give their initiatives, some research suggests that terminology is related to people's motivations. Terminology such as 'citizen sensor', 'volunteer' and others may alienate people and groups. See Eitzel, M.V., et. al. (2016): Citizen Science Terminology Matters: Exploring Key Terms. Available at: https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.96/

Such a view may be useful to illustrate some of the different steps which may be involved in citizen-generated data. For instance, some initiatives may gather new data in the field. Others might compile formerly unstructured data in the form of a structured database. Yet others may classify data, format, annotate, mediate, translate, or otherwise engage with it (see guide).

We may also study **possible dependencies** between these actions **and where CGD depends on other data infrastructures**. For example, CGD projects trying to measure government performance may **depend on** the availability of governmental **administrative records in order to define data points** and develop plans for on-site observations. Citizen-generated data projects may **enrich already produced data**, published by governments as open data or otherwise. For example, public service plans or facility performance can inform social audits to get an updated picture of public sector performance. Geographic data can be ingested into OpenStreetMap (OSM) to be further annotated.

Initiatives may distribute roles and responsibilities differently. Thinking in terms of flows may help understand how CGD initiatives organise and distribute participation among different actors (see guide).

'Active' data production: a matter of people's agency over data production

Citizen-generated data is characterised by people's involvement in the data production process. Several models[7] have been developed to note different levels of engagement along the data production process, ranging from contributory to co-created initiatives (for example, see figure 2). These models emphasize that people can get enrolled at different stages in a CGD initiative, and are an intuitive way to differentiate more or less participatory models. We caution that these models may imply that a maximisation of participation is better. They can also hide how people can participate, how questions are defined, how people collect data, and how individuals relate to other actors. Instead, we emphasise to consider different ways of participating (see below).

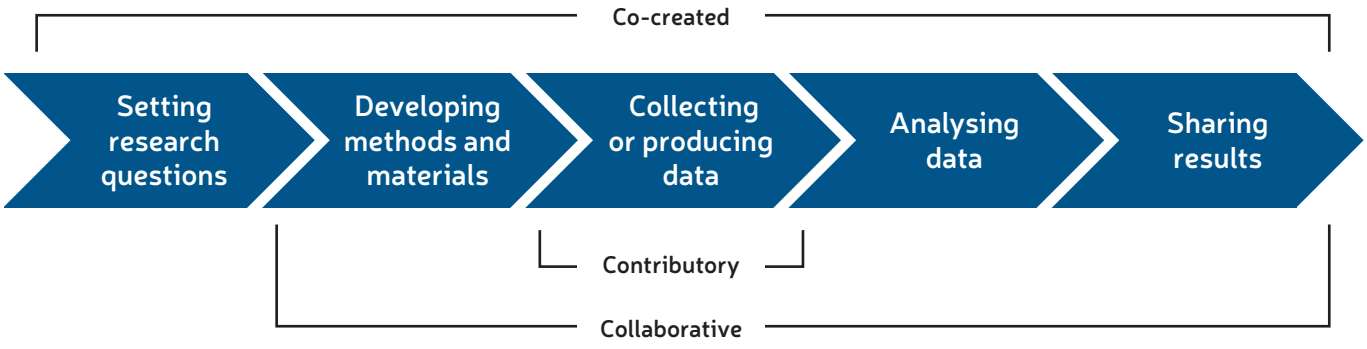


Figure 2: Stages of engagement in citizen science projects. Source: Shirk, J., et al. (2012)

7 Shirk, J., et al. (2012): Public Participation in Scientific Research: a Framework for Deliberate Design. Available at: https://www.ecologyandsociety.org/vol17/iss2/art29/
Haklay, M., (2013): "Citizen Science and Volunteered Geographic Information – overview and typology of participation". In Sui, D.Z., Elwood, S. and M.F. Goodchild (eds.), 2013. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice . Berlin: Springer. pp 105-122. Available at: https://povesham.files.wordpress.com/2013/09/haklaycrowdsourcinggeographicknowledge.pdf

Understanding active data creation as a matter of participation, it becomes clear that CGD is not only limited to methods that require people's attention or active involvement in the data production process. For instance, a group of citizens may **passively produce data** by using tracking devices or by installing air pollution sensor networks that gather data on their behalf. In these cases, people should be engaged in other stages of the data production process, such as defining the goals, or analysing data.

We may distinguish CGD from initiatives such as big data applications which track and analyse transactions and interactions as a by-product of people's interactions with digital media. This may include social media, call detail records or other data sources. In these cases citizens may not have much agency around what data gets collected, or when to stop the production of data.

What happens once citizen-generated data is produced?

Much attention is paid to define CGD based on who produces data and how. Less visible are debates around how CGD is used once individuals have produced it.

Several questions arise, including undue appropriation of knowledge and the **acknowledgement of data authorship**.[8] In the field of citizen science but also in work with indigenous communities, there have been questions how to best credit the authorship of individuals. One concern is that the more the data circulate, the more the credit will go to more powerful or visible institutions and not the citizens who generated the data.

Similarly concerns have been raised concerning the **ownership of data and data infrastructure** by which CGD can be produced. For example, Tomnod was once an open platform which was later sold to a private company. The platform uses crowdsourcing to annotate geographic information, a process similar to OpenStreetMap. Some authors raise concerns regarding not only accessibility of data, but also the idea of monetising free labour.[9] Other projects share their data only under licences with non-commerciality clauses to prevent monetisation of data and to ensure that everyone can freely access the data.

Ultimately, an increasing amount of research is dedicated to **privacy protection** and related data governance arrangements. This line of thought is concerned with the idea that CGD does not only render populations and their problems visible but can also make them more vulnerable[10]. Several ways to govern data access and use have been discussed elsewhere, ranging from licence agreements and data sharing licenses[11], to more institutionalised options like data trusts[12]. Different data protection principles are important to prevent unauthorised access to data, or undue and illegitimate data use, but also to ensure privacy by design or by default.[13]

8 The creation of citizen databases can bring with it questions of intellectual property. See Scassa, T., Chung, H. (2015): Managing Intellectual Property Rights In Citizen Science. A Guide for Researchers and Citizen Scientists. Available at: https://www.wilsoncenter.org/sites/default/files/managing_intellectual_property_rights_citizen_science_scassa_chung.pdf

9 Concerns include that CGD "is increasingly seen as valuable. Tomnod was recently bought by a large satellite operator, meaning now that the generous volunteered contributions by people trying to help in a disaster situation are at risk of being commercially exploited. For such situations clear transparency rules are needed." See also Foody, G., et al. (2015): Current status and future trends in crowd-sourcing geographic information. Available at: http://eprints.nottingham.ac.uk/29082/1/agi%20foresight-eprints.pdf

10 Taylor, L., Floridi, L., Van der Sloot, B. (eds.) (2016): Group Privacy. New Challenges of Data Technologies. Available at: https://www.springer.com/gb/book/9783319466064; McDonald, S. M. (2016): Ebola: A Big Data Disaster. Privacy, Property, And The Law Of Disaster Experimentation. Available at: https://cis-india.org/papers/ebola-a-big-data-disaster

11 See also the work of the Bristol Approach, available at: http://www.bristolapproach.org/

12 See McDonald, S., M., Wylie, B. (2018): What Is a Data Trust? Available at: https://www.cigionline.org/articles/what-data-trust

13 Privacy International (2018): The Keys to Data Protection. A Guide for Policy Engagement on Data Protection. Available at: https://privacyinternational.org/sites/default/files/2018-09/Data%20Protection%20COMPLETE.pdf

Participatory ambitions: Data literacy and institutional literacy?

Participation can take many forms and formats, from group consultations, to more substantive involvement. The educational and participatory ambitions of CGD projects can vary significantly. Some initiatives may consult groups about their priorities and design a project according to their needs. CGD initiatives may be an opportunity to "upskill" people by **increasing technical and data literacy**, instructing how to use tools or teaching methodological knowledge.

Other initiatives work to **enhance institutional literacy**, educating people about government institutions, and their rights in relation to these (e.g. environmental protection, right to basic services, etc.) as well as the role data can play to enable these rights. This may include a critical or creative engagement with official data practices (e.g. understanding drawbacks in existing governmental standards).[14] Likewise, governments may use different engagement strategies which can overlap with the educational goals of CGD initiatives. Governments might want to commission CGD projects to educate and engage the public or they might want innovate their institutional practices and open up institutional governance (more details can be found in our guide).

Citizen-generated data - fitness for purpose is key

An increasing amount of literature **rejects essentialist notions of data quality**. Instead, data can have many 'qualities' (see figure 3) which contribute to our understanding of whether a dataset is **sufficiently useful**.[15] Fitness for purpose means data is relevant and usable enough to provide answers to a particular problem.

CGD can complement existing information, increase the granularity or coverage of data collections, provide a comparative view and help cross-verify the accuracy of official data, or provide more updated information. In some cases, CGD may develop multi-purpose data, such as geographic basemap data, or pollution data which can be used for several types of analytics. 'Just good enough'[16] data is a moving target, but CGD can be designed to address questions of beneficiaries or decision-makers. Perceived usefulness, shared acceptance, and an appreciation for the value of data are key. Different processes have been practised and are currently experimented with to identify and ensure the relevance of CGD.

Figure 3 provides a list of quality dimensions that might become relevant in a certain setting for different people. It can be used to test existing data collections and identify how CGD enhances these (for more detail, see also the accompanying guide). Contextual quality dimensions depend on the use purpose and this is where CGD can make a difference by providing more relevant, complete, granular and timely data.

14 Gray, J., Gerlitz, C., Bounegru, L., (2018): Data Infrastructure Literacy. Available at: https://journals.sagepub.com/doi/10.1177/2053951718786316

15 Wang, R. Y., Strong. D. M. (1996): Beyond Accuracy: What Data Quality Means to Data Consumers. Available at: http://mitiq.mit.edu/Documents/Publications/TDQMpub/14_Beyond_Accuracy.pdf

16 Gabrys, J., Pritchard, H., Barratt, B., (2016): Just good enough data: Figuring data citizenships through air pollution sensing and data stories. Available at: http://journals.sagepub.com/doi/abs/10.1177/2053951716679677

Intrinsic quality attributes	Contextual quality attributes	Representational quality attributes
Trustworthiness	Relevance	Interoperability
Accuracy	Completeness	Representational consistency
Reliability	Granularity	Ease of understanding
Representativity	Timeliness	Structure

Figure 3: Data quality framework for CGD, derived from Wang, R. Y., Strong. D. M. (1996)

Intrinsic data quality factors such as trustworthiness, accuracy, reliability and representativity can be comparable to professional or scientific data collections provided task difficulty is appropriate and sufficient trainings are conducted.[17] As CGD methods expand the instruments, research protocols, and analytical tools to create data, they may not abide by statistical processes and standards, but may have their own quality assurance conventions in place.

Examples of representational quality attributes include interoperability (are two digital systems able to process different datasets?), representational consistency (do all data points within and across datasets use compatible and comparable sign systems?), understandability (can the data be interpreted by people?) and processability (is the data machine-readable?). Significant differences exist across CGD initiatives concerning the degree of standardisation of CGD databases and collection tools.

17 Kosmala, M., et al. (2016): Assessing data quality in citizen science. Available at: https://doi.org/10.1002/fee.1436

Case studies

In this section we explore the role CGD can play for implementation and monitoring of the SDGs by describing case studies that are occupied with particular types of information in the areas of:

- Geographic information
- Government policies, programmes and performance
- Environmental monitoring

Our selection of case studies cover prominent uses of CGD from the spaces of participatory mapping, social accountability and citizen science. The case studies also address related government administrations: geographic agencies, statistical offices, public service providers and environmental agencies.

For each case study, we highlight the situation in which a CGD method operated, describe the data generated, the workflows used and explore how the participation of citizens made a difference as opposed to relying on official data. Ultimately, each case study reflects on the role of CGD for implementing, monitoring and achieving the SDGs.

Our examples demonstrate that CGD can help address several SDG targets, increase disaggregation and granularity of information, and contextual information beyond siloed indicators. Beyond data for decision-making, the case studies demonstrate how CGD opens up space for participation, collaboration, education, and draw our attention to intricacies that might otherwise have been overlooked.

Case studies

Geographic information

Participatory mapping[18] enables individuals and groups to be involved in the process of making maps in a range of ways.

At one end of the spectrum, projects like OpenStreetMap (OSM) focus on geographic baseline information, including building footprints, infrastructure, or points of interest. Their objective is to build collaborative maps that can be used and on which they can overlay data. And at the other end, a plethora of geo-referenced applications have been developed to map violence during elections, plot people's opinions, highlight public service issues and to display photographs or other information on maps.[19]

The following case studies present some use cases of cartographic baseline information for the SDGs, focusing on OSM and the related Humanitarian OpenStreetMap Team (HOT), a large scale mapping community with roughly 140,000 contributors registered worldwide, and several chapters in place around the world.

18 It is sometimes also known as 'volunteered geographic information' (VGI), 'community-based mapping', or 'critical cartography'.

19 Some related projects include SafetiPin, Ushahidi, Caminos de la Villa, Harassmap, or Utunzi.

Geographic information

Humanitarian OpenStreetMap Team (HOT) Indonesia: Using participatory mapping to identify disaster exposure and prepare against disasters

Situation

In order for cities to plan and calculate risk exposure, urban decision-makers need information on building footprints and the number of houses that might be affected by a disaster as well as detailed infrastructural features. Granular data on the number of houses or infrastructural conditions may not be available which can leave decision-makers and planners in the dark as to the actual impact of a possible disaster.

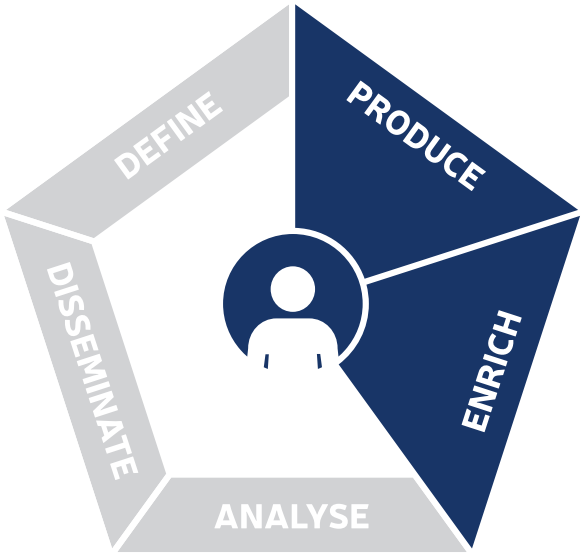
In Indonesia, for example, the national mapping agency does not hold the remit to produce cartographic information below the village level and cities may not have the capacity to tender data from private vendors. HOT Indonesia therefore partnered with government at various levels (including local and national disaster management agencies) to compile and annotate existing geo-data from basemaps and local surveys in order to calculate disaster exposure in different cities.

Workflow

HOT Indonesia has been running disaster risk mapping initiatives since 2011, using a common workflow used by the global Humanitarian OpenStreetMap Team. Firstly, the areas of intervention were scoped and available data gathered by volunteers from a range of sources including paper maps, satellite imagery, and GPS traces. Once these data were integrated back into the OpenStreetMap database, students annotated and traced maps via HOT's web editor tools. This resulted in updated basemaps enriched by hyperlocal building footprints. The information was put into the disaster risk exposure software InaSAFE[20] which allowed its users to better predict and understand risk of flooding.

20 See http://inasafe.org/

A flooded street in Jakarta. Photo by Seika / CC BY 2.0



According to HOT Indonesia's country manager Yantisa Akhadi, the resulting data provided a more accurate number of houses in sub-district areas, thus helping disaster management agencies in calculating the impact of disasters. Even if data was not entirely complete, the enhanced granularity was seen as a significant benefit. HOT's open mapping provides more detailed and accurate information of the number of houses, providing information for decision-makers who otherwise would have based their planning on intuition.

The project engaged people via workshops and mapping contests to spark interest in web-based and field-based mapping. The national disaster management agency BNPB helped run mapathons to gather basemap[21] information. The first iterations of the project let people contribute to the data generation in order to

21 Basemaps help orient the location on a map and can include various geographic layers, including parcels, administrative boundaries, relief, waterways, or satellite imagery.

rapidly generate data.

The project was able to create a large amount of sufficiently granular data in short time. Relevance and trustworthiness were ensured as HOT sought collaboration with government agencies to define a relevant data model. Several data quality assurance mechanisms - such as validation, technical supervision of field surveys and final data cleaning - ensured that the data was accurate, and that personal and regional biases were reduced. Ultimately, OSM data is interoperable with disaster risk software, readily machine-readable and can therefore be used by modeling systems.

What difference did the involvement of citizens make?

As Akhadi argues, for governments the prospect of cost savings and fast data collection were two main arguments in favour of experimentation. The speed of data gathering, and the overall quality of the data generated has motivated the National Disaster Management Agency (BNPB) to replicate the approach, and to run a pilot in disaster prone areas. Afterwards, HOT has collaborated with BNPB to embed the mapping approach further in government work processes.[22]

How does CGD relate to the SDGs?

This case study shows that HOT data has increased the granularity of available cartographic information on buildings and infrastructure. The data has been used as part of disaster risk exposure models. Gathering more granular information on the amount of buildings, the types of buildings and their condition helps strengthen the foundations of urban and national disaster planning, by highlighting new aspects of disaster exposure.

HOT data has increased the granularity of available cartographic information on buildings and infrastructure.

Using participatory mapping to improve disaster risk models may also help cities and human settlements at large to implement SDG 11.b[23]. This SDG target encourages cities and other human settlements to adopt and implement integrated plans and policies towards mitigation and adaptation to disasters, and to implement holistic disaster risk management at all levels. Participatory mapping may be particularly useful in cities lacking high-resolution survey data, or which do not have the means to update cartographic information about fast evolving settlements.

22 See https://crowdgov.wordpress.com/new-updated-case-studies/community-mapping-for-exposure-in-indonesia/

23 SDG 11.b: "By 2020, substantially increase the number of cities and human settlements adopting and implementing integrated policies and plans towards inclusion, resource efficiency, mitigation and adaptation to climate change, resilience to disasters, and develop and implement, in line with the Sendai Framework for Disaster Risk Reduction 2015–2030, holistic disaster risk management at all levels"

Statistics Canada: Using OSM to test the suitability of crowdsourced housing data for statistical purposes.

Situation

In Canada, there is currently no regularly updated unified register of housing. Statistics Canada (StatsCan) is in the process of developing a national Register of Buildings. The Data Exploration and Integration Lab, housed within StatsCan, explored the potential of crowdsourcing data to see if it can be used to complement and supplement other statistical information. Alessandro Alasia, Chief of the Data Exploration and Integration Lab, says: *"The idea of this pilot project was not merely to collect data - of course it was about collecting data - but the intent was primarily to test the idea of crowdsourcing. So the focus was on building related data but not on a specific type of information."*

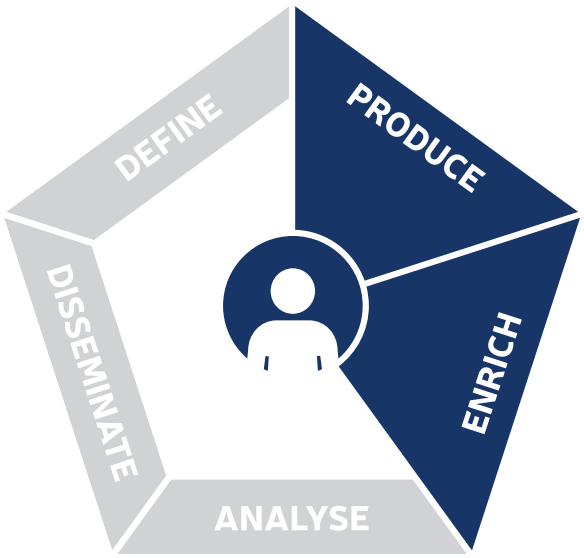
"The focus was on building related data but not on a specific type of information."

Workflow

StatsCan primarily used OSM to collect data on any type of building, including their locations and physical attributes. The agency wanted to have a collection tool ready quickly, so they used a customisation of OSM's iD editor[24] which they made available online. The StatsCan team simplified the OSM tagging system in order to show only tags related to information on buildings. OSM online documentation was used to train inexperienced mappers. They also sought collaborations with the

24 See https://www.openstreetmap.org/login?referer=%252Fedit%253Feditor%253Did

existing OSM communities in the region. This included to present project objectives to OSM's board of directors as well as the local OSM community in Ottawa in order to engage them in the mapping activities.



Canada's Ottawa-Gatineau region was the focus area of the mapping initiative.

StatsCan carried out a dedicated communication campaign, including universities, civic technology groups and journalists to increase press coverage. Once the project was over, StatsCan also presented the results of the project[25] at the Humanitarian OpenStreetMap Team Summit 2017[26] which took place in Ottawa in partnership with StatsCan.

In addition, StatsCan collaborated with the municipalities to unlock the necessary data for the mapping initiative, specifically for the import of municipal open data onto OSM. The OSM project took place from 2016 to early March 2018 and was funded through an internal initiative of StatsCan called the Big Idea Conference where innovative ideas within the agency were submitted and selected.

What difference did the involvement of citizens make?

Roughly 200 contributors took part in the initiative, with about 10 'power contributors' accounting for roughly 90 per cent of the data collected. StatsCan managed to map virtually 100% of their targeted area in four months, which exceeded the expectations of the team.

StatsCan ensured relevance by defining the data model themselves, before engaging with the Canadian OSM community. Ingesting maps onto OSM can sometimes cause legal interoperability issues if the government data is licensed under terms conflicting with OSM's Open Database Licence 1.0 (ODbL 1.0).[27] In this case, however, StatsCan government data was licensed under compatible terms.

StatsCan needed to spend as little time and energy as possible on the technical process and data infrastructure in order to concentrate on the outreach and engagement with Canadian citizens. As the official website of StatsCan states:

> *"As a project contributor, you can help create a free and open source of information on commercial, industrial, government and other buildings in Canada. We need your support to close this important data gap! Your work will improve your community's knowledge of its buildings, and in turn inform policies and programs designed to help you."*

25 See https://www.youtube.com/watch?v=EDljLFR7mr4&index=11&list=PLb9506_-6FMF-vBIRjvFziELghU6Y5BKb

26 See http://summit2017.hotosm.org/

27 See https://www.opendatacommons.org/licenses/odbl

Victorian houses, Montreal. Photo by Maëlick / CC BY-SA 2.0

According to Alessandro Alasia, this experience helped the agency realise that future crowdsourcing efforts need to be underpinned by specific skills to engage with the public, different from established communication approaches with the public. As Alasia states:

> *"The communication and engagement part is one of the key components of the project [...] We do engage with our clients and Canadians at large, but this was unique experience [...] The technology is important, but the area in which you really have to invest is outreach and engagement".*

How does CGD relate to the SDGs?

The project was considered to be a success and demonstrated to StatsCan that crowdsourcing can be accurate. It also inspired StatsCan to organise other follow-up projects. For instance StatsCan launched a project to crowdsource the price of cannabis throughout Canada before the legalisation of the cannabis market this October, for which no other way to collect data existed since it had previously been an illegal market. Currently the agency is exploring the possibility of using crowdsourcing to create a unified national register of buildings across Canada.[28] StatsCan is also examining how to bring together information from provincial and municipal sources to follow up on their crowdsourcing project.

The project was considered to be a success and demonstrated to StatsCan that crowdsourcing can be accurate.

The case shows how working with CGD may create the socio-technical infrastructure needed to handle and valorise 'underused' official data. Currently the agency is exploring whether they want to continue using OSM's methods and platform, or develop their own.

28 See https://www.statcan.gc.ca/eng/open-building-data/index

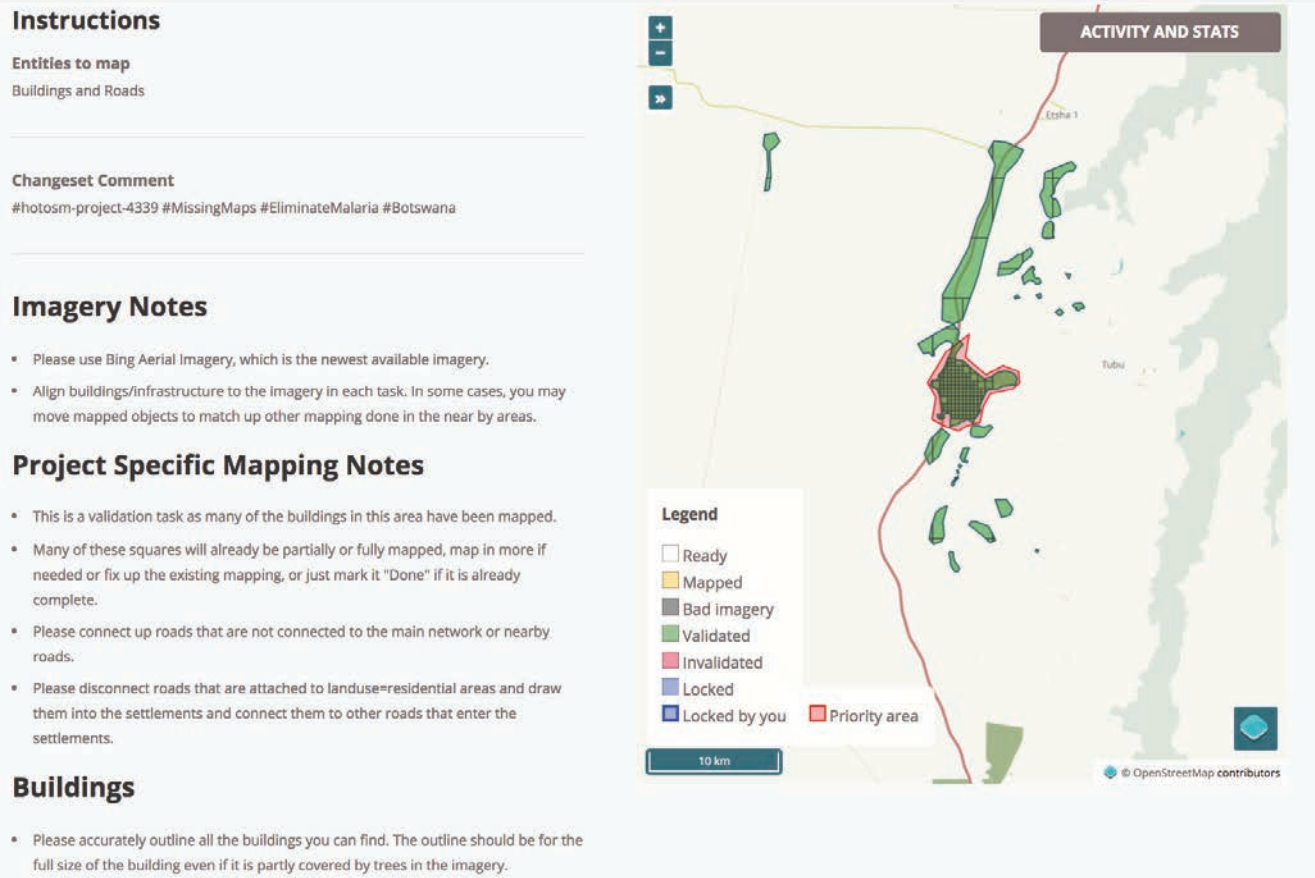
Humanitarian OpenStreetMap Team: Using OSM to plan a national malaria elimination programme in Botswana

Situation

Malaria is still endemic in several areas in Botswana. The Ministry of Health and Wellness has set a goal of eradicating the disease in the country by 2020. Through the National Malaria Programme (NMP), the ministry organises 'vector control'[29] activities - a mix of Indoor Residual Spraying (IRS) and long-lasting insecticidal mosquito nets - to fight the transmission of the malaria parasite. Through the District Health Management Team (DHMT), the Ministry needs to locate residential and community buildings to coordinate interventions. This includes understanding how all buildings in a district or village are being used, as well as their infrastructural condition.

Combining remote mapping with on-site field surveys, the open mapping methodology was adapted to understand people's exposure to mosquitos, not only by locating houses but also by checking whether these have complete walls and roofs. HOT led field mapping activities, working with the NMP and Clinton Health Access Initiative (CHAI) in six malaria endemic areas in Botswana. An initial remote mapping of buildings and roads using the HOT Tasking Manager and other OSM editing tools took place between November 2016 and February 2017. This was followed by another round of remote mapping to validate the initial mapping using newer available satellite imagery in March 2018. Further field mapping in selected areas was conducted between April and June 2018.

29 See https://www.who.int/malaria/areas/vector_control/en/

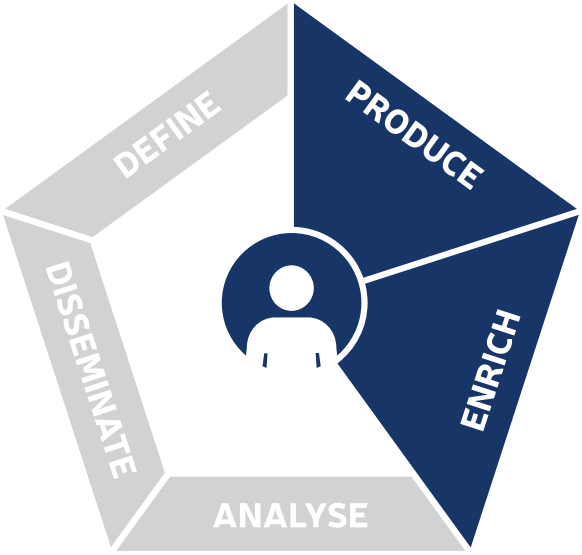


Malaria remote mapping area displayed in HOT's Tasking manager. Copyright: OpenStreetMap contributors, Source: Humanitarian OpenStreetMap Team

Workflow

The area of interest for the field mapping activities was provided by the ministry. The ministry was interested not only in the geolocations of buildings but also auxiliary information such as the number of rooms, how many rooms are painted and the quantity of sleeping spaces in a building. HOT inquired what data would be needed by the Ministry and, in collaboration with NMP and CHAI, developed a data model which was used for the field mapping activities. A partial data model can be found on a dedicated wiki page.[30]

30 See https://wiki.openstreetmap.org/wiki/Botswana_National_Malaria_Programme_(NMP)_Mapping



A data model was agreed that satisfied the requirements of the ministry, whilst requiring a feasible amount of work and time during the data collection exercise. David Luwata, Project Manager and Technical Advisor at HOT, said: "Combining our expertise with the ministry's needs, enabled us to agree on a data model which struck a balance between the needs, time and resources available." For instance, gathering outdoor observations requires less time than gathering information about indoor conditions, which usually requires conversations with the people using the building. Having iterated the method several times, HOT members knew how focussing on easily observable features could increase the reliability and scale of the data collection.[31]

"Combining our expertise with the ministry's needs, enabled us to agree on a data model which struck a balance between the needs, time and resources available."

Once the regions of intervention were scoped, NMP recruited members of the local communities they would operate in. The ministry signed short-term contracts with 59 community members to cover costs and as a means to provide work in these communities. The ministry hired these as 'research assistants' and ensured that those selected met the necessary educational requirements. Trainings were conducted in how to use the Open Data Kit[32], OpenMapKit[33] and OSMTracker[34] applications for mapping, as well as to familiarise the research assistants with workflows and the data model. To ensure data quality and a complete understanding of the OSM and open mapping workflows, HOT ran trainings for the research

31 The OpenStreetMap community has developed conventions of what data can be gathered in the database or not. A recent discussion revolved around the idea whether restaurant owners may be allowed to add client satisfaction, or other non-observable data points. Community members emphasised that OSM data needs to be physically observable, hence the narrower focus on geographic basemap information. Any additional mapping data can be added by using tools like Open Map Kit, Mapbox, or other tools.

32 See https://opendatakit.org/

33 See http://openmapkit.org/

34 See https://wiki.openstreetmap.org/wiki/OSMTracker_(Android)

assistants on how to do remote mapping, field mapping and field data cleaning.

Teams were formed which consisted of an experienced HOT mapping supervisor and four research assistants. Each team was assigned to areas and their data collection efforts were coordinated via social media and in-field supervision by the HOT staff. Once the data was gathered offline and moved onto an offline server, the data was validated and cleaned by selected research assistants and HOT staff by cross-verifying that the collected data matched the data model.

In total, the project took roughly seven months to conduct three mapping phases of the relevant six malaria endemic areas in Botswana, including the entire district of Chobe and five smaller villages in other sub-districts. According to David Luswata, these efforts could be replicated and expanded if additional funding was available. HOT operated on a contractual basis for these interventions.

What difference did the involvement of citizens make?

According to David Luswata, engaging people from local communities benefits projects as community members have better knowledge of the areas they live in. Community members also have personal relationships to other locals, and may be more trusted than outsiders. As the problem was defined by the ministry in the case of the NMP, HOT did not involve local community members in focus-group discussions to set targets.

How does CGD relate to the SDGs?

HOT recently delivered the data to the ministry which is using the maps generated and the data collected to plan future resource allocation including the distribution of nets and the 2018 round of IRS. Even though it is too early to review how the ministry has used the data, the case suggests that geographic information can be used to locate, plan and distribute health interventions. HOT's data model accommodates sufficiently detailed data on buildings and their infrastructural condition to inform health programmes. Participatory mapping can play a supporting role for programmes to reduce the number of malaria cases, therefore driving progress around SDG target 3.3 and SDG indicator 3.3.3 ('Malaria incidences per 1000 population').[35]

35 SDG 3.3: "3.3 By 2030, end the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases"
SDG 3.3.3: "3.3.3 Malaria incidence per 1,000 population"

Case studies

Monitoring and informing government policies, programmes and services

CGD initiatives may target governments and public sector activities. Initiatives could want to educate citizens about their rights to services, understand government performance, improve service quality or increase accountability within government. These initiatives might be driven from the bottom-up ('upwards accountability')[36] or led by governments to monitor their own actions ('downwards accountability'). What unites these approaches is their focus on government policies, programmes and services. The following section illustrates social audits, community scorecards and (household) surveys as the data generated from these methods is related to various SDGs which call on institutions to become more transparent and to enhance service delivery.

36 Peixoto, T., Fox, J. (2016). When Does ICT-Enabled Citizen Voice Lead to Government Responsiveness? Available at: https://openknowledge.worldbank.org/handle/10986/23650

Monitoring and informing government policies, programs and services

Black Sash: Using social audits and community scorecards to assess maternal and child care services.

Situation

In South Africa, public services such as health services do not reach all women and children. Some regions are underserved while other times women do not use services because they perceive their quality as low.[37]

Black Sash, a South Africa-based human rights organisation, developed a community-based monitoring (CBM) programme to assess health service performance, accessibility and the most pressing service issues experienced by women. To do so, Black Sash ran a social auditing programme and community scorecards. Through the social auditing process, community members learn about government services, their endowments and rights to services. Data on the real performance of services is also collected. Community scorecards enable community groups to assess government services based on what an ideal service would look like. Both methods result in action plans and agreeable commitments that communities and service providers can implement.

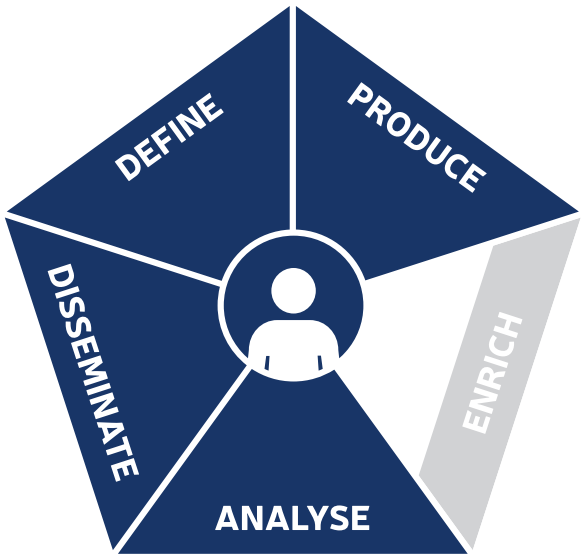
Both methods emphasise collaboration, deliberation, and dialogue between service users and providers. Black Sash implemented the social audit in 2010 in 74 clinics in 28 districts, and the community scorecards in 2013 together with several clinic committees[38]. As well as community based organisations they collaborated with government agencies like the National Department of Health (NDoH), health care workers, caregivers and district clinical specialists.

37 See https://www.blacksash.org.za/index.php/sash-in-action/community-based-monitoring/rmch

38 Clinic Committees are accountability institutions mandated by law to bridge between health service users and providers.

Workflow

Both approaches involve partnership building with local community-based organisations, governments, and public service providers.[39] To prepare the social audit, legal information, policies and service descriptions were gathered to understand what rights people have in relation to health services and to understand how government plans for services. From these documents, service performance indicators were derived related to accessibility (e.g. people's travel time to a facility), service management, provision (e.g. availability of drugs) and others. 86 community members living close to facilities were selected and trained to conduct surveys (they were called 'community monitors') with service users, as well as facility staff.



The data emerging from the social auditing had mixed quality in the first collection phases. Black Sash, in collaboration with Hivos and Code for South Africa (now OpenUp), simplified the survey by implementing closed instead of open questions and simplifying the interpretability of results. These changes also made it easier to do year-by-year comparisons.

Attention was paid to recruit community members in rural areas as well as urban areas to have a more representative vision of public services in different areas. This was important as Black Sash and its partners wanted local communities to conduct the surveys themselves, and to adopt the model in the future.

After community monitors finished the survey process, Black Sash's partner organisations gathered the results for different facilities. This information was translated into individual site reports, carrying the branding of the respective community partner. More aggregated provincial reports were also developed. Black Sash served as central coordinator, supervising the definition of the survey items and survey standards used to make the data comparable.

The community scorecard project included planning and preparation, scorecard meetings with

39 See https://www.blacksash.org.za/images/campaigns/rmch/Final-Black-Sash-Baseline-Report-04.04.2014.pdf

RMCH Action Group turning prioritised MNCH challenges into measurable indicators during the Community Scorecard meeting. Source: Black Sash

community and facility staff, as well as a multi-stakeholder meetings to present the findings. Trusted and respected community members and service staff were recruited to form 'action groups'. Forming these groups transferred ownership to communities and service providers to mobilise the larger community and other staff members to participate in the process. Communities and facility staff were split into separate demographic groups to identify the challenges they are facing with health services and to translate these into 'key indicators' or 'things they want to see changed'.

Each group collectively scored several criteria on a scale from one to five. Once the results were collected, broader stakeholder meetings were organised which included government representatives. From this an action plan was developed, laying out how services could be improved, who should be responsible and the timeframe for expected improvements.

What difference did the involvement of citizens make?

In both approaches, community members learned about the rights and services they are endowed with and how these are delivered in reality. According to the project leaders, communities involved in the project showed a heightened awareness of their rights to services after the intervention. Black Sash worked mostly with women as part of their social audit to educate them about their rights as well as

teaching them technological skills and survey methods.

Black Sash noted that a lack of stipends can negatively affected volunteerism and commitment to social auditing particularly if community members from poor communities are recruited who tend to prioritise paid employment opportunities.

Community scorecards are a method of collectively creating performance indicators representing people's most urgent problems everybody can agree upon. A limited set of measurable indicators must be found that reflects people's perceived problems and that everyone agrees with. The value of community scorecards can be that they provide a venue to let multiple realities surface, to spark discussion and disagreement around what counts as a good service and for whom. Yet, facilitation skills need to be high, and power dynamics within groups must be considered. For instance it was hard to get facility staff to speak openly in front of management representatives.

How does CGD relate to the SDGs?

The social audit method surfaced the fact that Emergency Medical Services (EMS) staff often don't reach communities quickly enough. The community scorecard process foregrounded reasons that women feel uncomfortable visiting health services. Social audits can provide data on actual service performance in the absence of available information, helping to achieve SDG target 1.4 which aims to ensure that everyone has basic access to services.[40] Social audits and community scorecards revert the perspective. Rather than telling us how many people in households have access to basic services (SDG 1.4.1), they get to the bottom of the challenges service users and facility staff members face in accessing and delivering on these services.

Social audits can provide data on actual service performance in the absence of available information.

There is potential for social audits to be considered as complementary to other administrative records to provide context on SDG 1.4.1. This SDG indicator refers to the "proportion of population living in households with access to basic services".[41] Data for this indicator is planned to be drawn from household surveys and censuses, administrative records of service providers, as well as from local governments where applicable. Social audits can provide information on the real status of public service delivery and provide updated data for service providers, particularly when service plans are not regularly monitored by providers.[42] This may be the case when governments don't include regular performance monitoring in their terms of reference with service providers. As social audits rely on the accessibility and analysis of

40 SDG 1.4: "By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance".

41 See also https://unstats.un.org/sdgs/tierIII-indicators/files/Tier3-01-04-01.pdf

42 Whilst this is not the case for our Black Sash case study, other projects working with social audits detected that governments may not have monitoring schemes in place for public service providers. See also the case of Social Justice Coalition (SJC) and Ndifuna Ukwazi in Cape Town: Gray, J., Lämmerhirt, D., (2017): Data and the City. How Can Public Data Infrastructures Change Lives in Urban Regions? Available at: https://blog.okfn.org/files/2017/02/DataandtheCity.pdf

government records and legal code, they benefit from public access to administrative data. Governments can support these initiatives by providing administrative records, thus also implementing SDG 16.10[43] and improving facility-level monitoring systems.

Both methods are examples of how citizens' voices can be heard in public decision-making to inform the actions of both government and communities, thus contributing to SDG 16.7.2[44]. Regarding SDG indicator 16.6.2 (proportion of population satisfied with their last experience of public services), community scorecards help not only to count, but also to translate information on people's perceptions into action plans and commitments at the facility level. Both methods compliment existing governmental accountability mechanisms.[45] For instance, the community scorecard pilot inspired NDoH to develop an 'ideal clinic' intervention, which included the suggestion to implement community scorecards in the overall management of health facilities.[46]

43 SDG 16.10: "Ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements"

44 SDG 16.7.2: "Proportion of population who believe decision-making is inclusive and responsive, by sex, age, disability and population group"

45 South Africa's Department of Monitoring and Evaluation has spearheaded an initiative to implement both social audits and community scorecards in government. See https://www.dpme.gov.za/keyfocusareas/cbmSite/CBM%20Documents/Framework%20for%20Strengthening%20Citizen-Government%20Partnerships%20for%20Monitoring%20Frontline%20Service%20Delivery.pdf

46 Information retrieved from an internal grant report Black Sash shared with the research team.

Monitoring and informing government policies, programs and services

Citizen Engagement Programme: Developing a standardised approach for community scorecards to assess school and health care services in Mozambique.

Situation

In Mozambique, a number of national NGOs had implemented social accountability programmes using community scorecards, social audits and community dialogues to improve the quality of health services and schools. These initiatives were focused on local facilities, but were not joined up so, while improvements and changes were seen locally, the knowledge gained did not trickle up to a higher level of government where it could have helped design policies and better allocate budgets to improve service performance.[47] To advocate for the improvement of service quality at the district, provincial and national levels, the Citizen Engagement Program (CEP) was initiated. The programme was a collaboration between COWI[48], Centro de Aprendizagem e Capacitação da Sociedade Civil (CESC), N'weti, the Institute for Development Studies, Kwantu, Oxford Policy Management and Save the Children.

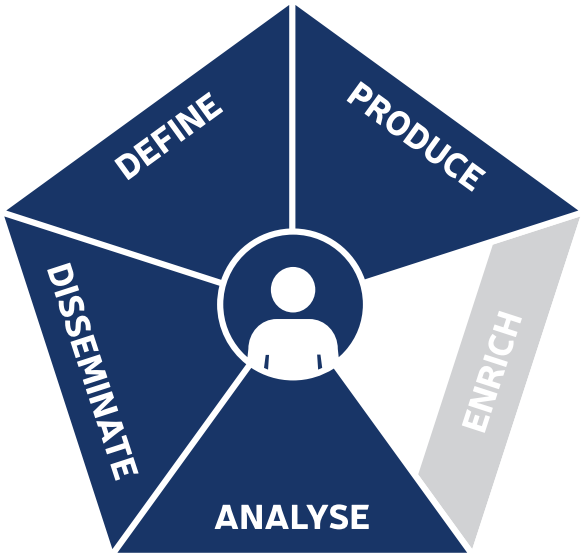
The programme sought to standardise the kind of data collected through community scorecards and how this data was coded. The intention was to create reusable data standards that would enable programmes to collect data in a comparable way. The programme also sought to develop a standard application that would provide a common data format and a shared data hub in which data from different programmes could be centralised. Implementing standardised data models into the process of community scorecards was aimed to foreground local performance issues in the health and education sectors with key stakeholders at provincial and national levels. Building the programme on top of community scorecards sought to educate service users of their entitlement to better health and education services. The programme ended in 2017 and the dataset was passed to CESC, a national NGO who now manage it.

47 Fox, J., (2016): Scaling accountability through vertically integrated civil society policy monitoring and advocacy. Available at: https://jonathanfoxucsc.files.wordpress.com/2011/11/fox_scalingaccountability_online5.pdf

48 Cowi is a Danish organisation with a longstanding presence in Mozambique.

Workflow

The programme collaborated with local civil society organisations to conduct two community scorecard cycles in eight districts across four provinces to conduct community scorecards. Local community members facilitated community sessions, working in around 120 health facilities and 200 schools. Around 20,000 citizens were involved in the community scorecard sessions.



Similar to the Black Sash case, the community scorecard model used a deliberative approach that engaged citizen groups in dialogue around problems affecting quality of service delivery.

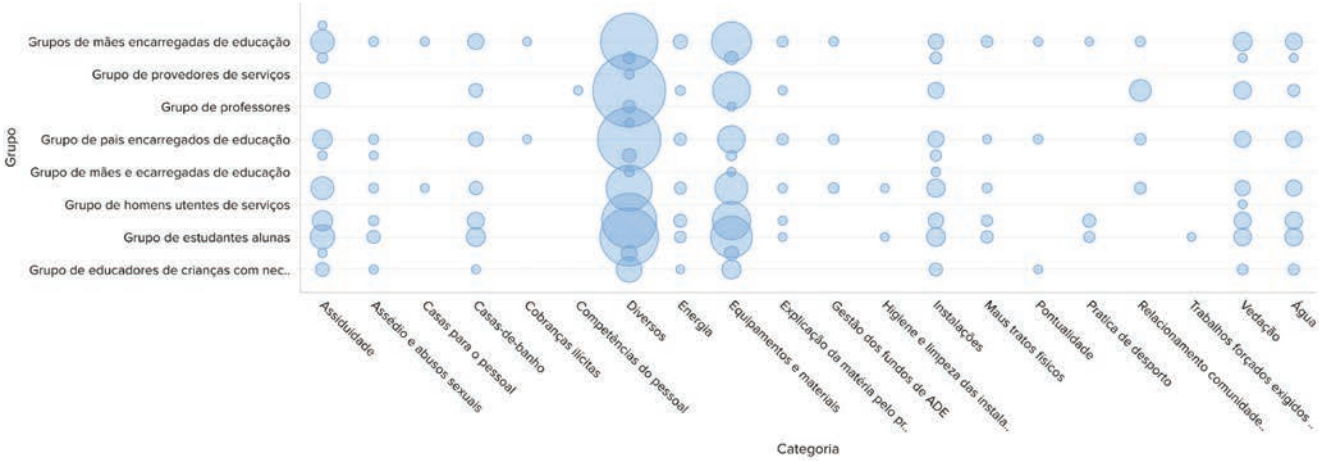
District level community-based organisations were contracted to facilitate the community scorecard process in the eight target districts. They engaged with communities using the facilities to identify and recruit people to take part in between five and six citizen groups for each scorecard cycle. Service users were grouped by sex, age and other criteria to understand how different groups experience services. This included at least one group representing facility staff.

Each session gathered basic information on the school or health facility in which the scorecard is being conducted, the groups rating the service and the ratings of the severity of service delivery issues raised by groups. A trained facilitator led the process, being aware of power dynamics in the group so as to provide a space in which each person felt able to raise issues they felt important. This helped ensure that the issues surfaced were reliable.

Action plans identified the most important issues and whether changes had been implemented. After the first cycle of the community scorecard process was completed, CESC and N'weti analysed the data gathered in all local interventions. A taxonomy of issues and actions was derived from the issues raised by citizens in each facility. Applying this taxonomy to granular information on problems helped detect commonly shared problems across facilities, the top problems raised by citizens, as well as the problems requiring action from higher level government. It linked issues and action categories on a facility level to national indicators, such as teacher absenteeism rates, and provided a way to relate issues raised by citizens - in their own words - to national level issues. As such, interoperability with national indicator systems could be ensured. This analysis informed research and advocacy campaigns.

What difference did the involvement of citizens make?

Like other community scorecard interventions, the programme facilitated dialogue between communities and local service providers. Citizens were invited to initiate interventions in the health and education sector and to hold public institutions accountable for their promises.



Graphic showing most common issues identified by each type of group in the education sector, Source: Citizen Engagement Program

The CEP aims to increase the availability of relevant information for citizens so they can better access and benefit from services delivered by the public sector, improve citizens' capacity to assess service delivery, and engage service providers to improve the quality of health and education services. The civil society organisations - CESC and N'weti - were able to aggregate emerging issues which could not otherwise have been detected. This informed decision-making within regional and national government.

How does CGD relate to the SDGs?

As CEP builds on similar methods as Black Sash, they are similarly relevant for the SDGs targets including SDG 16.7.2, SDG 16.6.2, SDG 16.10, and SDG 16.7.2.

Around 60% of the problems identified by the groups participating could be solved by the end of the programme. The CEP explored ways of collaborating with government. This was realised at local levels where citizen data was used to inform local planning and development processes.

In comparison, many health issues require attention over a longer timescale. A number of studies show that significant health outcomes can be achieved through use of community scorecards, including greater use of health services, increased satisfaction with health services, improved relationships between service providers and users, and a decrease in child mortality.[49] On a national level, the Ministry of Health requested information collected in the CEP data on citizen satisfaction to include it into the national health observatory. The observatory is a database which informs the ministry's policy and planning. CESC and N'weti successfully lobbied the ministry to include community scorecards as one approach to monitor health service quality from a user perspective. The ministry is currently planning to collaborate with local CSOs on using community scorecards in three provinces to identify well-performing health facilities. The goal is to allocate performance-based funding to health facilities so that they can improve the quality of their services. This system of community-based input to assess service quality and allocate funding accordingly relates to several goals expressed in SDG 3.8, such as achieving universal health coverage, with access to high quality essential healthcare services and quality medicines and vaccines.

49 See Gullo S., et al. (2017): Effects of a social accountability approach, CARE's Community Score Card, on reproductive health-related outcomes in Malawi: A cluster-randomized controlled evaluation. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171316

Monitoring and informing government policies, programs and services

Uganda Bureau of Statistics: Using a community information system to collect local survey data for local decision makers.

Situation

In Uganda, household surveys are primarily conducted on the national or district level and therefore do not adequately reflect the situation on sub-district administrative levels, such as parishes and villages. This became particularly significant after Uganda's devolution process in which local level government gained more responsibilities. In 1997, Uganda's Bureau of Statistics (UBOS) decided that available statistics and sampling sizes needed to be complemented by more granular information to help inform local decision making such as budget planning and service execution.

UBOS developed the Community Information System (CIS) as part of several development strategies.[50] Its goal was to increase the granularity of surveys and organise the collection of survey indicators on local levels, which are standardised and comparable across administrative levels and regions. To do so, UBOS expanded collaborations by providing training and guidelines to actors from civil servants all the way to parish chiefs, as well as individuals from village communities. This cascading training model was able to "give capacity to local people" and ensure "households and communities have access to their own information, and make use of it at that level."[51] CIS is spearheaded by UBOS, backed by a national statistical directorate and staffed with several statisticians supervising the programme. It includes mobilisation and awareness raising within local communities, training and data collection as well as database management of all results collected.

50 These include Uganda's Rural Development Strategy of 2005 as well as Uganda's Poverty Eradication Action Plan, and builds on development plans of district statistical capacity development. See also: https://www.ubos.org/onlinefiles/uploads/ubos/Amolatar.pdf

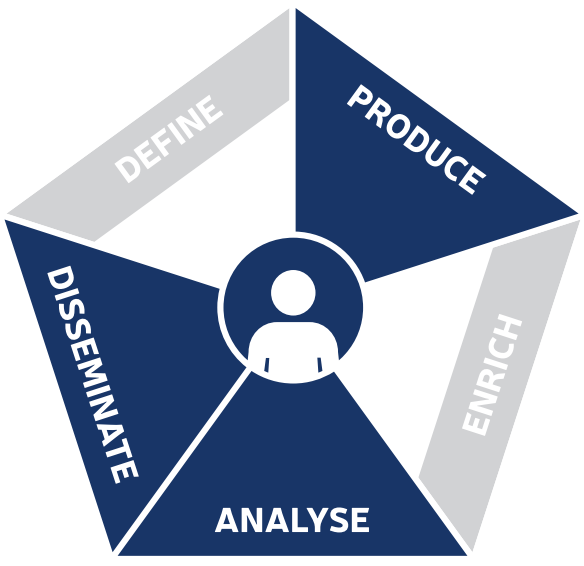
51 See https://www.ubos.org/onlinefiles/uploads/ubos/pdf%20documents/UBOS%20SSPS%202013-18.pdf

Mayur Nayi, Tanzania Bora Initiative from Data Zetu / CC BY SA

Workflow

UBOS 'trains the trainers' within government, meaning that one government agency assigns and trains a responsible person who stands in contact with and trains civil servants on lower administrative levels. The training programme draws on the Fundamental Principles of Official Statistics, as well as Uganda's National Statistics Indicator Framework.

The lowest level data is collected at the household, parish and sub-county level. Community facilitators are recruited from the villages and trained by parish chiefs to collect data via mobile phone apps or on paper. They conduct household surveys collecting socio-economic characteristics related to household welfare (household assets, income sources, agricultural production, access to water and sanitation, and others). The answers of each survey form are enumerated and aggregated in a village summary sheet and handed over to local chiefs. Using a sampling process adjusted to the size of the administrative zone ensures representativity.



Parish chiefs supervise community facilitators and aggregate information related to the entire parish. The data is summarised in parish registers and can be used by parish committees and others as a basis for local government planning.

Local household surveys adjust the sampling process to smaller administrative zones and can provide more nuanced information on household parameters and parish information. Bottom-up summarisation of data has provided immediate feedback at all levels of administration. Yet village data collectors

may make arithmetical errors when summarising survey results, and the workload may be bigger in some villages, which mean the process can take longer. As a representative of UBOS says, CIS has provided more relevant information to local decision-makers, and seeks to provide consistent, reliable, representative and accurate data abiding by governmental standards. CIS uses a standardised survey and hence does not allow for adjustments which could be relevant to foreground other local problems. Some local chiefs showed reluctance to engage with the system but reports by UBOS suggest that the system broadly has increased the knowledge base about household conditions and provided information for local decision-making.

What difference did the involvement of citizens make?

Community chiefs are elected representatives of their communities and are well placed to understand the problems specific to their communities. They may be more motivated to conduct surveys as the results are also usable by the chiefs themselves. Chiefs may also command familiarity and trust within their communities which can be helpful in getting people to conduct the surveys.

Government also sees CIS as was way to upskill parish chiefs and village data collectors by imparting knowledge around technical skills, statistical thinking and data analysis. However, the process is labour intensive and some parish chiefs have been slow to accept the value of adopting the approach in the long term.

As the initiative is conducted through village collectors who are trained by government, some people were concerned that data is used for taxation or other purposes. This highlights the importance of trust and its effects on the reputation of the initiative and its leading organisations.

How does CGD relate to the SDGs?

Usually information about the usage of services and the welfare of households across villages was not known by district government. Gathering village-level household data increases the level of detail about welfare, poverty, and access to services. What is the average level of education across villages? How accessible are basic services to individual villages? The patterns emerging from comparisons across villages can be used for planning within district level government. For instance, CIS can be used for used for local decision-making, to identify local government development priorities and to calculate resources needed for the coming fiscal year. This information could feed into several SDG targets and indicators relying on household surveys including SDG 1.4.1 which measures the proportion of the population living in households with access to basic services.

CIS can be used for used for local decision-making, to identify local government development priorities and to calculate resources needed for the coming fiscal year

Monitoring and informing government policies, programs and services

Data Zetu and HOT: Using geolocated household surveys to understand the accessibility of health services and HIV prevention in Tanzania.

Situation

In Tanzania, provision with health services does not always reach all community members. Some communities struggle to physically access health services, limiting their ability to get health advice or treatment. Sometimes health services may be prohibitively expensive, understaffed or not stocked with sufficient medicine for different health issues.

Community-based health services and frontline workers complement health services, for example providing testing and counseling services, support to vulnerable children or initiatives to counter gender-based violence. For example, to fight HIV infections, community workers develop interventions and coordinate their actions based on where transmission is most prevalent.

Data Zetu is a project that organises participatory work in Tanzania. It is implemented by the International Research & Exchanges Board (IREX) and five partners - Tanzania Bora, Sahara Sparks, HOT, SB Consulting (SBC4D) and Tehama Lab. The project's goal is "to help communities make better, more evidence-based decisions to improve their lives". This includes building skills and developing digital and offline tools that make information accessible to everyone. Data Zetu is funded by the Data Collaborative for Local Impact (DCLI) programme, a partnership between the Millennium Challenge Corporation (MCC) and the President's Emergency Plan For AIDS Relief (PEPFAR).

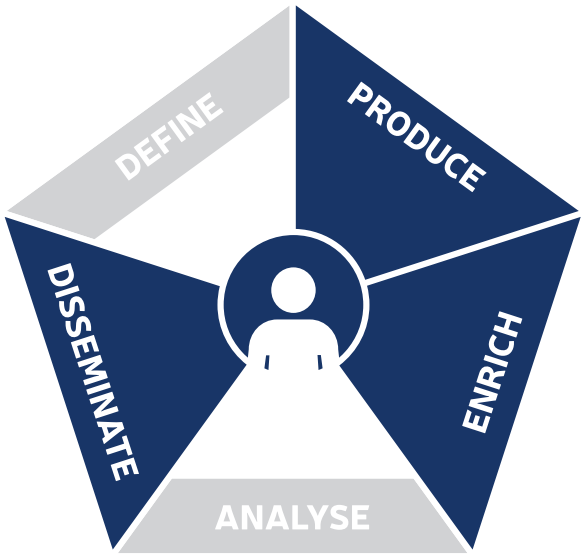
As part of this programme, Data Zetu organised several initiatives to improve how data is used in the provision and use of health services in Temeke, Mbeya, and Kyela districts. These districts are priority areas for PEPFAR due to their high HIV infection rates. Data Zetu seeks to inform decision making particularly in sexual and reproductive health. For instance, it helped local organisations fighting HIV/AIDS combining household surveys with geographic mapping to understand the distribution of health

services in combination with people's use of these services. In addition, Data Zetu organises so-called 'listening campaigns' - workshops consolidating problems from communities - to further understand problems around healthcare and other problems.

Workflow

Led by HOT, Data Zetu conducted a larger scale project to understand the accessibility to health care services. They gathered information on health service locations and basic infrastructure by partnering with HOT.

In addition, household surveys were organised in local communities, covering topics such as reproductive health and the level of accessibility to maternal care resources in communities. The data was collected for different districts and sub-districts in Tanzania. Survey data was geolocalised to enable follow-up analysis such as calculating the walking distance from individual households to health care services. The questionnaire included questions on actual health service usage, such as whether household members go somewhere to receive sexual reproductive health advice, whether elders and children receive access to free health services, and whether maternity medicine is available and its cost.[52] The project was designed in collaboration with local health officials, who are coordinating PEPFAR's DREAMS programme in the districts.[53] This ensured that data would be relevant to fighting HIV/AIDS.



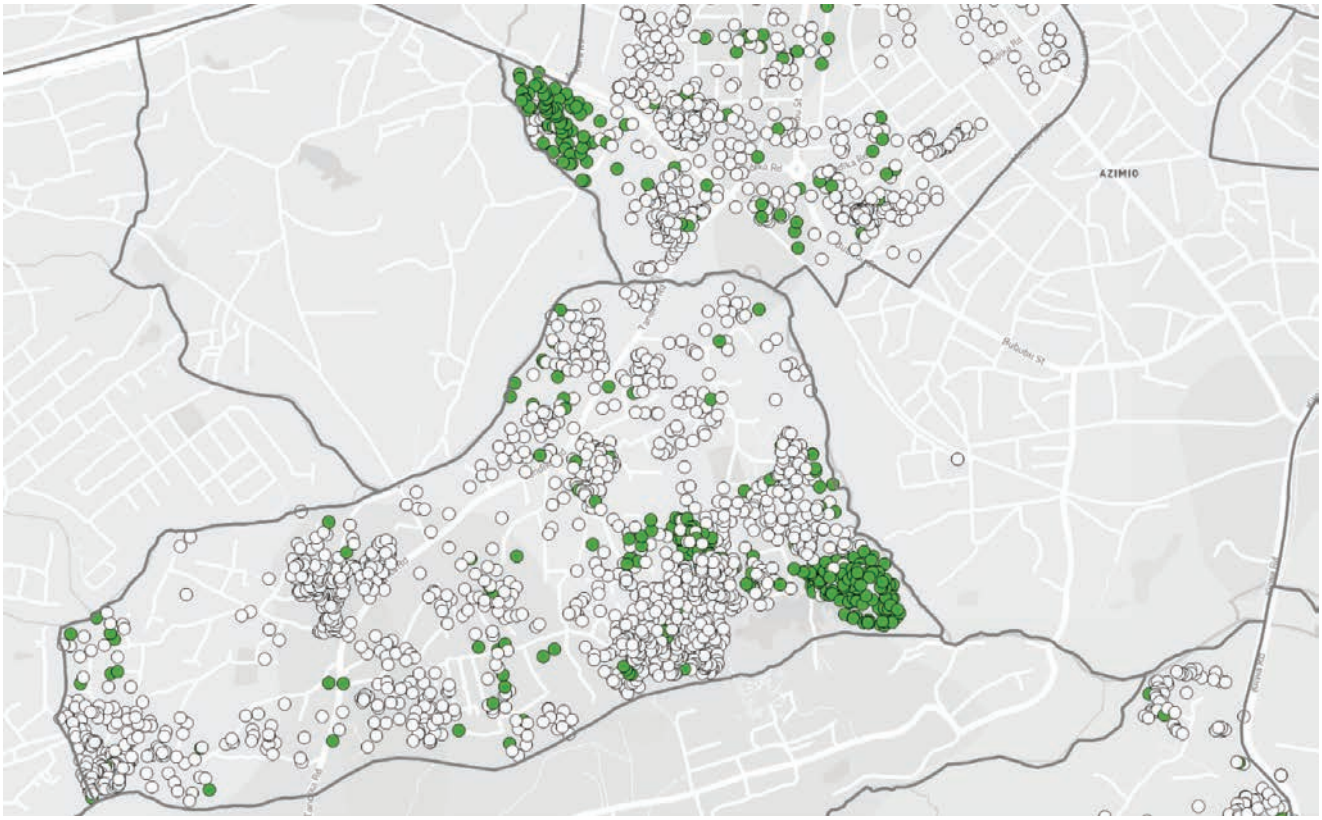
University students were engaged in the data collection process, carrying out the mapping as well as the surveys on Android phones using Open Data Kit, the OpenMapKit and the Kobo toolbox for surveys. Data was anonymised and analysed by the Data Zetu team using QGIS. Later the information was uploaded to the Humanitarian Data Exchange[54]. According to Samhir Vasdev, Advisor for Digital

52 See https://medium.com/data-zetu/community-generated-data-reveals-new-insights-about-gender-and-maternity-care-challenges-in-urban-24d0f20e53d6

53 See https://www.pepfar.gov/partnerships/ppp/dreams/

54 See https://data.humdata.org/

Development at IREX, the resulting dataset included survey results from more than 22.000 households with community members, as well as two million datapoints on health infrastructure and surveys. The programme serendipitously started mapping Shinas.[55] These are the smallest administrative units in Tanzania, grouping a small number of households under the care of a Mjumbe, a community-appointed representative. The map of Shinas had never been drawn before.



Hotspot map. Green spots indicate longer walking times to reach health services. Source: Data Zetu

In a later initiative, Data Zetu collaborated with a frontline organisation to collect information about hotspots of HIV transmission such as nightclubs and bars. They trained 300 field workers from community to map the location of night bars. The information enabled the frontline organisation to dispatch their community workers.

What difference did the involvement of citizens make?

Data Zetu does not frame its work around collecting citizen-generated data per se. Instead they try to "amplify better use of data amongst local leaders and local communities, whether that means: better management of data within a community organization, building skills to analyse and visualise data or communicate about data for advocacy or policy change, systems development around databases, and of course producing data to fill data gaps."

For the health survey initiative, Data Zetu engaged university students and community leaders, called wajumbe in the local Kiswahili language, to conduct the surveys. This helped build trust with

55 See https://datazetu.or.tz/wp-content/uploads/2018/06/Newly-revealed-Shina-boundaries-offer-unprecedented-hyperlocal-data-for-decision-makers.pdf

inhabitants and ensure that more accurate information is provided. Health issues, such as access to maternal, sexual reproductive and infant health centres, are often sensitive topics that people might feel more comfortable discussing with a local leader. During the household collection, a large part of the community members collecting data were women. According to Data Zetu, this helps women to learn about health care issues other women are facing. Data Zetu emphasises that the data created through their work are 'community insights'. Therefore, the initiative shares data with communities in several formats, from open data[56] to synthesised lists of issues to booklets or murals.

How does CGD relate to the SDGs?

Data Zetu's insights (information on households, the cost of medicines and services, as well as spatial accessibility of health services) relate to several of the goals expressed in SDG 3.8. This SDG target aims for universal health coverage, access to essential healthcare services and safe, effective, affordable essential medicines and vaccines. The data could help surface differences in health service accessibility due to travel time, as well as commonly faced issues such as underprovision of medicine at some facilities. According to Data Zetu, data from household surveys sparked ideas among the wajumbe to organise mobile clinics or community outreach in areas with low healthcare accessibility.

The data provided by Data Zetu is relevant for targeting the health interventions of community workers in locations with increased likelihood of HIV infections. The use of HOT's participatory mapping approach provided frontline workers with the necessary information. This may help achieve SDG 3.3 which seeks to end the AIDS epidemic by 2030 and which measures the number of new HIV infections.

The use of HOT's participatory mapping approach provided frontline workers with the necessary information. This may help achieve SDG 3.3 which seeks to end the AIDS epidemic by 2030 and which measures the number of new HIV infections.

Geographic information on shina boundaries was used to help a local hospital in Dar es Salaam digitise their patients register and modify the online registration system to add a field where patients choose their wajumbe/shina so the hospital betters understand "whom they are servicing". Data Zetu is currently exploring together with the National Bureau of Statistics, whether shinas can also be mapped in the remaining parts of Tanzania to inform the upcoming census. Collaborations with the National Bureau of Statistics helped vetting unofficial statistics in Tanzania. As Samhir vasdev says: "We have been balancing that with a good relationship with specifically the National Bureau of Statistics." As such, Data Zetu's data may more generally be helpful for increasing government's own data collections.

56 See www.bit.ly/dzinsights

Case studies

Environmental monitoring

Environmental monitoring is the practice of inviting citizens to participate in the observation, sampling and analysis of environmental parameters, be it wildlife observation or the monitoring of noise, air-borne pollutants or watersheds. Environmental monitoring has been described as one of the most prolific areas of citizen science, with a growing number of projects over the past decade.[57] Inviting citizens to participate in environmental monitoring has several well-documented cases in public policy and regulatory decision-making, and a long tradition in some countries such as Canada and the United States.[58] To illustrate this type of CGD, we discuss examples of water and air pollution monitoring, which have relevance for SDG 3.9[59], as well as showing how weather observation data have been used to monitor extreme weather.

57 Brett, A. (2017): Putting the Public on Trial: Can Citizen Science Data Be Used in Litigation and Regulation, Available at: https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1395&context=elj

58 For instance, the US Environmental Protection Agency has 'crowdsourced' citizen data for decades without calling it as such. This information has been used for decades in formal water quality assessments and is as an example how citizen-generated data can be verified and integrated into decision-making by regulatory agencies. See Brett 2017: Putting the Public on Trial: Can Citizen Science Data Be Used in Litigation and Regulation, Available at: https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1395&context=elj

59 SDG target 3.9: "By 2030, substantially reduce the number of deaths and illnesses from hazardous chemicals and air, water and soil pollution and contamination."

Case studies

Environmental monitoring

Atlantic Water Network: Monitoring the quality of open water sources in Canada.

Situation

Assessing water quality is necessary for many public policy areas, such as regulating traffic in waters and being able to assess the violation of water protection standards. In Canada, many watersheds are not monitored given the sheer size of the country. A 2017 World Wildlife Fund (WWF) report[60] concluded that Canada still lacks the significant amount of environmental data required to effectively inform public policy making. On the other hand, environmental monitoring initiatives that monitor water and wildlife are on the rise.[61] The Atlantic Water Network (AWN) is a Canadian network of water monitoring organisations comprising of several Atlantic Canadian provinces. AWN's mission is to build capacity among water monitoring organizations and to close "the gap of the extreme water quality data deficiency that exists across Canada".

Workflow

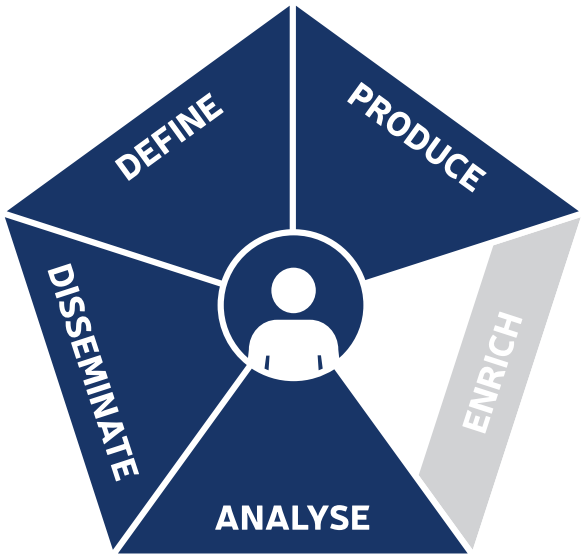
AWN works with diverse groups, ranging from entirely volunteer-led initiatives of concerned citizens to government organisations with full-time staff whose daily job involves water monitoring. Each organisation decides the scope of the area they cover and which data they want to upload onto the platform. Many of AWN's community-based organisations have years' worth of water quality data, often being collected during the summer field season. Collecting water pollution parameters can answer a

60 See http://assets.wwf.ca/downloads/WWF_Watershed_Reports_Summit_FINAL_web.pdf as well as https://www.thechronicleherald.ca/news/local/wwf-president-says-fresh-water-data-lacking-atlantic-canadas-in-particular-242499/

61 Carlson, T., Cohen, A. (2018): Linking community-based water monitoring to water policy: Perceptions of citizen scientists. Available at: https://www.sciencedirect.com/science/article/pii/S0301479718304596

range of questions and is therefore interesting for many diverse groups.

AWN lends accredited technology[62] via an Environmental Monitoring Equipment Bank and trains communities on how to use approved protocols. It also calibrates tools and offers an open access data hub, the Atlantic DataStream[63], which incorporates more than 5000 different parameters.[64] AWN focuses on a limited set of water quality parameters and trains people in using professional protocols which are also used by the Canadian government. Indicators of water quality include pH, temperature, dissolved oxygen, conductivity, salinity and total dissolved solids. Members of AWN stress the simplicity of measuring these parameters. "Once you do the appropriate training, people can easily find out these different variables within the watershed themselves."



What difference does engaging citizens make?

AWN emphasises that engaging communities with their local environment "is not only a tool to get more data online but it is also a tool to help people get more interested, engaged and passionate about their environmental surroundings." Collecting data would only be a small piece of the puzzle in their work. One strength of citizen-based monitoring efforts lies in mobilising local communities in regions governments could not reach otherwise, or which would be prohibitively expensive to cover by government scientists. As Jessie Smith, Project Officer at AWN, says, the Northwest Territories of Canada "is a lot more geographically isolated, so community-based water monitoring is essential because government bodies are not going to go out there very often to collect data." Government scientists rarely visit these isolated places. This is also a reason for the respective provincial government to include community-based water monitoring in their data collections.

62 They loan a so-called WET-Pro field kit, a professional YSI multiprobe which is calibrated by the AWN team to gather accurate measurements.

63 See http://atlanticdatastream.ca

64 Members of AWN emphasise the role of user-friendly interfaces such as excel sheets, which people are already familiar with.

Scatter plot visualisation, built into Atlantic DataStream, Source: Atlantic DataStream.

Community work is required to establish trust, in particular with indigenous communities who might hold relevant data, according to Jessie Smith. "The challenge is finding a way to integrate indigenous ways of knowledge into scientific databases. Indigenous communities have long experienced the appropriation of data for exclusive scientific endeavours, meaning that the inclusion of traditional knowledge within open access scientific databases can be undesirable for indigenous communities."

AWN publishes all community data on Atlantic DataStream, a technology spearheaded and implemented with the support of the Gordon Foundation. It enables communities to share only the data they are comfortable contributing. The data hub helps balance reuse via open access with the fears of appropriating indigenous knowledge by documenting provenance of data and requiring attribution of the data.[65] One of the main principles of Atlantic DataStream is 'ethical open access'. All data stored on the platform is made available in a timely manner for research, information, and educational purposes.

How does CGD relate to the SDGs?

The most important aim of AWN is to provide data that can inform public policy around environmental conservation and climate change adaptation. Currently the information aggregated on Atlantic DataStream is not used by government. AWN is working with provincial and federal governments in order to digitise and publish historical data from government and community-based water monitoring groups on the data hub.

CGD on water pollution affects several SDGs at once. It can for example help monitor SDG 12.4[66] by

65 In addition, funding from the RBC Foundation and the World Wildlife Fund Canada, the Gordon Foundation supported Atlantic DataStream with the implementation of Blockchain technology to increase the security of freshwater data while decentralising the tracking of data uploaded.

66 SDG 12.4 "By 2020, achieve the environmentally sound management of chemicals and all wastes throughout their life cycle, in accordance with agreed international frameworks, and significantly reduce their release to air, water and soil in order to minimize their adverse impacts on human health and the environment."

testing water quality of watersheds which are connected to sources of hazardous waste. This may provide sufficient evidence for follow-up monitoring or other actions by government.

Water monitoring may also directly contribute to monitor SDG 6.3.2[67] which estimates the number of open water bodies with good water quality. CGD can capture core parameters set by the UN and add complementary data to the sampled water sources that are proposed by the official SDG framework. The custodian agency UN Environment even encourages the use of CGD, stating that "additional [...] data from [...] citizen-science monitoring programmes can be used to supplement the available authoritative monitoring data, provided they are authorised by the national reporting agency."[68]

CGD on water pollution affects several SDGs at once. It can for example help monitor SDG 12.4 by testing water quality of watersheds which are connected to sources of hazardous waste. This may provide sufficient evidence for follow-up monitoring or other actions by government.

Water pollution monitoring might also be combined with surveys to help understand where people access drinking water and identify whether open waters are safe to drink from (relating to SDG 6.1.1).[69]

Ultimately, citizen-based water monitoring can help implement public policies to support SDG indicator 15.1.2[70]. Research further documents that citizen-based water pollution monitoring data can be used for public policy purposes.[71] Local governments can also use CGD as a "mechanism by which individuals and communities can meaningfully contribute to decisions and directions about water and sanitation management"[72] thereby directly supporting the implementation of SDG indicator 6.b.

67 SDG 6.3.2: "Proportion of bodies of water with good ambient water quality".

68 See https://unstats.un.org/sdgs/metadata/files/Metadata-06-03-02.pdf

69 SDG 6.1.1: "Proportion of population using safely managed drinking water services". https://unstats.un.org/sdgs/metadata/files/Metadata-06-01-01.pdf

70 SDG 15.1.2. "Proportion of important sites for terrestrial and freshwater biodiversity that are covered by protected areas, by ecosystem type."

71 Carlson, T., Cohen, A. (2018): Linking community-based monitoring to water policy: Perceptions of citizen scientists. Available at: https://www.sciencedirect.com/science/article/pii/S0301479718304596; Clearing the path

72 SDG 6.b "Support and strengthen the participation of local communities in improving water and sanitation management." See https://unstats.un.org/sdgs/metadata/files/Metadata-06-0B-01.pdf

Case studies

Environmental monitoring

Science for Change Movement: Measuring air pollution levels in Pristina, Kosovo.

Situation

Kosovo suffers from heavy air pollution, according to measurements by the U.S. Embassy[73], with air pollution sometimes hitting multiple times the levels deemed acceptable by the World Health Organisation (WHO). This is important because of the linkages between pollutant levels and diseases such as cancer, heart diseases, asthma or chronic bronchitis. High emission values are due to Kosovo's energy mix relying strongly on fossil fuels, non-modernised heating systems in people's homes, as well as traffic congestion. As part of negotiations with the European Union, Kosovo's Environmental Protection Agency (KEPA) has recently installed air pollution measurement stations across the country. But assessments by Kosovo's Supreme Audit Agency and the European Union question the accuracy of governmental measurements, and highlight the lack of timely and user-friendly information on air pollution.[74] KEPA installed a real-time air pollution monitoring system, but it is still being tested.[75]

To sensitise the public and to enable actions and campaigning around the problem, a number of partners have teamed up on 2014 to co-found Science for Change Movement, such as: UNICEF Innovations Lab Kosovo, Dr. Dan McQuillan, NGO Peer Educators Network, and Transitions Online. It consists of a self-organised collective of younger people, as well as three committees that organise educational work, monitoring and research methodologies for air monitoring, as well as campaigning and mobilisation.

73 See https://airnow.gov/index.cfm?action=airnow.global_summary#Kosovo$Pristina

74 See http://www.zka-rks.org/wp-content/uploads/2018/05/Report_air-quality_eng.pdf, as well as https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/key_documents/2016/20161109_report_kosovo.pdf; and https://eeas.europa.eu/delegations/kosovo_fi/42001/Kosovo%20and%20EU%20held%20the%20second%20SAA%20Sub-committee%20on%20Transport,%20Environment,%20Energy,%20Climate%20and%20Regional%20Development

75 See http://kosovo-airquality.com/secure/index2.html

Driving in Pristina, Kosovo. Photo by David Jones / CC BY 2.0

Grounded in the ideas of environmental justice and environmental racism, the group champions participatory air pollution monitoring as a way to "democratise air". This idea includes to make an otherwise invisible problem tangible, understandable, actionable and campaignable for young people or those who are particularly vulnerable to air pollution. The approach aims to spark thinking about air pollution and its sources and to show how air pollution relates to existing institutions, laws and public services. A representative of the group stresses that young people are engaged in all activities, from the selection of sites to take air pollution measurement, through to analysis, the setting up of action plans and campaigning.

Workflow

The Science for Change Movement has organised several air pollution measurement initiatives, based on pilot measurements to identify air pollution patterns in people's usual living environments, in Kosovo's capital city, Prishtina. The group calls this idea a way to create "measurements embedded into society". The collective made the decision to take long-term measurements in regular intervals in targeted locations with particularly high air pollution concentrations, namely an elementary school in Pristina and areas near the coal-powered power plants.

In the course of its measurement activities the collective was approached by Europe's Horizon2020-funded "Making Sense"[76] consortium. Difficulties with the first generation of their Smart Citizen Kit taught the collective that expectation management about the quality of the measurement tools is key.

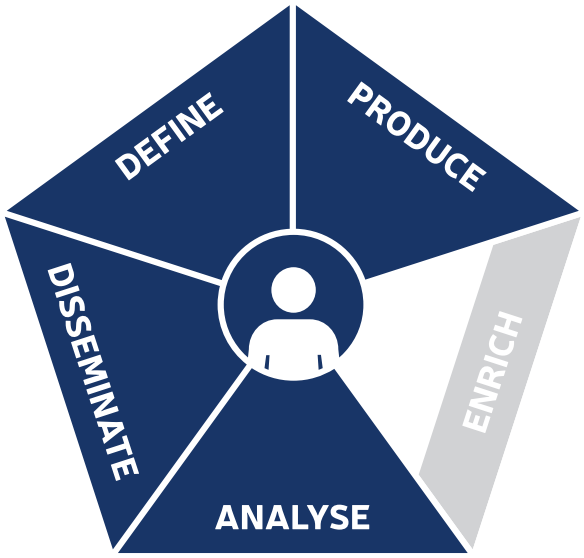
76 See http://making-sense.eu/

Inaccurate data can discourage people who are motivated by spotting pollution anomalies. Device calibration played a key role for the collective and its activities. Firstly, the requirements of calibration are part of the scientific education of participants and key aspects differentiating citizen science from other kinds of activity. Calibration enables the project team to make clear statements about the reliability of the measurements so that they can be used for campaigning. As Ron Salaj from the Science for Change Movement says, "calibration recognizes that policy impact and community engagement are related to trust in the sensor data". During the activity the following digital devices have been calibrated: Airbeam and Dylos DC1700. The calibration of the devices was done in close collaboration with the US Embassy in Kosovo.[77] In every case, the data was interpreted by scientists and taken back to the collective to develop strategies for campaigning and to organise workshops with young people. The data was published on GitHub[78], a collaborative infrastructure for developers to share code, and an EU website.[79]

What difference did the involvement of citizens make?

The collective emphasises the need to move beyond mere data production in order to turn data into actionable, campaignable data. Engaging young people helps 'democratising air' by enabling people to analyse information, discussing and reflecting upon the results, activating people to discuss how pollution affects their daily lives and health, and making spaces for people to engage around the otherwise invisible problem of air pollution.

The project combines data collection with analyses of the policies, regulations, commitments, and government information systems around air quality, a depth of engagement similar to social audits and other social accountability strategies. In 2016, measurements from the U.S. Embassy and the Science for Change movement sparked public protests.[80]



77. This included Airbeam Dylos DC 1700 sensor technology, certified by the US EPA. As the EU's Citizen Sense website notes "the Airbeam Dylos DC1700 is the latest iteration of the DC1100 which has a long track record of being used for low cost indoor PM measurements. The readings have been found to correlate well with the more expensive BAM monitors used for statutory measurements." http://making-sense.eu/campaigns/prishtina-air-quality/

78 See https://github.com

79 See http://ajriprishtines.info

80 See http://kosovotwopointzero.com/en/united-air-pollution

How does CGD relate to the SDGs?

The case of the Science for Change movement showed the political nature of data collections, and how Kosovo's Environmental Protection Agency did not accept the data as accurate, even though the group could reproduce the U.S. Embassy's records. The Science for Change network questioned the credibility of the environmental agency, based on EU assessments (see above) and argued that government has no accurate baseline to cross-verify and reject their data.

Similar cases of crowdsourcing air pollution data suggest how collaborations between government and crowdsourced data collections may support such efforts. By comparison, a recent citizen-generated large-scale air monitoring project in Flanders distributed chemical air quality sensors to more than 20,000 people. The goal was to use calibrated sensors to test predictive air quality models used by Belgium's environmental protection agency. First quality-assured results suggest deviations from existing air quality predictions and exceeding air pollution concentrations. Using tested sensors could help surface anomalies and hotspots, and possible reasons for air pollution, including building canyons and lack of ventilation, traffic density and traffic flows in urban areas and villages. These are now further tested and validated.[81]

In the US, local communities have used similar approaches to monitor air pollution in their immediate environment and located air pollution concentrations, based on accredited tools and protocols. This informed the United States Environmental Protection Agency to verify and act upon exceeding air pollution concentrations.[82] This suggests that, if accredited tools and data collection protocols are used, citizen-generated air pollution data may provide information to cross-verify existing data, which can help detect possible sources of air pollution, and may inform follow-up data collections by government. While being only one element to understand air pollution, this may help inform strategies around SDG target 11.6[83] in cities and complement local monitoring networks[84] in cities to support monitoring SDG indicator 11.6.2.

If accredited tools and data collection protocols are used, citizen-generated air pollution data may provide information to cross-verify existing data, which can help detect possible sources of air pollution, and may inform follow-up data collections by government.

81 See https://curieuzeneuzen.be. Information based on an interview with a project representative.

82 McElfish, J.; Pendergrass, J.; Fox, T: (2016): Clearing the Path: Citizen Science and Public Decision-Making in the US. Available at: https://www.eli.org/research-report/clearing-path-citizen-science-and-public-decision-making-united-states

83 SDG 11.4.: "By 2030, reduce the adverse per capita environmental impact of cities, including by paying special attention to air quality and municipal and other waste management."

84 The custodian agency World Health Organisation suggests following data sources for SDG 11.6.2: city-based pollution monitoring networks, satellite remote sensing, topography, and information on local monitoring networks. See also https://unstats.un.org/sdgs/metadata/files/Metadata-11-06-02.pdf

Case studies

Environmental monitoring

WOW Australia: Using national crowdsourcing of weather data to better understand extreme weather conditions.

Situation

Australia is continuously facing extreme weather conditions such as droughts. Some regions are particularly affected, including western New South Wales. The Australian Bureau of Meteorology (in the following the Bureau) is collecting weather-related data in Australia on a continuous basis to understand patterns of extreme weather and to prepare regions accordingly. Climate models gather data from multiple sources such as weather stations, satellite imagery, rainfall gauges, existing forecast models and cameras installed on the ground.

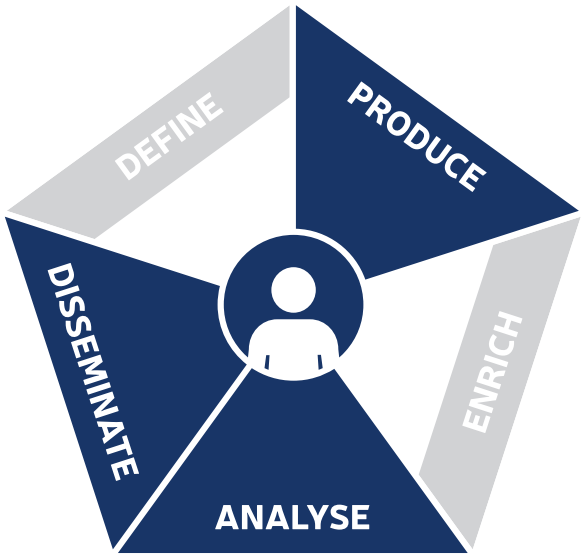
To collect crowdsourced weather observations, the Weather Observation Website (WOW) was developed as a central point to aggregate data. “We are more and more interested in citizen scientists tracking these events for us” said a representative of the Bureau. They also encourage farmers to install weather stations on their land and to send the data to WOW so that the Bureau can pull rainfall data from the website and incorporate it in climate models.

According to the Bureau, WOW is for everyone interested in weather who wants to share and exchange information in an online community. It is a resource for weather enthusiasts, educators, farmers, gardeners and anyone with an interesting weather photo or weather fact to contribute.

Workflow

WOW is a web-based service which allows anyone to enter CGD either as automatically gathered and regularly updated data from weather observation stations, or as field-based, manual observations. These observations can be submitted in form of photographs or by sending one-off 'quick' reports

such as "it is hailing now", "it is foggy" or "there is a thunderstorm". The information uploaded may also describe impacts of weather hazards, including property damage, safety issues or the disruption of utilities. The data collected is visualised on maps, tables and graphs, and people can compare their observations with local Bureau Automatic Weather Stations (AWS). In terms of scale, WOW is able to compile data from all territories in Australia. Since its implementation in 2013, the website had about 10,000 hits per month and currently has around 300 automatic weather stations continually uploading data.



To ensure that data is fit for purpose, the Bureau has drafted guidelines on specifications for instruments citizens should use in order to provide the Bureau with accurate data. For example, to correctly represent the rainfall amount per day the right calibration needs to be set (e.g. the rain gauge resolution should be set to 0.2 millimeters) and be captured in the right measurement intervals (on an hourly basis). The Bureau also provides information on how to maintain the equipment and captures metadata to judge the quality of the data uploaded.

Meteorologists working with the Bureau acknowledge the challenge of drawing data from multiple sources of data. Therefore, the Bureau is planning to develop a tiered approach for engaging with data sources. In this scenario, the Bureau's own climate sites serve as primary data sources for large-scale forecast modelling. CGD aggregated on WOW is used to increase situational awareness. The remaining data sources such as sensors and automated weather stations provide a lot of observations/noisy data which could be used for applications by third-party developers.

What difference did the involvement of citizens make?

As the WOW website states, "the independent information you submit to WOW increases knowledge of weather and climate, as well as raising awareness of unrecorded weather events. This leads to better informed communities, in remote, regional and local areas." According to a representative of the Bureau, "the primary reason to set up WOW was to engage more with the community" as opposed to filling the data gaps. The community mostly consists of individual enthusiasts but the project also targets people affected by extreme weather conditions, such as farmers affected by droughts. To that end, the Bureau

is developing a project "where we are trying to get more and more people particularly in western New South Wales that keeps experiencing significant drought at the moment".

Other interest groups include organisations such as fire stations, as well as people's associations along the east coast of Australia where sometimes no official weather stations are placed. For instance, a yacht club has installed its own network of weather stations and shares the data with the Bureau through WOW to benefit the public. Gathering data from individuals, groups and formalised organisations may encounter different legal restrictions. In order to gather data on droughts, the Bureau collaborated with individual farmers as well as associations of farmers with commercial interests. As these organisations were willing to share data with the Bureau confidentially, WOW needed to develop alternative ways of transmitting data which does not display data immediately.

How does CGD relate to the SDGs?

WOW provides data on extreme weathers at higher resolution, helps gathering data in regions where official weather stations are not installed and provides a platform for people to exchange information. CGD is seen as a valuable complement to other types of weather statistics. The government users of WOW's data are primarily forecasters and modelers. As one official, working with the Bureau said: "Forecasters will take anything and everything that gives them a better understanding of the weather. The key use of the data is for forecasters to use the information. We don't ingest it to the Bureau as yet, it is just viewed in WOW. We do have plans - working with the Meteorological Office - to provide some API to directly ingest the data into our forecasting systems."

The resulting data will ideally inform decision making on drought-related funding to farmers who are going to be affected by drought. A related SDG target could be SDG 2.4[85] referring to the implementation of resilient agricultural systems, adaptation strategies to extreme weather and drought. Monitoring the severity of droughts may help farmers as well as governments identify such strategies, including allocating resources to regions particularly affected, or supporting programs for experimentation with new farming techniques.

The resulting data will ideally inform decision making on drought-related funding to farmers who are going to be affected by drought.

As a side-effect, the launch of WOW was a way to invest in tools and devices to enable engagement in the meteorological community to make use of the enthusiasm to provide data to the Bureau.

85 SDG 2.4 "By 2030, ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, that help maintain ecosystems, that strengthen capacity for adaptation to climate change, extreme weather, drought, flooding and other disasters and that progressively improve land and soil quality".

How citizen-generated data can help achieve the SDGs

Citizen-generated data (CGD) can support the achievement of the Sustainable Development Goals (SDGs) in different ways. Based on our case studies, we outline three ways to consider the potential relationship between CGD initiatives and the SDGs:

1. CGD initiatives create new kinds of relationships, participation and public spaces between public institutions, civil society groups and citizens
2. Beyond closing data gaps, CGD initiatives also provide alternative ways of measuring and help informing, expanding, and improving existing monitoring efforts
3. CGD initiatives can inform public policy, behavioral change and other actions towards implementing and achieving the SDGs

Creating new kinds of relationships, participation and public spaces between government and civil society

Most of our case studies emphasise that CGD initiatives are a way to engage the public, to educate people by means of producing data or to improve the relationship between public institutions and the public.

For some governments, CGD initiatives are opportunities to support local development and education, as the above examples of UBOS and HOT Indonesia show. In both case studies, governments consider CGD as a vehicle to build capacity within institutions and among the public to collect and use data. CGD is also an opportunity for governments to initiate outreach and engagement with CGD communities (see StatsCan's engagement with the OSM community) and special interest communities (see the case of WOW, and how the Australian Bureau of Meteorology (BOM)'s sought to engage farmers). In both cases,

the success of the projects was primarily measured in terms of engagement (the number of participants for example). As Alessandro Alasia from Statistics Canada said: "The idea of this pilot project was not to collect data - of course it was to collect data - but it was to test the idea of crowdsourcing." Governments can use citizen-generated data as an enhancement of existing engagement strategies. This may blur the boundaries between inner workings of the government - which follow their own rules, logic and practices - and the 'outsider' practices of civil society organisations, so as to enable a common learning process. Engaging with CGD initiatives can help laying the foundation for organisational changes, and learning within government. Governments may also develop outreach skills and tools that are not necessarily in the array of public authorities' competences, such as innovating public relations work.

Some CGD cases demonstrate more substantive and ambitious ways of engaging with sustainability issues and how institutions can manage these. CGD initiatives grounded in the ideas of social accountability, environmental justice and rights-based advocacy don't stop at data literacy trainings. For them, CGD is part of strategies to collaboratively improve public institutions and public policy. This can include creating action plans, campaigns and public outreach, as well as defining commitments with government to improve the management of public institutions. Institutional willingness and openness towards citizen-generated data and public engagement are key, and can be organised in different ways.

CGD initiatives therefore dovetail with several SDG targets stressing the inclusion of citizens and civil society groups in public decision-making. SDG 16.7 calls on countries to "ensure responsive, inclusive, participatory and representative decision-making at all levels." SDG 6.b encourages "mechanisms by which individuals and communities can meaningfully contribute to decisions and directions about water and sanitation management"[86]. Ultimately, CGD initiatives may be an opportunity to invest in civil society partnerships (SDG 17.17[87]).

Beyond data gaps: providing alternative and complementary ways of measuring SDGs

Often CGD initiatives are considered as opportunities to close existing data gaps. In practice, the relationship between CGD and the data necessary to monitor the SDGs is more nuanced.

CGD initiatives can produce alternatives to official data in regions otherwise not reachable by government: Governments might not always find themselves able to produce official data collections for all regions in a country. This can have different reasons. In several cases, public data infrastructures are run by high-level government agencies. These gather data up to a certain administrative level, but do not capture granular data to cater to local decision-makers and management. In Indonesia, local governments receive geographic survey data from private vendors (HOT Indonesia), and in Uganda, national household welfare averages do not reflect parishes and villages (UBOS Uganda). In the case of Canada, the sheer size of the country prevents government from gathering complete datasets on environmental pollution. In several cases, CGD can provide granular data which can either complement

86 "Support and strengthen the participation of local communities in improving water and sanitation management." See https://unstats.un.org/sdgs/metadata/files/Metadata-06-0B-01.pdf

87 SDG 17.17: "Encourage and promote effective public, public-private and civil society partnerships, building on the experience and resourcing strategies of partnerships".

data from other regions, or create more granular versions of existing datasets.

Some CGD initiatives can provide updated and detailed data at lower costs and more quickly than established data collections: Some official data collection methods, such as household surveys and geographic surveys are very expensive. The speed of data gathering, and overall sufficient data quality provided by CGD initiatives is an argument for some governments to experiment with the CGD. Despite the benefits of gathering more granular data faster and at lower costs, CGD is not cheap. Interventions usually require outreach, organising with communities, support, training, physical resources and coordination.

CGD initiatives can identify patterns and phenomena normally hidden behind averages: In many cases, CGD deliberately does not aim for exhaustive data collections, or representative samples. Instead, the goal is to get to a sufficiently close representation of a phenomenon of interest. We might think of this as pattern identification. In Pristina, Kosovo, repeated air quality measurements identified pollution hotspots where people are directly exposed. Information gathered by social audits and community scorecards - if aggregated - can foreground patterns of service delivery problems across facilities. Concerns such as self-selection and regional bias (only popular locations are covered) may be partly addressed through coordination and the identification of intervention areas from the start. Sometimes collecting citizen-generated data may require comparative data collection to cross-verify measurements or to ensure data quality. This may lead to the creation of 'data doubles' by CGD initiatives and governments (see also next point).

CGD initiatives can guide official monitoring systems, and improve their efficiency to detect issues: CGD does not only increase the breadth, timeliness or granularity of data. CGD may also help plan, support or guide governmental data collection. Data Zetu and Tanzania's National Bureau of Statistics are currently exploring how local administrative boundaries can be used to plan the upcoming census. Whilst undocumented in our case studies, literature discusses how environmental monitoring[88] and social auditing[89] could spark governmental data collections. By collecting comparative data on a problem (creating 'data doubles'), governments can determine not only the quality of CGD but also of its own data collections and predictive models. For instance, CGD can be used to test the extent to which government's predictive models reflect problems on a local level (as in the case of air pollution monitoring).

Official data and monitoring systems may be essential to support the production of CGD: CGD and official data sometimes exist in a symbiotic relationship. For example publicly accessible official data collections can provide baseline data for CGD projects to start their data collections. Governments may want to enrich their data, or open up data for collaborative auditing projects. They can find demand for (open) data in CGD initiatives. In other cases, government data and CGD can cross-verify one another.

88 CGD has been used by the U.S. EPA as part of the Clean Water Act to respond to possible water pollution anomalies and to initiate government data collections. This duplication of data serves the purpose of "peer review" and verification of CGD for regulatory choices. See: Brett, A. (2018): Putting The Public On Trial. Can Citizen Science Data be Used In Litigation and Regulation? Available at: https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1395&context=elj

89 See a case of social auditing in Nigeria. Gray, J.; Lämmerhirt, D. (2017): Data and the City. How Can Public Data Infrastructures Change Lives in Urban Regions? Available at: https://blog.okfn.org/files/2017/02/DataandtheCity.pdf

CGD initiatives can complement official monitoring systems and open them for public deliberation: Social audits and community scorecards can render output-oriented performance monitoring public. Public participation in auditing may not only close 'monitoring gaps' and provide more accurate reflections of public services. Basing monitoring on how service providers and users perceive problems can open new ways of seeing 'good service quality'. Governments can gather context around a problem otherwise unnoticed in output-oriented measurements. Community scorecards in particular reflect people's lived realities which are translated into measurable indicators, making this approach a much more deliberative process to generate metrics.

Political dynamics might come into play if CGD represents a different type of benchmarking system. For example, governments may prioritise institutionalised output-oriented performance metrics over the insights of participatory methods). In other cases, established measurement systems may be contested.

The role of CGD for implementing and achieving the SDGs

Driving progress around the SDGs requires a whole array of actions, including policy making, regulatory decisions, problem identification and planning in cities and beyond, expanding baseline research and monitoring, changing individual behaviour, but also better and different ways of managing services. Beyond participation and more inclusive decision-making, this may include several public policy areas, including gathering baseline research data, monitoring of performance and service management, planning and resource allocation.

When is CGD fit for purpose?

Recent research starts to provide a more nuanced answer to the question when citizen-generated data is good enough.[90] Legal and public policy-oriented research suggests that higher quality is required the more likely the governmental use case involves legal and policy actions.

Several pieces have looked at the ways the US Environmental Protection Agency (US EPA) assesses the suitability of citizen data for different operations. The required data quality is lower when the initiative has an educational purpose or shall foster community engagement. Gathering indicators for baseline research requires better data quality, whilst legal and policy actions affecting the rights of persons require the highest quality standards.

The US EPA has used for over twenty years a tiered approach to assess and triage citizen science data (see more information in our guide). Soon, they will publish guidelines outlining quality assurance processes for different public policy use cases.

90 Gabrys, J., Pritchard, H., Barratt, B., (2016): Just good enough data: Figuring data citizenships through air pollution sensing and data stories. Available at: http://journals.sagepub.com/doi/abs/10.1177/2053951716679677

Our case studies show that CGD is particularly useful when users of the data have a clear remit or responsibility to manage a problem, and when data feeds into actionable tasks. CGD might be less helpful if actors have conflicting interests, if the responsible agency cannot easily be determined, or if problems require multiple related actions.[91]

Our case studies show that CGD is particularly useful when users of the data have a clear remit or responsibility to manage a problem, and when data feeds into actionable tasks.

Here, we illustrate some use cases where CGD can play a role:

Baseline research: CGD may provide otherwise unavailable data for local decision-makers by gathering new baseline information or by increasing the granularity of existing data. WOW Australia as well as Atlantic DataStream aggregate citizen information in the absence of situational data. In Canada, baseline data is still largely missing for many open waters. Currently both WOW and AWN explore the use of CGD for further research.

Planning and strategy development: Participatory mapping and using data for disaster modelling can support in urban and national disaster planning. OSM data may provide data where alternative data at similarly high resolution is otherwise not available. Household data, gathered as part of the CIS in Uganda, contributed to local development plans and budgetary decision-making, which affect rural development more broadly.

Resource allocation: HOT's building data may serve to inform health programmes to reduce communicable diseases - in this case particularly as malaria infections depend on how people protect themselves at home. Participatory mapping can support programs to reduce the number of malaria cases, therefore driving progress around SDG target 3.3 and SDG indicator 3.3.3 ('Malaria incidences per 1000 population'). In Mozambique, the Ministry of Health explores the use of community scorecards to allocate financial resources based on facility performance.

Improving access to services: Our case studies suggest several ways how CGD can be used to improve access to services, including using CGD as basis for budget planning (household data), for health care planning (CEP), or for facility level actions (Black Sash). CGD initiatives can also multiply the perspectives on access to basic services (SDG 1.4), and reframe what counts as an 'accessible' service. Geographic access to services could be mapped via participatory mapping[92], whilst household surveys can provide representative data how population samples use services. Social audits, and community scorecards can reframe what counts as a well-performing service, based on people's perceptions. As CEP's case shows and additional literature shows[93], translating people's perceptions into action plans and commitments may directly address these problems, and enhance the use of services. It may also help increase the proportion of population satisfied with public services (SDG 16.6.2).

91 See Wesselink, A.; Hoppe, R.; Lemmens, R. (2015): Not Just a Tool. Taking Context into Account in the Development of a Mobile App for Rural Water Supply in Tanzania. Available at: https://core.ac.uk/download/pdf/31152392.pdf

92 UN Habitat acknowledges remote sensing as one form of understanding access to basic services.

93 Gullo S., et al. (2017): Effects of a social accountability approach, CARE's Community Score Card, on reproductive health-related outcomes in Malawi: A cluster-randomized controlled evaluation. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171316

Next steps to better support CGD

The following section illustrates five areas outlining how CGD could be better supported in the future. These are derived from the case studies included in this report, as well as a broader literature review and additional interviews. In order to 'scale' CGD, a shift towards an enabling socio-technical infrastructure is needed which can include:

1. Legal and policy frameworks
2. Inclusion of CGD into administrative processes
3. Infrastructure
4. Resourcing
5. Coordination

While the first two areas primarily address governments to lay the foundation for CGD, the remaining areas outline some elements of CGD which define how CGD is organised, and can be replicated and scaled to gather data relevant for different levels of government.

We propose these as points conversation starters because emerging experimentations in the field are only emerging, because research has just started to address some of these topics, and because a further elaboration of these aspects has been outside of the scope of this report.

Legal and policy frameworks

Some governments have established different legal and policy frameworks to determine how governments can engage with unofficial data or citizen data. On an international level, the Fundamental

Principles of Official Statistics are a key policy framework to guide the rationale for using unofficial statistics. Principle 5 includes a provision arguing that data can be gathered from any available source, given that data quality, costs, timeliness and burden on respondents are appropriate.

Governments have started to provide for the use of citizen-generated data directly. In the US, the Clean Water Act regulates the Environmental Protection Agency's (EPA) responsibilities to monitor water sources. As part of this framework, the EPA has developed regulations requiring each state to develop assessment plans for all existing data sources including from 'citizen monitoring groups', and 'members of the public' and to actively solicit for research they may be conducting or reporting. In addition to developing assessment plans, the EPA also provides resources such as quality assurance project plans and quality assurance/quality control protocols the public can use to ensure data quality.

Other governments base their engagement with citizen-generated data on laws and policies defining the citizen's role in relation to the public sector more broadly. The South African Department of Monitoring and Evaluation (DPME) based the development of a government-wide 'citizen-based monitoring' framework[94] in the South African Constitution, and the National Development Plan from 2012 underscoring the importance of public participation in policy development, service delivery design and accountability at the point of delivery. The DPME identified a range of institutions that could adopt more participatory monitoring approaches.

CGD does not only mean using citizen data, but also requires enabling CGD initiatives through open government and open data initiatives. Citizen-generated data often depends on the availability of government data, as in the case of social auditing efforts. Laws, policies, and decrees fostering the publication of open government data, as well as freedom of information/right to information are critical enablers for people to underpin their CGD initiatives.

Recommendations

Research may replicate existing studies of how governments currently regulate engagement of particular agencies with citizen-generated data. This may include comparative legal analyses of the provisions in legal and policy frameworks, the institutional processes they mandate, and the roles citizens play. Emerging ideas, such as the right of citizens to produce data, should be considered, to explore new forms of citizenship and civic engagement through data.

94 See https://www.dpme.gov.za/keyfocusareas/cbmSite/CBM%20Documents/Framework%20for%20Strengthening%20Citizen-Government%20Partnerships%20for%20Monitoring%20Frontline%20Service%20Delivery.pdf

Inclusion of CGD methods into existing administrative processes

Governments have many options to expand their processes and participation models with CGD. South Africa's DPME developed a framework for public institutions on how to include citizen-based monitoring[95] into existing planning, budgeting and evaluation processes. The approach emphasises action learning, root cause analysis and collaborative decision-making and complements rather output-oriented performance tracking. DPME helps governments adopting the approach, providing support through advice, tools and in-field assistance to government entities and helps shifting auditing cultures of administrations towards more deliberative models.

Other governments have found different opportunities to include CGD in their processes. In Mozambique, the national Ministry of Health and Wellbeing is currently exploring ways of integrating community scorecards into performance-based funding for health facilities. UBOS in Uganda has built its CIS on top of existing administrative structures to develop a cascading capacity building model within government to make decisions based on statistical information.

Recommendations

Governments and civil society may explore how existing public institutions allow for citizen participation, and how they help voicing people's concerns around topics. Governments should explore in what ways administrations already provide for public participation, and what requirements for organisational change would be needed to meaningfully engage with citizens and their data.

Infrastructure

Different CGD initiatives require different types of technical and organisational infrastructure to collect data and to scale. Infrastructure, coordination and resourcing overlap in many ways, and can enable one another differently. Different fields of research, from infrastructure studies, to platform studies or media studies have started to ask whether new data production methods cause a fragmentation or "splintering" of data infrastructures and the effects this has.[96]

There are several ways one might start thinking about citizen-generated data infrastructure and how it enables to coordinate and resource CGD initiatives. This can include providing infrastructure to connect CGD initiatives in order to extract insights from spatially or otherwise aggregated data, and to increase their interoperability.

95 See https://www.dpme.gov.za/keyfocusareas/cbmSite/Pages/CBMFramework.aspx

96 For broader context on the discussion how infrastructure refers to platforms see: Plantin et al. (2016): Infrastructure studies meet platform studies in the age of Google and Facebook. Available at: https://doi.org/10.1177/1461444816661553

Photo from Data Zetu K15 Photos / BY SA NC

At a tool level, several solutions for interoperability exist. The workflows of HOT and OSM can make use of open source, customisable Open Data Kit, and OpenMapKit, as well as the Kobo Toolbox. The resulting data can be pushed into Mapbox and other applications, and data collection is possible offline as well as online. Black Sash has also developed a survey tool which is open source and can be adjusted to customise surveys for mobile phones. Kwantu has developed a paid-for application to standardise the way community scorecard results can be captured.

At a dataset level, organisations have started defining reusable standard data models to ensure comparability across initiatives. For instance, the Everyone Counts consortium has developed a data standard for social auditing and community scorecards. The standard provides a common way of recording the issues raised by the group, the score given to the issue, the groups scoring the issue (using Leave no one behind categories), and the joint actions agreed to be implemented as follow up. This enables to filter problems emerging from individual scorecard processes by geographic area and sector as well as leave-no-one-behind categories such as gender, age range, disability status and community role.[97]

Other initiatives are dedicated to develop a common metadata model that can be used by initiatives to increase their findability. Fields like citizen science, but also CGD initiatives from fields like social accountability are scattered, and CGD initiatives in general decentralised. Up-to-date information about citizen science projects is not only important for coordinating citizen-generated data, but is also valuable to find available data for the SDGs. The Citizen Science Association, its Data and Metadata Working Group, and the Wilson Center have developed the PPRS Core Metadata standard to help cataloging

97 See http://www.kwantu.net/community-scorecard-app/

projects. The goal is to develop a common taxonomy for all types of citizen science activities, and areas of research with the goal to increase their findability across repositories, and for search engines.[98]

Several models exist to work with CGD, beyond aggregating data. Databases have been developed across a variety of areas, from cartography (OSM) to environmental monitoring (DataStream for water in Canada, or Cornell's Ornithology Lab) to land ownership (Land Matrix initiative and LandMark).[99] Websites such as CitSci, SciStarter and World Vision's Citizen Voice and Action Database not only collect data, but also function as project registers. World Vision is currently testing its database to gather all data from its programmes, and to develop a standardised data model allowing to filter projects by regions of intervention and community engaged, as well as to provide more analytical capacities, such as analysis of how well public facilities comply with government plans, based on social auditing and scorecard information.[100]

Some websites organise community in different ways by enabling matchmaking, coordination, or task distribution. Matchmaking websites can serve as platforms for individuals to find one another and to form groups. Other websites offer integrations with apps and other systems, becoming platforms for other infrastructures. Crowdsourced databases such as OSM could be seen as a resource for governments and others if used to publish and store data, and use it as 'platform as a service'.

Recommendations

Additional research can help understand the different ways how CGD relates to government data, and under what circumstances they can benefit from one another.

Further work should expand on existing research around standards and their relationship to CGD communities and their stakeholders. This can include to test the perceived benefits and drawbacks of using standards, or investigating participatory models and institutions to design data standards together with communities.[101]

Beyond storing data, web technologies include components to transfer, organise and distribute data. Follow up work can further explore the role of web technologies in enabling and coordinating data collections, including ways of networking communities 'on the ground' as well as their data.

98 Bowser, A., et al. (2017): Citizen Science Association Data & Metadata Working Group: Report from CSA 2017 and Future Outlook. Available at: https://www.wilsoncenter.org/article/citizen-science-association-data-metadata-working-group-report-csa-2017-and-future-outlook

99 See http://www.landmarkmap.org/data

100 See http://www.wvcva.org/ The database is currently in a testing phase.

101 Martin, V.Y., Ramirez-Andreotta, M. D., Goebel, C. (2017): Citizen science data and metadata standardization across the globe: What are the issues for stakeholders? https://www.researchgate.net/publication/316275350_Citizen_science_data_and_metadata_standardization_across_the_globe_What_are_the_issues_for_stakeholders

Resourcing

Initiatives require beyond technical infrastructure different resources to collect data, including technical assistance, funding, or training. Various organisations provide resources, international donors, to national NGOs, FabLabs and maker labs, universities, and governments.[102]

Funding sources can come in various forms, such as project-based grants (Black Sash), short-term contracts between government and central coordinating organisations (HOT), as well as government budgets (UBOS). Money can be further distributed to established communities who receive grants to get support for their local mapping activities. If data collection requires on-site visits, stipends for travel and data collection or other compensations may be needed.

Open government data can in itself be seen as a crucial resource for citizen-generated data. The case study on StatsCan's but also the examples of social audits demonstrate how governments can make data openly accessible, and thereby support the production of updated and more detailed data.

Some organisations have dedicated their operations to provide civic groups and individuals with technology. Equipment Banks can be considered as one model to loan tools to communities. This can ensure that tools are calibrated before being used, and can create interfaces between technical experts and the communities who want to use such tools. Another model may be the provision with open source tools. The Public Lab has developed several open source technologies and guidance material for people to run their own project.

Some data collections may be paid for, coordinated and distributed by a central organisation. For instance, a university in Flanders, Belgium, has recently started a large-scale air pollution project, distributing sensors to more than 20.000 people in Flemish cities to understand air pollution distribution.[103] Likewise, Australia's BOM has achieved to get farmers involved in acquiring weather monitoring stations that can send data to BOM's WOW database.

Training resources include information about methods, protocols or tools to use. Some governments provide entire web pages dedicated to training resources. This can also include training about people's rights to interact with government and to contribute data to public policy processes. A good example is again the US EPA, which has developed an entire catalogue of training material for different types of environmental monitoring endeavours.

102 Examples include the Atlantik Water Network (AWN), the Canadian Aquatic Biomonitoring Network (CABIN), the US-based River Network and the Public Lab.

103 See https://curieuzeneuzen.be/

Recommendations

Beyond volunteerism, explore existing economic models of different CGD initiatives and how those have benefited the initiatives. Inquire what resources governments could provide in relation to available resources from other actors. This can include to study the organisational requirements to embed supporting structures for CGD initiatives into existing administrative processes.

Coordination mechanisms

CGD initiatives, communities and community engagement rarely emerge spontaneously but often need to be nurtured and guided in one way or another. This can have several reasons, including aligning what data gets collected and who collects or verifies the data across distributed organisations. Sometimes data collection needs to take place in specific geographic areas and requires targeted data collection.

In our case studies, several organisations coordinated the recruitment of citizens to alleviate issues around self-selection and regional biases or gaps in the data. This had several reasons, including finding trustworthy individuals who could gather data. In other cases, the project aims to include a specific group of people, either during goal setting, data collection, or other stages of an initiative. For example, HOT runs workshops and other engagement strategies to broaden and balance involvement of certain populations in underrepresented locations (e.g. targeting women, young people and different geographical regions).

Some CGD initiatives coordinate the data production process through a mix of institutional processes. For example, UBOS organises data collection centrally, through a government-internal data collection training to implement local statistical monitoring. This includes to implement technical working groups down to the district levels to help guide the production and use of statistics.

Governments start to collaborate nationally and transnationally. For example in the US several governments have formed learning groups to exchange best practices on how to set up and engage with different forms of CGD. On an international level, the Australian Bureau of Meteorology has replicated technology from the United Kingdom's Meteorological Office to install WOW.

Recommendations:

Governments and CGD projects may explore existing coordination models, and when they are most useful, be it during data definition, during the collection of data, or during dissemination. This may include to study different types of large-scale projects and how they achieve to coordinate data collections.

References

Bowser, A., et al. (2017): Citizen Science Association Data & Metadata Working Group: Report from CSA 2017 and Future Outlook. Available at: https://www.wilsoncenter.org/article/citizen-science-association-data-metadata-working-group-report-csa-2017-and-future-outlook

Brett, A. (2017): Putting the Public on Trial: Can Citizen Science Data Be Used in Litigation and Regulation, Available at: https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1395&context=elj

Carlson, T., Cohen, A. (2018): Linking community-based water monitoring to water policy: Perceptions of citizen scientists. Available at: https://www.sciencedirect.com/science/article/pii/S0301479718304596

Center for Open Data Enterprise (2018): Strategies for SDG National Reporting: A Review of Current Approaches and Key Considerations for Government Reporting on the UN Sustainable Development Goals. Available at: http://reports.opendataenterprise.org/CODE_StrategiesforSDGreporting.pdf

Degrossi, L. (2018): A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/tgis.12329

Eitzel, M.V., et al. (2016): Citizen Science Terminology Matters: Exploring Key Terms. Available at:

https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.96/

Fish, A. (2016): Birds of the Internet: Towards a field guide to the organization and governance of participation. Available at: https://recursivepublic.net/wp-content/uploads/2010/11/PartPublicPaper-JCE-V1.6.1.pdf

Foody, G. et al. (2015): Current status and future trends in crowdsourcing geographic information. Available at: http://eprints.nottingham.ac.uk/29082/1/agi%20foresight-eprints.pdf

Fox, J., (2016): Scaling accountability through vertically integrated civil society policy monitoring and advocacy. Available at: https://jonathanfoxucsc.files.wordpress.com/2011/11/fox_scalingaccountability_online5.pdf

Gabrys, J., Pritchard, H., Barratt, B., (2016): Just good enough data: Figuring data citizenships through air pollution sensing and data stories. Available at: http://journals.sagepub.com/doi/abs/10.1177/2053951716679677

Gray, J., Lämmerhirt, D., (2017): Data and the City. How Can Public Data Infrastructures Change Lives in Urban Regions? Available at: https://blog.okfn.org/files/2017/02/DataandtheCity.pdf

Gray, J. (2018). "Citizen generated data" according to Google Search, 30th July 2018. figshare. https://doi.org/10.6084/m9.figshare.7376243.v2

Gray, J., Gerlitz, C., Bounegru, L., (2018): Data Infrastructure Literacy. Available at: https://journals.sagepub.com/doi/10.1177/2053951718786316

Gullo S., et al. (2017): Effects of a social accountability approach, CARE's Community Score Card, on reproductive health-related outcomes in Malawi: A cluster-randomized controlled evaluation. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171316

Haklay, M., (2013): "Citizen Science and Volunteered Geographic Information – overview and typology of participation". In Sui, D.Z., Elwood, S. and M.F. Goodchild (eds.), 2013. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice . Berlin: Springer. pp 105-122. Available at: https://povesham.files.wordpress.com/2013/09/haklaycrowdsourcinggeographicknowledge.pdf

Kosmala, M., et al. (2016): Assessing data quality in citizen science. Available at:

https://doi.org/10.1002/fee.1436

Martin, V.Y., Ramirez-Andreotta, M. D., Goebel, C. (2017): Citizen science data and metadata standardization across the globe: What are the issues for stakeholders? Available at:

https://www.researchgate.net/publication/316275350_Citizen_science_data_and_metadata_standardization_across_the_globe_What_are_the_issues_for_stakeholders

McDonald, S. M. (2016): Ebola: A Big Data Disaster. Privacy, Property, And The Law Of Disaster Experimentation. Available at: https://cis-india.org/papers/ebola-a-big-data-disaster

McDonald, S. M., Wylie, B. (2018): What Is a Data Trust? Available at: https://www.cigionline.org/articles/what-data-trust

McElfish, J., et al. (2016): Clearing the Path. Citizen Science and Public Decision Making in the United States. Available at: https://www.wilsoncenter.org/sites/default/files/clearing_the_path_eli_report.pdf

Meijer, A. (2015): Citizen generated open data. Available at: https://dspace.library.uu.nl/handle/1874/356763

Peixoto, T., Fox, J. (2016). When Does ICT-Enabled Citizen Voice Lead to Government Responsiveness? Available at: https://openknowledge.worldbank.org/handle/10986/23650

Plantin, J., et al. (2016): Infrastructure studies meet platform studies in the age of Google and Facebook. Available at: https://doi.org/10.1177/1461444816661553

Privacy International (2018): The Keys to Data Protection. A Guide for Policy Engagement on Data Protection. Available at: https://privacyinternational.org/sites/default/files/2018-09/Data%20Protection%20COMPLETE.pdf

Rogers, R. (2013). Digital Methods. Available at: https://mitpress.mit.edu/books/digital-methods

Scassa, T., Chung, H. (2015): Managing Intellectual Property Rights In Citizen Science. A Guide for Researchers and Citizen Scientists. Available at: https://www.wilsoncenter.org/sites/default/files/managing_intellectual_property_rights_citizen_science_scassa_chung.pdf

Shirk, J., et al. (2012): Public Participation in Scientific Research: a Framework for Deliberate Design. Available at: https://www.ecologyandsociety.org/vol17/iss2/art29/

Taylor, L., Floridi, L., Van der Sloot, B. (eds.) (2016): Group Privacy. New Challenges of Data Technologies. Available at: https://www.springer.com/gb/book/9783319466064

The DataShift (2015): What is Citizen-Generated Data And What Is The DataShift Doing To Promote It?, Available at: http://civicus.org/images/ER%20cgd_brief.pdf

Wang, R. Y., Strong. D. M. (1996): Beyond Accuracy: What Data Quality Means to Data Consumers. Available at: http://mitiq.mit.edu/Documents/Publications/TDQMpub/14_Beyond_Accuracy.pdf

Wesselink, A.; Hoppe, R.; Lemmens, R. (2015): Not Just a Tool. Taking Context into Account in the Development of a Mobile App for Rural Water Supply in Tanzania. Available at: https://core.ac.uk/download/pdf/31152392.pdf

Zeelenberg, K.; Booleman, M.; (2013): Quality reporting and sufficient quality. Statistics Netherlands. Available at: https://www.cbs.nl/nl-nl/achtergrond/2013/21/quality-reporting-and-sufficient-quality

Appendices

Methodology

This report started with a literature review to scope existing ways of describing and classifying CGD projects. Methods:

- **Search term-based research** of academic and grey literature
- **Document analysis**
- **Snowballing** to identify relevant concepts and studies in initial text body.
- **Iteration of search** with more targeted search terms.

If new search terms/concepts or authors were identified during document analysis, they were documented in our search term list, and used for another search.

We queried Google Scholar using the search term "Citizen-generated data" (using quotation marks searches narrowly for this term, not synonyms). By using this term we understand often the term is used in academic literature. It also helps to identify how "citizen-generated data" is associated to other concepts. Our literature corpus resulted in 362 mostly peer-reviewed academic publications. We filtered results by skimming abstracts. Irrelevant[104] content was filtered out, resulting in 122 articles at least mentioning citizen-generated data once.

104 Our exclusion criteria included: No social media and big data articles if they discuss analytics of social media data without the involvement of individuals, like here, here, here, here (Digital Earth - big data project), or here. We did not consider studies of user-generated content and socio-demographics, see here. We excluded papers on user-generated content for newspapers, or articles dealing with broader ideas of development, governance of development aid, et al. such as here.

As our report aims to be illustrative rather than comprehensive, in order to begin exploring the breadth of this field we started with a list of over 230 projects that were associated with the term "citizen-generated data" on Google Search, using an approach known as "search as research" (Rogers, 2013). This means that we focus mainly on English language examples that have been explicitly associated with the term, as well as initiatives suggested by the GPSDD's citizen-generated data task team.

Methods:

- **Iterative, structured keyword search** of projects.
- **Targeted search** in a selection of repositories and selected literature to complement project list.
- **Analysis of cases** based on preliminary distinction criteria
- **Extraction** of further distinction criteria
- **Repeat: Iterative, structured keyword search** of projects.

We note following challenges when scoping citizen generated data projects: There are many alternative names and concepts for CGD, including citizen science, community monitoring, and others. Our query design only covers projects from English speaking groups. We first gathered a set of CGD initiatives, extrapolated from these methods to 'generate data', and later summarised these data operations in a list of methods. To understand scope and replication of methods, we did not consider terminology such as "community-based monitoring" but instead only looked at the actual labour to generate data.

In order to understand how governments can engage with citizen-generated data, we conducted desk research of government documents, and interviews with several government officials and civil servants in administrations and statistics offices (see acknowledgement section).

Acknowledgments

The authors would like to register our gratitude to the following people whom we interviewed, consulted or otherwise drew inspiration from in this project.

Yantisa Akhadi, Humanitarian OpenStreetMap Team

Alessandro Alasia, Statistics Canada

Rupert Allan, Humanitarian OpenStreetMap Team

Mara Balestrini, Ideas for Change

Jay Benforado, United States Environmental Protection Agency

Anne Bowser, Woodrow Wilson International Center for Scholars

Herbert Buyondo, Uganda Bureau of Statistics

Janet Chapman, Crowd2Map Tanzania

Lonny Chen, Morogoro Youth Initiative

Gillian Cooper, Commonwealth Foundation

Jack Cornforth, CIVICUS Alliance

Sam de Craemer, University of Antwerp

Sarah Cummings, The Demand Institute

Rebecca Firth, Humanitarian OpenStreetMap Team

David Gooding, Australian Bureau of Meteorology

Erica Hagen, Map Kibera

Mordechai (Muki) Haklay, University College London

Poncelet Ileleji, Gambia YMCA

Kshitiz Kamal, Kathmandu Living Labs

Edgar Kisembo, World Vision International

Alexa Lane, Black Sash

Claudia Lopez, UN Women

David Luswata, Humanitarian OpenStreetMap Team

Maria Rosa Mondardini, Citizen Science Center Zurich

Dave Mitchell, Connected Bits

Besinati Mpepo, World Vision International

Bernard Justus Muhwezi, Uganda Bureau of Statistics

Zachary Mwangi, Kenya National Bureau of Statistics

Evashnee Naidu, Black Sash

Denise Namburete, Nweti

Laura Oremland, United States National Oceanic and Atmospheric Administration (NOAA)

Noel Perez, Philippines Statistics Authority

Alexandre Pólvora, Joint Research Centre, European Commission

Lejla Sadiku, United Nations Development Program

Ron Salaj, Science for Change Movement

Omar Seidu, Ghana Statistical Service

Biondi Sima, Humanitarian OpenStreetMap Team

Jessie Smith, Atlantic Water Network

Jonathan Timm, Department of Monitoring and Evaluation South Africa

Johannes Tonn, Global Integrity

Samhir Vasdev, International Research & Exchanges Board

Jennifer Walker, Freelance consultant

Maike Weißpflug, Naturkundemuseum Berlin

Hannah Wheatley, CIVICUS Alliance

Marcos Wolff, TECHO

Robert Worthington, Kwantu

List of Abbreviations

API	Application Programming Interface
AWN	Atlantic Water Network
AWS	Automatic Weather Stations
BNPB	National Disaster Management Agency
CBM	Community-based monitoring
CIS	Community Information System
CEP	Citizen Engagement Program
CESC	Centro de Aprendizagem e Capacitação da Sociedade Civil
CGD	Citizen-generated data
CHAI	Clinton Health Access Initiative
CSO	Civil society organisation
DCLI	Data Collaborative for Local Impact
DHMT	District Health Management Team
EMS	Emergency Medical Services
EPA	Environmental Protection Agency
EU	European Union
GPSDD	Global Partnership for Sustainable Development Data
HOT	Humanitarian OpenStreetMap Team
IREX	International Research & Exchanges Board
IRS	Indoor Residual Spraying
KEPA	Kosovo's Environmental Protection Agency
MCC	Millennium Challenge Corporation
NDoH	National Department of Health
NGO	Non-governmental organisation
NMP	National Malaria Programme
ODbL	Open Database Licence
OSM	OpenStreetMap
PEPFAR	President's Emergency Plan For AIDS Relief
SDGs	Sustainable Development Goals
StatsCan	Statistics Canada
UBOS	Uganda Bureau of Statistics
UNICEF	United Nations International Children's Emergency Fund
WHO	World Health Organisation
WOW	Weather Observation Website
WWF	World Wildlife Fund

Danny Lämmerhirt (Open Knowledge International) was Principal Investigator, leading the project, its conceptualisation, case study development, framing, and overall analysis. Jonathan Gray (King's College London), Tommaso Venturini (CNRS) and Axel Meunier (Sciences Po, Paris) at the Public Data Lab led the mapping of citizen-generated data (CGD) projects, undertaking interviews and empirical work, contributing to research design, framing, and the first draft of several sections in the report and guide (with an emphasis on different ways of doing CGD, inspired by recent research in science and technology studies, new media studies and critical data studies). Members of the Global Partnership for Sustainable Development Data (GPSDD) CGD Task Team provided input and feedback throughout the process.

Global Partnership for Sustainable Development Data

OPEN KNOWLEDGE INTERNATIONAL

Public Data Lab