MAKING CITIZEN GENERATED DATA WORK

TOWARDS A FRAMEWORK STRENGTHENING COLLABORATIONS BETWEEN CITIZENS, CIVIL SOCIETY ORGANISATIONS, AND OTHERS

Danny Lämmerhirt
Shazade Jameson
Eko Prasetyo

DATASHIFT   OPEN KNOWLEDGE

Danny Lämmerhirt is research coordinator and executive researcher at Open Knowledge International. **Shazade Jameson** is an independent social science researcher on governance issues. **Eko Prasetyo** is a practitioner focusing on ICT4D project implementation and evaluation.

MAKING CITIZEN GENERATED DATA WORK

TOWARDS A FRAMEWORK STRENGTHENING COLLABORATIONS BETWEEN CITIZENS, CIVIL SOCIETY ORGANISATIONS, AND OTHERS

Danny Lämmerhirt

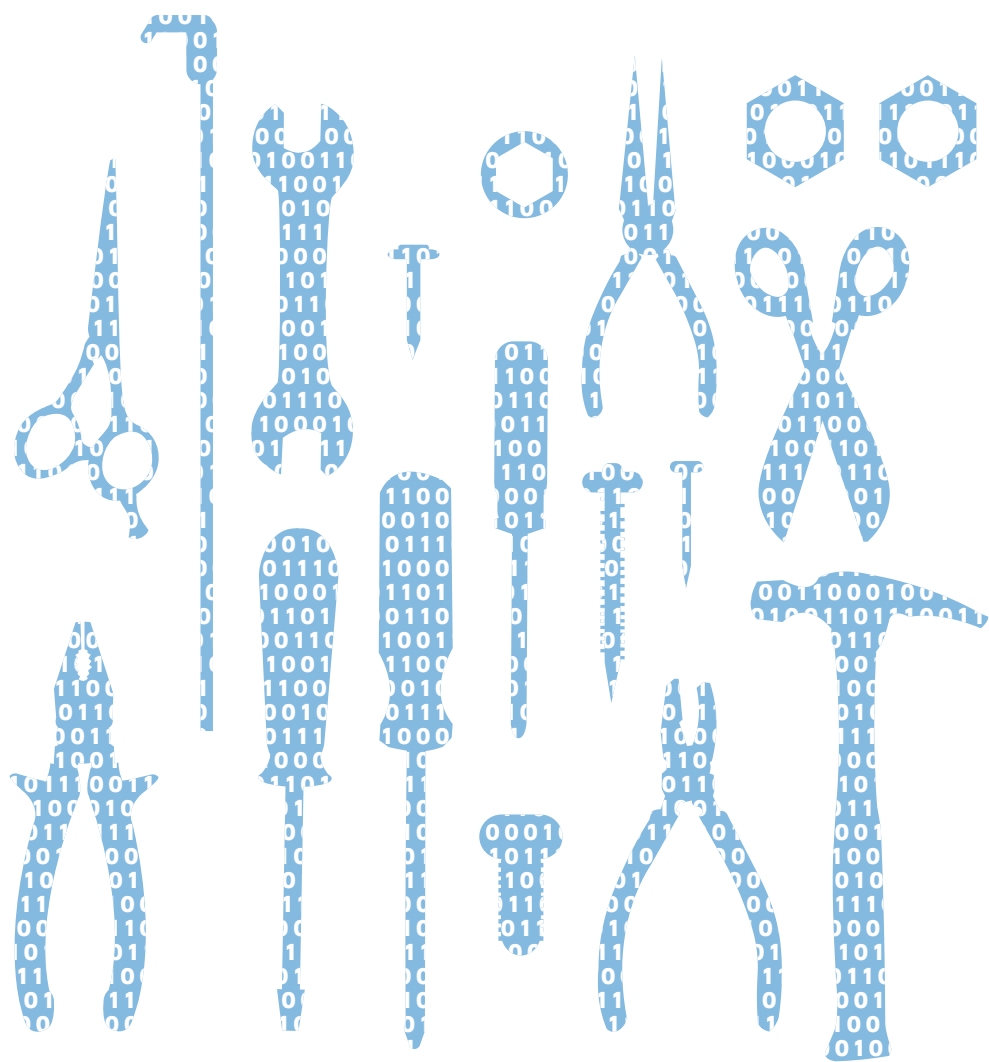Shazade Jameson

Eko Prasetyo

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

RECOMMENDATIONS

On the basis of our case studies and interviews for this report, we suggest that everyone interested in working with citizen-generated data should mind the following points:

- **ALIGN INTERESTS AMONG KEY STAKEHOLDERS TO ENCOURAGE PARTNERSHIPS.**
  Successful CGD projects bring together actors with different interests in the same data. Data serves as common ground for actors and is the focal point of collaborations. There is often a difference between the benefits associated with production, use, and uptake. Different actors can value different aspects of the data; understanding how actors perceive this value is key to build multi-stakeholder partnerships. Furthermore each CGD project should communicate goals and benefits of a collaboration clearly.

- **CITIZEN-GENERATED DATA SHOULD BE USABLE IN MULTIPLE WAYS TO MAXIMISE UPTAKE AND IMPACT.**
  The more ways a dataset can be used, the more different types of actors will become interested in the data. To facilitate different use cases by different actors data needs to be accessible and presented in an interoperable format.

- **TAPPING INTO EXISTING RESOURCES AND PROCESSES MAKES IT EASIER**
  to produce and use CGD effectively. This includes using technology citizens already use, as well as building on established routines and group dynamics, such as existing bureaucratic processes or community forums.

- **CONSIDER THE SPECIFIC INCENTIVES THAT DEPEND ON THE CONTEXT AND THE GOAL.**
  Key dimensions to consider include whether the project aims to link up with government directly or not, and the socio-political and governance environment. This includes, amongst others, whether the government is responsive, whether there is a strong legal framework and high levels of trust, or whether there is adequate information about the issue.

FOREWORD
MONITORING SUSTAINABLE DEVELOPMENT FROM THE BOTTOM UP

The Sustainable Development Goals (SDGs) are the largest effort in history to measure, understand, and drive sustainable development around the world. With 17 goals, 169 targets, and 230 indicators the SDGs aim to reflect the state of the Earth's social, economic and environmental dimensions more holistically than their predecessors, the Millennium Development Goals (MDGs). An additional self-declared objective for the framework is "to leave no one behind" by capturing the living conditions of the most vulnerable and marginalised people. The SDGs aim to monitor progress over 15 years, meaning SDG data needs to exist for a long time, be methodologically sound and comparable across nations and, regions and over a long time span.

Likewise, monitoring progress on the SDGs cannot only be borne on the shoulders of national governments and National Statistics Offices (NSOs). The costs of equipping NSOs to capture information on all indicators are estimated to require excessive funding from international donors, while funding for statistical offices remains low especially in countries with weak statistical systems.[1] In many cases no data yet exists, nor does a standardised methodology to capture the data.[2] Most importantly, by deciding what to measure and how, governments and NSOs, willingly or not, paint a certain image of humanity's development towards sustainability. This image may risk overseeing the issues most important to citizens and civil society.

The MDGs demonstrated that official national aggregates might mask inequalities between populations. The United Nations Human Rights High Commissioner's Office states that "SDG data collection should capture not only national averages or aggregate statistics, but also the situation of the most disadvantaged or deprived,

1   See also: Badiee, S. et al.(2016): The State of Development Data Funding 2016. Available at: http://opendatawatch.com/wp-content/uploads/2016/09/development-data-funding-2016.pdf as well as Partnership for Statistics in Development in the 21st century (2016): Support to statistics remains low, putting SDG monitoring at risk. Available at: http://www.paris21.org/sites/default/files/PRESS-2016-web-final.pdf

2   The SDG indicators are split into three tiers, to reflect their methodological maturity. Tier I indicators are based on historical data, and NSOs have established, standardised methods to collect them. For Tier II indicators either no historical data records exist, or no methodology is agreed upon to collect data on them. Finally, Tier III indicators are those for which neither data nor agreed methodology exist yet. See also: Badiee, S. et al. (2016): The State of Development Data Funding 2016.
Available at: http://opendatawatch.com/wp-content/uploads/2016/09/development-data-funding-2016.pdf

as well as the inequalities among social groups"[3]. Beyond the concerns of gaining detailed enough data there is a second concern: official statistics cannot or may not want to capture certain issues or disparities, such as the incidence of corruption or human rights abuses. Data created by citizens can be an important instrument to highlight new or under-reported issues, express counter-narratives of sustainable development, and truly make "all voices count".

This research series investigates how citizen-generated data can be used to monitor progress around the SDGs. This first piece *Making Citizen-Generated Data Work* elaborates different ways projects mobilise partners and resources to produce citizen-generated data.

Which incentives do different actors have to join a collaboration and produce or use citizen-generated data? Who has to be engaged, and how, in order to ensure that a collaboration achieves its goals? How do projects organise themselves to do this? Which resources and expertise do they bring into a collaboration? And what can other projects learn to support and benefit their own citizen-generated data initiatives? The piece stresses the importance of partnerships, being able to harness citizen-generated data to tackle issues within different contexts. In this way, the piece responds to the possible issue that projects overemphasise the design of data production, whilst not paying enough attention to how others can make use of the data, and how the data feeds into different organisational processes to alleviate issues.[4]

The research series was commissioned by DataShift, an initiative that builds the capacity and confidence of civil society organisations to produce and use data, especially citizen-generated data, to drive sustainable development. It also builds on former research by Open Knowledge International on what can be done to make the "data revolution" more responsive to the interests and concerns of civil society.[5]

---

3   See also: United Nations (2015): Ensuring no one is left behind: A human rights-based approach to data.
    http://www.ohchr.org/Documents/Issues/MDGs/Post2015/DataDisaggregation.pdf
4   Datashift (2016). Using Citizen Generated Data to monitor the SDGs: A Tool for the GPSDD Data Revolution
    Roadmaps Toolkit. Available at:
    http://www.data4sdgs.org/s/Data4SDGs_Toolbox-Citizen_Generated_Data_for_SDGs.pdf
5   Gray, J. (2015). Democratising the Data Revolution: A Discussion Paper. Open Knowledge.
    Available at: http://blog.okfn.org/2015/07/09/democratising-the-data-revolution/, Gray, J., Lämmerhirt, D.
    (2015): Changing What Counts: How Can Citizen-Generated and Civil Society Data Be Used as an Advocacy Tool
    to Change Official Data Collection. Available at: http://civicus.org/thedatashift/wp-content/uploads/2016/03/
    changing-what-counts-2.pdf; as well as Gray, J. and Lämmerhirt, D. (forthcoming): Data And The City. How Can
    Public Data Infrastructures Change Lives in Urban Regions?

1 INTRODUCTION

We live in a datafied world. Governments use data to measure the impact and efficiency of evidence-based policies. Businesses mine vast amounts of data to research markets and understand their clients. Individuals are organising their social life through media enabling to capture, collect and communicate more information on their beliefs and actions than ever before. The United Nations called for a "data revolution" to leverage new technologies and tap into this vast trove of existing and emerging data to help the Sustainable Development Goals (SDGs), which seek to measure humankind's path towards sustainability with a more expanded array of data sources while aiming "to leave no one behind".

Data is not only a mere camera to look unto the world. By determining what is measured and how, data writes a certain story–and leaves out many others that could be written. Civil society increasingly recognises the value that data holds to tackle the issues affecting it–be it by collecting evidence on oil spills in the Gulf of Mexico, running surveys to understand the satisfaction of local communities with health facilities, or challenging existing statistics about politicised topics.[6] The data to do so can already be held by larger NGOs and civil society organisations–or it can be **citizen-generated data**, which is actively created by citizens, often to directly monitor, demand or drive change on issues that affect them.[7]

While projects creating "citizen-generated data" (in the following: CGD) are important to reflect citizens' perspectives, they are often short-lived or cease to exist if they don't achieve their desired outcomes. Some projects may depend on external funding from donor bodies and development agencies. Once funding runs out, the project stops. Other projects struggle to keep citizens engaged with collecting data over a long time–a problem gaining importance if this data shall describe developments around an issue.[8] Yet in other cases, data may be fed onto a website and reside there silently without being used by the public or government.

6   See also a report by Civicus (2015): What Is Citizen-Generated Data And What Is DataShift Doing To Promote It? Available at: http://civicus.org/images/ER%20cgd_brief.pdf

7   See also: Higgins. K., Cornforth, J. (2015): Civil society and SDG monitoring: Harnessing Civil Society and Citizen-Generated Data.
    http://civicus.org/images/SDG%20Monitoring%20-%20DataShift%20Background%20Note%20-%20Final.pdf

8   See also: Irby, E. (2015): Kenya Learning Paper 2015–Closing the Loop: Combining Community Integrity Building with Integrity Education. Available at: http://integrityaction.org/publication/kenya-learning-paper---2015-closing-loop-and-combining-community-integrity-building

Given the important role citizen-generated data can play, how can the success of these projects be encouraged? Who has to be engaged, and how, in order to ensure that a project achieves its goals? How do projects mobilise resources to produce data as well as engage with interest groups that can make use of the data? Former work often regarded one element in the pathway of data production and uptake. How projects organise themselves, why, and what the chain of actions between production and uptake is often remains opaque.[9]

This report seeks to shed light on these questions by investigating a broad variety of citizen-generated data projects. It starts with two assumptions: 1) citizen-generated data is created through **different project setups**, or chains of data exchanges. These setups may involve cross-sector collaborations–including individual citizens, civil society organisations, funders, public institutions, policy makers and private companies; 2) project setups operate in **different contexts** that shape the incentives to produce and use citizen-generated data. Based on these assumptions two research questions are addressed:

■ **What are the incentives that encourage different interest groups to produce and use citizen-generated data?**

■ **What are the methods, strategies, and resources that ensure the uptake of citizen-generated data projects?**

After a brief methodology section, the paper presents how projects operate in different contexts (Section 3). The section highlights factors which appear to support strategic uptake and success in the short and long term, and discusses to what extent these factors are context-specific. Section 4 elaborates on the commonalities between projects–including how to engage citizens to participate in different project setups, as well as general practices to facilitate uptake and longevity. Chapter 5 discusses key take-aways, as well as risks and limitations when running CGD projects.

---

9   For instance research examined the incentives of volunteers to participate in citizen science projects, or regarded the methodological issues of surveys done by citizens. For further details, see: Wiggins, A., Crowston, K. (2012): From Conservation to Crowdsourcing: A Typology of Citizen Science.
Available at: http://andreawiggins.com/research/Wiggins2011HICSS.pdf. For an overview of the challenges to do surveys and social observations, see: Purdam, K. 2014: Citizen social science and citizen data? Methodological and ethical challenges for social research. Available at: http://csi.sagepub.com/content/62/3/374.abstract.

RESEARCH METHODOLOGY

To answer the research questions, desk research was conducted looking at 160 projects known by DataShift. A total of 14 case studies were selected for in-depth semi-structured interviews. In order to cover diverse incentives and organisational models, following criteria informed the selection process. A citizen-generated data project needed to demonstrate an active and purposeful contribution of data by citizens, as well as one or two of the following:

- A unique method to produce and use data

- A distinct way of engaging a variety of different actors

Interviews with practitioners and experts as well as grey literature uncovered the perspectives of stakeholders involved to produce and use citizen-generated data. The questionnaire built on prior reports by Open Knowledge International and The DataShift. A literature review served to understand different dimensions[10] impacting on people's incentives to produce and take up citizen-generated data. The questionnaire was refined in pilot interviews (for the final questionnaire, see Appendix).

Beyond the limitations inherent to a sample, one methodological limitation of this research was the non-consistent definition of success. For most surveyed projects a long project lifespan and longevity are key criteria for success, but there are different ways to envision longevity (see Section 3.5).

---

10 The reviewed dimensions are among others: degree of civic participation in a project, the role of mutual benefits when participating in a Citizen-generated data project and the different accountability strategies that can be designed through citizen-generated data. See also Wiggins, A., Crowston, K. (2012): From Conservation to Crowdsourcing: A Typology of Citizen Science. Available at: **http://andreawiggins.com/research/Wiggins2011HICSS.pdf**; Haklay, M., et al. (2014): Crowdsourced Geographic Information Use In Government. Report to GFDRR (World Bank). Available at: **https://www.gfdrr.org/sites/gfdrr/files/publication/Crowdsourced%20Geographic%20Information%20Use%20in%20Government.pdf** As well as Peixoto, T., Fox, J. (2016). When Does ICT-Enabled Citizen Voice Lead To Government Responsiveness?. World Bank Group. Available at: **https://openknowledge.worldbank.org/bitstream/handle/10986/23650/WDR16-BP-When-Does-ICT-Enabled-Citizen-Voice-Peixoto-Fox.pdf**

Definitions of success also depend on when researchers conduct their analysis.[11] Certain projects within the sample have been running for a much longer period of time than others. Some have achieved more of their goals than others, and some of their goals have shifted over time.

As such, it is not the point of this report to portray impact as a measure of success, but instead to look at the incentives and drivers that shape the opportunities and blockages throughout the processes of a project. Nevertheless, in order to get some measure of success, the questionnaire allowed the interviewees' own opinions to surface about what they themselves considered to be the successes and challenges of the project.

---

11   See also: Heeks, R. (2002). Failure, Success and Improvisation of Information Systems Projects in Developing Countries. Development Informatics Working Paper Series, No.11/2002. Manchester: Institute for Development Policy and Management. Available at: http://ictlogy.net/bibliography/reports/projects.php?idp=199

CONTEXTUAL DRIVERS

The context within which a CGD project happens shapes the incentives producing and using citizen-generated data. An incentive is understood as any inducement encouraging an actor to participate in a CGD project. This report looks at two larger contexts that are important to help CGD projects be effective in the long term.

IS THE CONTEXT CHARACTERISED BY STRONG OR WEAK GOVERNANCE AROUND THE ISSUE?

Governance is a useful framing to analyse our case studies because all observed cases revolve around governance issues or are affected by them. It is a contested term with many interpretations. In general it describes "the structure and processes by which people in societies make decisions and share power".[12] Governance includes government as well as private actors and civil society and their relationships to one another.[13] Governance is a useful framing to analyse our case studies because all observed cases revolve around governance issues or are affected by them. A broader and more inclusive perspective of governance enables to understand which roles governments, public-private partnerships and civil society play in producing and using CGD. In a similar vein it is acknowledged that multiple actors have to be mobilised to monitor progress around the SDGs. A broader, multi-actor governance framework enables to point at the roles different actors may play in contributing CGD to the SDG monitoring. This report looks at the elements of governance that are relevant to helping CGD projects be effective in the long term. This includes: the availability of information to make decisions, the strength of laws, the level of government responsiveness, the level of trust, and the level of informality around the particular topic.[14] Importantly, in some cases, there is a strong governance context within a country with a weaker state, or vice versa (e.g. see Section 4.1). In these cases, it is the governance context *around the particular issue* that matters most.

---

12 See also: Folke, C., Hahn, T., Olsson, P. and Norberg, J. (2005) Adaptive Governance of Social-Ecological Systems. Annual Review of Environment & Resources 30, no. 1 (2005): 441–73.

13 See also Rhodes, W. (1996): The New Governance: Governing Without Government. Available at: https://www.researchgate.net/profile/R_A_W_Rhodes/publication/227979762_The_New_Governance_Governing_Without_Government/links/5405e8150cf2bba34c1ded30.pdf. See for further details: International Council for Science (2015): Review Of Targets For The Sustainable Development Goals: The Science Perspective. Available at: http://www.icsu.org/publications/reports-and-reviews/review-of-targets-for-the-sustainable-development-goals-the-science-perspective-2015/SDG-Report.pdf

14 We defined these elements of governance based on established categories defining "good governance". These elements have informed an analysis of governance elements projects are facing. For a broader reflection on governance, see for instance: United Nations Economic and Social Commission for Asia and the Pacific (n.y.): What Is Good Governance?, Available at: http://www.unescap.org/sites/default/files/good-governance.pdf

Graphic 1 Governance context

STRONG		WEAK
Highly responsive government		Unresponsive government
Strong collaborations		Weak or no collaborations
Strong legal framework	⟷	Weak legal frameworks
High trust		Low trust
Low informality		High informality
Adequate information base		Information base lacking

DOES THE PROJECT DRIVE CHANGE WITH A DIRECT OR INDIRECT LINK TO GOVERNMENT?

The goal of CGD is to monitor, advocate for, or drive change around an issue important to citizens, and successful projects establish a particular strategy to do so. Part of that strategy is the way in which to engage government, either directly or indirectly. The choice of working with government or not is influenced by whether a project aims to be complementary to government monitoring strategies, contribute to government processes and monitor performance of government[15], or whether the issue is being dealt with outside of government. Sometimes the actors able to deal with an issue are not found within government. In other cases government or private companies perceive of an issue differently than civil society does, thereby hindering partnerships from emerging.[16] Some projects may emerge out of government itself seeking to work with CGD, whereas others are grassroots initiatives complementary to government, and some represent (social) businesses working on development issues using CGD. The extent to which projects link up with the government affects the types of incentives that come into play because each project requires a different type of partnership.

15 Such as social audits, community-based monitoring, or similar projects.
16 For instance the project "Feowl" failed to stimulate uptake of its electricity outage data because neither policymakers nor electricity providers had a use for this data. Source: Interview with Nicolas Kayser-Bril, Journalism++.

HOW DOES THIS CONTEXT AFFECT THE INCENTIVES AND PROJECT SETUPS?

Combining these two dimensions creates four different types of context. Graphic 2 and 3 describe which projects operate with which strategy in different contexts.

Graphic 2 Types of projects within different contexts

		GOVERNANCE	
		Strong	Weak
LINK TO GOVERNMENT	Direct	E-government projects, helping government improve efficiency	Citizen monitoring to raise issues
	Indirect	Knowledge platforms and data repositories	Intermediaries building data infrastructures

There are several answers to what can make a project successful. Graphic 3 below describes the **four distinct approaches to CGD project success** in their corresponding contextual cluster.

Graphic 3 Approaches to CGD project success in different contexts

		GOVERNANCE	
		Strong	Weak
LINK TO GOVERNMENT	Direct	Institutionalising CGD projects into government processes by building on and restructuring existing methods	Building government and community capacities to create new processes for CGD take-up
	Indirect	Maximising economies of scale through user interactions, creating open data repositories	Projects are focused on governance issues due to information asymmetries. Data generation tools are drawn from existing media channels or provided through project partners

It is possible that these projects can be applied in other contexts, using slightly modified strategies. It is therefore not the goal of this section provide a representative clustering, but rather to spark imagination how CGD collaborations develop strategies to operate within different contexts. The following section compares different projects within each group, discussing key takeaways as well as contrasting case studies.

3.1 STRONG GOVERNANCE CONTEXT, DIRECT LINK TO GOVERNMENT

The common characteristics of the observed projects in this category are: (1) a strong partnership with key government agencies and (2) project activities aligned with existing government processes in order to complement and improve their work. The data may be handled at the discretion of government and be dedicated for government-internal use.

The direct link with government is achieved through engaging government officials who become the main users or target audiences of the data. Both parties may share financial costs or provide human resources. Contrary to a context with weak governance and a link to government, projects falling in this context benefit from existing regulatory frameworks and established government processes.

The examples for this category are projects implemented by **HUMANITARIAN OPENSTREETMAP TEAM** (**HOT**) **INDONESIA** and **MYSOCIETY**. HOT is placed in a strong governance context because although the Indonesian state may be characterised otherwise, the project itself emerged out of a strong collaboration with AusAID and regional government agencies and explicitly aimed to collect open data on Disaster Risk Reduction (DRR). This data fits into the risk models of the Disaster Response Agency's at national and local level. The project has evolved over the years, from urban mapping workshops and mapping competitions (2011) to building the capacity of government officials and universities (2015).

FIXMYSTREET is a reporting/mapping system built by the charity mySociety. It is designed to fill the gaps of existing government reporting system where citizens are unable to report problems for not knowing how to send reports or to whom they should be addressed. Likewise, local government needs to easily locate and respond to reports as well as notify the senders–functionalities that are usually lacking in the government reporting system. Since its successful implementation in the UK, the platform has been adapted by local governments and CSOs worldwide.

The observed projects last long because of political buy-in, linked to expected efficiency gains for public sector operations. The projects transfer knowledge to government, building capacity within government to independently perform or replicate the projects. This knowledge transfer requires the government to invest, either in the form of training staff members, project funding, or technology.

---

17 See also: Sjoberg, F. et al (2015): The Effect of Government Responsiveness on Future Political Participation. World Bank Group. Available at: **https://www.mysociety.org/files/2015/03/SSRN-id2570898.pdf**

Since the citizen-generated data projects aim to be integrated into government processes, it is paramount to ensure sufficient human capacities, an adequate organisation of labour and technical interoperability between civic technology and government information systems.

KEY DIFFERENCES	HUMANITARIAN OPENSTREETMAP TEAM INDONESIA	FIXMYSTREET UK
Approach	HOT solves the problem of lacking DRR data with three strategies: providing data collection and mapping platforms, developing modules, and training users to use effectively these tools	mySociety built a single platform with versatile functionalities serving both residents and government reporting needs. Governments using FMS for Councils, for example, have the option to fully integrate it with their Customer Relationship Management systems
Data uptake	DRR data is mainly used by government and university for disaster risk mapping and city planning	Reports are used in various ways: public authorities use it to locate and fix problems, citizens use it for tracking reports/response, and academic research uses it to study, for example, the correlation between government responsiveness and citizen participation[18]
Users of the data	Main users and/or stakeholders are government agencies and universities that have strong interest in disaster management	Main users of FMS are local residents and the government (i.e. local council)

HOT focusses on government capacity building for self-sustained DRR data management. The technological interoperability of FixMyStreet enables government to easily link and operate the system themselves. Both projects face some challenges such as a short span of interest from key stakeholders, shortage or high turnover of those involved directly in the project, and sometimes a lack of political will from policy makers to actually use the data to inform decision making. High-level policy formulation has to be accompanied by more complex adaptations of work routines among civil servants and be supported by the government department implementing the project. In cases where governments perceive the issue differently than civil society, one strategy is to commonly shared interests between the two perspectives. This makes it easier to fit CGD into government planning, as the case of HOT exemplifies.

CONSIDERATIONS WHEN BUILDING PARTNERSHIPS

- ■ Build on positive relationships that already exist between parties
- ■ Understand if citizens have a strong demand for the public service intended to be improved
- ■ Make the data capture tools easy to integrate into existing government processes
- ■ Search for alliances with benevolent bureaucrats or government entities whose data could be complemented by citizen-generated data
- ■ Define the problem so that it answers a question bureaucrats are already asking themselves
- ■ Encourage collaborations by allowing the data to be used by multiple actors

3.2 STRONG GOVERNANCE CONTEXT, INDIRECT LINK TO GOVERNMENT

Case studies falling within this category operate in a context with strong legal regulations There is a tendency that a great amount of information is already officially available. Projects aim to fill possible data gaps within this information. The case studies in this category tended not to seek a direct link to government in order to make change. This is partially dues to a less significant need to do so, coming from highly-developed countries. Instead, the projects are based on platforms, using economies of scale by facilitating interactions between citizens to collect citizen-generated data. In this model, the more people actively participate, the more valuable it becomes for the entire network of citizens to participate and share information, and the more granular data can be captured through the system and used for analyses of the issues the platform addresses. Diversified data can be analysed and catered to different parties interested in the issue, and so the platform serves multiple user groups simultaneously. These "network effects" are characteristic for platforms which are a "two-sided market"[18], meaning that both supply (users interacting with the platform) and demand (users interested in these interactional data) increase the value of a network. The platform model can be applied across many sectors as social traffic networks like Ma3Route and Waze, or social networks like WeFarm demonstrate. A strong regulatory framework is an environment conducive to business, and so within this context there is an incentive to monetize these platform models. The choice of whether to do so or not depends on the goals of the project.

───

18 See also: Eisenmann T., Parker, G., Van Alstyne, Marshall (2006): Strategies for Two-Sided Markets. Harvard Business Review. Available at: **https://hbr.org/2006/10/strategies-for-two-sided-markets#**

PATIENTS LIKE ME is a for-profit patient social network and real-time research platform responding to the lack of information and slow research progress on rare diseases such as amyotrophic lateral sclerosis (ALS). The platform enables citizens to track and discuss their experiences with others, building a strong support community. Patients Like Me frames its business around the idea of "give something, get something". Patients who donate data (give something), immediately receive some kind of analytics chart about how their data compares to the community's data (get something). Patients not only learn from one another but also actively contribute to research, as the project works closely with the US Food and Drug Administration (FDA) and the pharmaceutical industry. Through real-time surveys the platform receives feedback on how medication is working and symptoms are evolving. An interviewee woking with Patients Like me stated that the social network may improve the efficiency of medical research, getting sound results in 7 weeks whereas clinical trials can take 7 years.

SAFECAST is an open data platform for radioactivity data providing citizens with sensor technology to create and upload the data. The platform was initiated in response to the lack of information after the Fukushima disaster. At that time, people were mainly interested if their street was contaminated. Yet, official data only represented contamination levels on the city level Geiger counters, used to measure radioactivity, were soon sold out due to high demand and little availability. Safecast used this momentum and provided open source hardware enabling everyone to build their own measurement tool. It thus mobilized communities to capture large amounts of hyperlocal contamination values.

As the project leads mention, Safecast is nowadays seen as a trusted intermediary (see Section 4.4) for different interest groups. This stems from its commitment to neutrality towards the politically sensitive topic of nuclear energy. Safecast remains a neutral provider of data, while other interest groups can use the data for any purpose, including campaigning and lobbying government. Its database is also recognized as valid by the Fukushima regional government and has inspired other Environmental Protection Agencies to open up some of their data too.

Both projects are some of the longest lasting out of the sample as they have managed to position their platform as a trusted source of information. To do so, they pay special attention to the methodological rigour of their data collection, the granularity and the reliability of their data. Both have several methods of checking the data for validity, including manually analysing outliers and cross-referencing the data sets with environmental and insurance data. The openness of Safecast's data encourages trust because it allows everybody to verify information. Both projects position themselves as 'ethical middlemen' because they work on sensitive topics. Another commonality is the use of a 'pull' rather than a 'push' model: beyond basic outreach and marketing, for the most part citizens come to them because they are interested in the project's vision.

KEY DIFFERENCES	PATIENTS LIKE ME	SAFECAST
Advocacy styles	Patients Like Me engages directly with the medical industry to integrate their research findings and promote ethical medical research practices.	Rather than advocating for a particular stance on nuclear policy, Safecast enables citizens and other organisations to create and use the data freely.
Value exchanges	As a for-profit organisation, Patients Like Me sells most of its data to pharmaceutical companies. Researchers have access to information in order to run analyses and enrich the data provided on the platform. Collaborations with nonprofits are rare because they cannot provide commercial value.	As a volunteer-run open data organisation, Safecast focusses on providing information freely to anyone. The transparent and open architecture increases the acceptance and neutrality of data across user groups because the data are not associated with the politics of specific group.
Sources of knowledge and resources	Emerging out of an ALS clinical research trial foundation, Patients Like Me had the money and know-how in house to hire data analysts.	Emerging out of the DIY and hackerspace culture, Safecast drew resources in from a wide network of whoever could and wanted to contribute.

However, this type of CGD project depends on the resources for data analytics and to build the platform itself. For some projects this is a significant barrier because of the financial costs associated with platform development. For example, Civic Action Group seeks to provide openly available data, but the website is not yet as interactive as intended, simply because the project doesn't have to funding to pay a developer yet. Because they're based in India with laxer environmental regulations, Civic Action Group also cannot yet tap into financial incentives for businesses to use their platform and value the demand side of their analytics.

CONSIDERATIONS WHEN BUILDING PARTNERSHIPS

- Build on existing discussions on the issue
- Help people to feel part of a community around this issue
- Encourage collaborations by allowing the data to be used by multiple actors
- Provide people with useful information after contributing to the data
- Check if legal regulations and compliance frameworks support your case
- Financial penalties linked to non-compliance may support uptake of data, if the data help to comply to legal regulations

3.3 WEAK GOVERNANCE CONTEXT, INDIRECT LINK TO GOVERNMENT

WEAK GOVERNANCE CONTEXT, INDIRECT LINK TO GOVERNMENT

In this context information asymmetries across different actors play an important role. These information asymmetries are due to a lack of human capacities to monitor and capture data, a lack of trust across different actors, as well as missing organisational processes or information and communications technologies (ICT) able to capture and communicate necessary information. Indirect links to government may exist through 1) government as a funding source, 2) a project organization not primarily involving government actors but open to collaboration, 3) data that can yield value for actors both within and outside of the public sector.

Projects[19] like Premise or Africa's Voices Foundation operate in this environment as intermediaries between governments, businesses and civil society. They tap into existing media resources and communication networks to set-up information infrastructures across these actors. These infrastructures may serve individuals or communities to exchange information and connect to each other (Africa's Voices Foundation). Both examples deliver commissioned research to decision-makers aiming to adapt their funding strategies and operations in remote areas that are otherwise to hard to reach.

AFRICA'S VOICES FOUNDATION is a social enterprise providing tailored research to international NGOs, development agencies and other decision makers who are interested in the opinions and perceptions of citizens affected by their policies. To do so, Africa's Voices Foundation collaborates with media partners, such as trusted radio stations, in different regions across East Africa to engage citizens in interactive discussions and gather their opinions via SMS and social media channels. While decision makers partnering with Africa's Voices Foundation often have quantitative data about the outputs of their policies, they may lack contextual information and insight into how citizens perceive issues, socio-cultural barriers to behaviour adoption, or norms that might impact an intervention's success. By engaging citizens in conversational forums Africa's Voices Foundation enables them to articulate their concerns and share opinions. A multidisciplinary analysis of citizens' responses (local language, digital data) can uncover hidden perceptions or knowledge complementing,contextualizing and deepening existing quantitative statistics held by the clients. The drawbacks of this approach are similar to other survey-based investigations–socially sensitive topics such as gender-based violence are hard to explore.

---

19 Another project in this cluster is WeFarm. For an explanation of WeFarm's model see Box 1: How WeFarm replaces external funding by monetizing data: the social network platform model.

PREMISE is a for-profit research startup using a customizable app and a network of paid contributors in order to crowdsource data on diverse topics ranging from price fluctuations to the quality of health facilities. The data are catered to the needs of clients who otherwise do not have the capacities to reach or cover these areas. To recruit its contributors Premise also engages with community-based organisations and NGOs working in the respective regions.

KEY DIFFERENCES	AFRICA'S VOICES FOUNDATION	PREMISE
Models of incentivising	Interactive conversations on radio and social media tailored for target audiences increases relevance for the listeners. When citizens share their own voice in these forums, they feel empowered. Having their message read on-air gives them public recognition which is a key internal motivation to participate.	Paying people is an incentive for local community members to collect data that is otherwise not relevant to them. The difficulty to find and access the information is compensated through extra payments.
Channels of engagement	Interactive radio shows serve to engage citizens in debates and to understand the issues affecting them. Radio stations may receive training by Africa's Voices on how to improve their moderations and can improve the engagement with their audiences. Depending on the research question Africa's Voices collaborates with different radio stations. Participants engage by sending free SMS or social media messages.	Premise customizes its application and engages with local organisations to plan when, where and how often how many contributors have to collect information (flexible scalability).
Data produced	Mix of qualitative and quantitative data: individual and collective beliefs and behaviours, as well as socio-demographics information.	Mix of quantitative and qualitative data: geo-locations, images, and text-based descriptions.
Data uptake	Africa's Voices employs survey-style investigations into people's perceptions. Gaining a depth of understanding of issues for different groups is more important than creating representative data. The clients interested in this kind of data may already have quantitative (output) indicators. Africa's Voices provides the data to understand the "why" behind these numbers better (for instance, why do policy interventions not bring expected results?).	Multiple data points, some data is comparable and being used as baselines for quantitative assessments, such as price indicators for taxes. Clients use the data to understand patterns such as spreadings of diseases or market price fluctuations.

Both projects answer immediate strategic and policy needs of policy-makers, international NGOs, business (in the case of Premise) without necessarily having the intention to undertake long-lasting data collection.[20] Instead, the goal is to provide infrastructure that can connect clients with citizens on the ground who are the recipients and data producers by building on existing resources: Premise uses smartphones as people go about their lives for a secondary income, and Africa's Voices Foundation uses the popular radio platform, mobile phones, and social media. Besides some methodological drawbacks, providing tailored research and analytics may also cause high "first-copy costs"[21]. Africa's Voices Foundation therefore seeks to establish long-term partnerships with its clients–not only to reduce the overhead costs tied to the development of a research project, but more importantly to enhance the impact of their insights through more engaged, collaborative partnership.[22]

CONSIDERATIONS WHEN BUILDING PARTNERSHIPS

- Ensure that data collected by citizens will not be used against them
- Use data to help citizens in a crisis
- Gain political support from prominent organisations
- A rigorous methodology can promote transparency,
  political neutrality & trust
- Increase local acceptance by using local employees & language
- Provide adequate offline support for citizens to engage
- Provide people with useful information after contributing to the data
- Check if legal regulations and compliance frameworks support your case
- Financial penalties linked to non-compliance may support uptake of data,
  if the data help to comply to legal regulations

20 As described on page 30, WeFarm aims at building a sustained social network gaining its value from user interactions. These interactions enable farmers to tackle the issues they are facing by themselves (self-regulating system) while allowing other top-level decision makers to understand the issues they are facing and adjusting their supply chains and interactions with smallholder farmers accordingly.

21 Costs for producing a final version of a research analysis which can later be replicated and disseminated at significantly lower costs.

22 Overhead costs can come from iterative research design to understand a client's needs, the insights that matter most to it, and how insights should be communicated.

3.4 WEAK GOVERNANCE CONTEXT, DIRECT LINK TO GOVERNMENT

This context cluster is defined by a lack of information flows and organisational processes put into place to capture information. The goal is to improve government processes. The projects tend to employ differing forms of community-based monitoring initiatives in developing countries, and require multi-partnership collaborations to strengthen the governance context.

The public private partnership of **MOBILE BIRTH REGISTRATION** between UNICEF, the Government of Tanzania, the mobile phone operator Tigo and VSO International, aims to simplify birth registration for children under age 5. Usually parents who want to register newborn children need to travel to a district registrar with often prohibitive travel costs. Mobile Birth Registration places two registration officers in local hospitals and health clinics. This minimises travel distance for parents who can register their children during their regular health check-up. Trained registration officers send birth information via SMS to a central database at national government (e.g. Registration, Insolvency and Trusteeship Agency "RITA" HQ) in real time. The project also works with regulators on the legal framework for replication and scaling up the system nationwide.

DEVELOPMENTCHECK is an app developed by the organisation Integrity Action, aiming to foster the capacities of local and national organisations to do community-based monitoring. Integrity Action primarily collaborates with organisations who are already familiar with community-based monitoring strategies. The app DevelopmentCheck allows these projects to capture information on the transparency and efficiency of public service delivery. Integrity Action transfers monitoring responsibilities to community monitors with good reputation in the communities. The goal is to stimulate a culture of accountability between community and government by sensitizing local communities for their rights to public services, building confidence within local groups to monitor the issues surrounding them, and actively fostering collaborations with government actors. Community monitors are trained by a community-based organisation. They visit public facilities and survey other community members, providers of public services, or take account of public service delivery.

KEY DIFFERENCES	MOBILE BIRTH REGISTRATION	DEVELOPMENTCHECK
Accountability	Upwards accountability: citizens do not directly use the system, but rather data are internally analysed and used for government tracking of birth registration. Civil servants are accountable to higher-level administration.	Downwards accountability: the local community engages with government through the support of the app to monitor issues and work collaboratively to solve them. Government is accountable to citizens, and non-effective delivery increases political costs.
Support	After piloting in one region, national government gave political backing and was replicated in other regions. Strong ownership from the government results in more data collection.	Community monitors have to build trust and authority within a community and when interacting with civil servants. This can stimulate a "culture of accountability", yet relies on individual motivation.
Costs	Citizens who routinely visit a health facility can automatically register their child , integrating the service into already existing processes and reducing government costs. Collaboration with Tigo supports the technological capacity.	Because monitors have to be physically present, Integrity Action pays community monitors a stipend for travel costs. There is also an investment in a minimum level of education and training needed to ensure a proper data collection.

Longevity of these projects is dependent on expanding their geographic coverage and institutionalizing their uptake into government processes. Because of the nature of monitoring local issues, community-based monitoring initiatives are highly context sensitive and therefore difficult to replicate exactly across contexts. An interviewee from Integrity Action states that the successes in improving public services differ across geographic region and public sectors activities monitored.[23] Mobile Birth Registration is equally sensitive to local political contexts. The project started with small-scale pilots to gain acceptance from local government first before being replicated across regions.

Particularly in developing contexts such as in Tanzania, the laws shaping birth registration are based on colonial legacies and therefore do not capture all that is necessary for the local context. Therefore, Mobile Birth Registration developed a proof of concept, not so much to show that a process works, but rather to change the government's mindset and the political context. As interviewees state, despite a law prescribing the registration of newborn children, the government lacks adequate registration process. Birth registration data is patchy and Tanzania's government not

23 Interview with Integrity Action, October 2016.

experienced handling and using this type of data. Part of Mobile Birth Registration's proof of concept is to demonstrate the applicability of registration data while at the same time lobbying for a more supportive legal framework prescribing fine-grained data collection.

Building trust and partnerships with key government, CSOs and private sectors help to drive these changes. This process is encouraged by not trying to reinvent the wheel but rather building on existing government processes: done very explicitly in the Mobile Birth Registration and more subtly in Development Check, which starts from a shared acknowledgement of a problem among community-based organisations and government.

CONSIDERATIONS WHEN BUILDING PARTNERSHIPS

- Government must feedback to citizens about how data is used
- Use opportunities for the government to get 'good publicity'
- Increase geographic coverage & ownership through decentralisation
- A rigorous methodology can promote transparency, political neutrality & trust
- Increase local acceptance by using local employees & language
- Provide adequate offline support for citizens to engage
- Search for alliances with benevolent bureaucrats or government entities whose data could be complemented by citizen-generated data
- Define the problem so that it answers a question bureaucrats are already asking themselves

OVERARCHING INCENTIVES

The surveyed CGD projects often deal with cross-cutting incentives which recur across multiple contexts. In the *short-term*, CGD projects need to consider the supply side of data. Who is able to provide the data and how can the provision of data be incentivised? In the long-term, citizen-generated data projects have to build mechanisms that allow for a flexible platform development and knowledge sharing. Since funding is one of the key elements supporting project success, we have dedicated a section to funding models.

4.1 ENGAGING CITIZENS TO PRODUCE DATA

Because citizen-generated data is consciously given by citizens, each project has to consider the incentives for data contribution. Citizens will contribute data if it benefits them. Crucially, the CGD project should have a clear answer to the following question: 'What does the person get in exchange for contributing data?'. The issues and data can be defined from the bottom-up by citizens so they can encourage other actors to take ownership of the problem. Issues can also be defined from top-down so non-civic actors already have ownership of the problem. Their task is then make citizens sensitive towards the issue. In the latter case, CGD projects have to pay attention to the transactions that activate citizens to contribute data. In some cases **citizens do not necessarily need to "own" the problem to start collecting data** when other incentives and benefits are strong enough.

Effective projects build a chain of exchanges: each 'link' in the chain is an exchange between different actors who each gain value at that particular exchange. All surveyed projects made us of such **exchanges**. In an exchange, however small it may be, the two exchanging parties both give and receive something of value to them. Exchange means being on equal footing because the process is beneficial to both parties, often for different reasons. This does not mean that both parties have equal power, but that both benefit in a way that they value. Effective projects incorporate an appreciation for this exchange into the design of the data collection and use.

CGD projects should consider how to fit within the citizen's worldview. It is important to explain what a project wants to achieve and to demonstrate how

citizens can be useful in achieving the goal.[24] For instance, virtual projects aiming to analyse data on hydraulic fracturing (also known as 'fracking') do not only have to be interesting for experts like geologists, but can also resonate with the concerns of laypersons and the issues within their daily lives.[25] In some cases, communities can emerge from the data that offer citizens a platform to exchange knowledge and experiences (see for example Patients Like Me, or WeFarm).
On a very practical level, there is the question of **ease of use**: how easy is the technology for CGD? Is it something already embedded in people's lives within that context, or do they have to make an extra effort in order to engage with the chosen technology?

4.2 FINANCIAL INCENTIVES FOR CITIZENS

The choice whether and how to remunerate citizens for participating within a project needs a nuanced analysis. There are three options for dealing with monetary incentives.

1. PAY PEOPLE TO COLLECT DATA

Often, smaller stipends are paid to cover operational costs that arise for citizens during the data collection. In other cases, payments are used to incentivise individuals to collect data which is otherwise of little concern for them. These payments may require significant financial resources, which is why large-scale data collection efforts may involve clients who subsidise the data collection around a specific question (cf. Premise). Investing in staff members may be a prerequisite for running a CGD project, as exemplified with the recruitment of registrar assistants for the Mobile Birth Registration project.

Deliberate payments may jeopardise sustained and autonomous data collection by citizens if problem-ownership has to be transferred to them (institutionalising, for example, their role as community monitors). As a representative of Integrity Action states, payments exceeding small cost compensations may make community monitors overly dependent on repeated payments. Once they have received payments for doing their job as monitors, it will be hard to cut these payments without stopping their participation. Payments may also undermine the monitor's

---

24 This issue-driven approach to citizen-generated data is repeatedly found throughout the crowdsourcing literature. See for further details: Haklay, M. (2013): Haklay, M., et al. (2014). Crowdsourced Geographic Information Use In Government. Report to GFDRR (World Bank). Available at: https://www.gfdrr.org/sites/gfdrr/files/publication/Crowdsourced%20Geographic%20Information%20Use%20in%20Government.pdf as well as Brabham, D. 2013. Using Crowdsourcing in Government. Collaboration Across Boundaries. IBM Center for The Business of Government: Washington, DC. Available at: http://www.businessofgovernment.org/report/using- crowdsourcing-government
25 See also: Doherty, J. (2016): Community Matters. SciFabric shares the secret to successful citizen science projects. Available at: https://scifabric.com/blog/2016/10/28/Community-matters.html

identification as a trusted authority who can hold government to account. However, our findings are different to past research arguing that citizens' motivation to produce data can change over time. Volunteers may develop a passion to continue collecting data if they sense a strong value in the work they are doing. Similarly, research suggests payments can be an initial incentive before being replaced by intrinsic motivations.[26]

2. PEOPLE COLLECT OR PRODUCE DATA FOR FREE

When people donate data, they feel they are contributing to a cause. Social values and feelings of community-belonging become important (see section 4.1). This is also why some projects monetising data do not want to pay users. It 'changes the dynamics'[27] between users. Employing a voluntary data collection model also reduces the cost overhead and enables projects to mobilize larger crowds in data collection. However, this means that data collection is dependent on other incentives and, especially, how interested people remain in the project goals over the long term.

3. PEOPLE HAVE TO INVEST IN DATA COLLECTION

When people invest money, time, or other resources into the data collection, they may care about it from the beginning. On the contrary, requiring citizens to invest in resources can negatively affect who is able to participate, skewing social representativity of the data towards higher socio-economic demographics. Different investments are conceivable: Safecast observed that their users stay engaged over time by paying for the data capture device. The purchase means that users are already motivated to capture specific information before the actual data collection. Another interviewee working with Integrity Action argues that communities may fundraise monitoring projects after an initial grant has helped design a proof-of-concept for their monitoring activities. This model is, however, only applicable to citizens in urban areas with higher socioeconomic status.

26 See Haruvy, E. Wu, F., Chakravarty, S. (2005): Incentives for Developers' Contributions and Product Performance Metrics in Open Source Development: An Empirical Exploration.
Available at: http://www.iimahd.ernet.in/publications/data/2005-03-04sujoy.pdf.
27 Interview with Patients Like Me, October 2016.

4.3 ALIGNING SHARED INTERESTS

Shared interests and values are the foundation of effective partnerships. Finding key points of overlap in shared interests enables different actors with different interests to collaborate. The case studies show that some projects struggle with take-up because they fail to provide data that can align the interests of many actors.[28]

Aligning interests is especially important when data is political. In the absence of governance mechanisms capable of alleviating an issue, data may divert actors from taking up data. Issues requiring follow-up actions can increase political costs of the responsible organisation if it is not capable of or unwilling to tackle the issue.

A positive example of working with politicised data is the Land Matrix Initiative, which is an independent project to promote transparency and accountability in decisions on land purchases and investments in low- and middle-income countries. Many different types of actors are interested in land, and so the platform includes about 100 different variables in order to capture as much information that could be useful to a wide variety of people. Because large-scale land deals are mostly opaque, Land Matrix Initiative only captures a fragment of these. Instead of aiming to be comprehensive or reliable, the database draws its value from indicating tendencies and trends, and land data can be complemented by its users with various other data to gain better knowledge around issues related to land transactions.[29] Thus Land Matrix Initiative offers a starting point of inquiry for academic research, but also for campaigners, governments and companies.

Particularly for projects engaging government, an effective strategy is to **complement existing work processes and organisational routines**. For example, Fix My Street aims to make it easier for citizens to report local problems by directly linking reports to existing information management systems within government. This means they help support the aims of citizens to get their problems solved as well as government's aim in greater efficiency in collecting citizen reports, without stepping on any toes.

---

28 Feowl for instance allowed citizens to report power outages by SMS. The service seeked to map outages users of energy grids are facing (persons who take electricity from the grid, without necessarily being customer). However, top-level actors like energy providers did not see the need for this data, because their primary concern is customer support and satisfaction.

29 For instance a research project investigated to what extent land acquisitions are aiming towards groundwater acquisitions. See also: Breu. T. et al. (2016): Large-Scale Land Acquisition and Its Effects on the Water Balance in Investor and Host Countries.
Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150901

One effective strategy is to **use the same data to frame multiple issues**. In the case of Civic Action Group creating flood risk maps in Chennai, one individual working with a community-based organisation was able to mobilise different communities around flood risk maps by framing the task as important for drinking water scarcity, which was perceived by many interest groups to be a urgent issue.

A similar, but slightly different strategy, is to recognise and value **different interests for data production and use** by different actors along the chain of data exchanges. For example, some projects like Patients Like Me and WeFarm use social networks (see section 4.2.2) to build a community between citizens who benefit from interactions with their peers, as well as providing relevant data for other actors (such as private companies and researchers). In these cases, citizens and researchers have different interests in interacting with data at different steps in the chain of data exchange. Since Patients Like Me treats sensitive information of personal health conditions, their information policy not only informs patients that their data serves to advance medical research, but citizens perceive their contribution to research as an active long-term benefit.

4.4 FLEXIBILITY AS KEY FOR LONG TERM ENGAGEMENT

Projects that are resilient and able to flexibly adapt to changing contexts may increase their longevity. Flexibility is an offshoot of being issue-driven rather than tech-centred. When the issue matters more than the how, it allows the how to change over time as relationships and technologies evolve.Flexible projects also employ data to align different interests and propose that the data they collect may evolve together with changing issues over time.

STRATEGIES FOR FLEXIBILITY

We found two common threads to encourage flexibility in CGD projects. Here are the implications of each.

COMMITMENT TO OPENNESS & TRANSPARENCY

As a principal philosophy behind an organisation or project, an open communication of the data collection methods and methodological weaknesses, may breed trust from all actors involved, especially from citizens contributing their data. Particularly in contexts of weak governance, or when business actors are involved, trust helps the organisation to maintain an ethically sound reputation.

An open and transparent communication also helps projects to remain neutral around politically or personally sensitive topics, preventing political concerns from stopping the CGD project. For example, Safecast facilitates the open data collection

on radiation and air pollution measurements as a reaction to the lack of information after the Fukushima nuclear disaster in Japan. They explicitly position themselves as neutral, not favoring either side of the debate about nuclear energy, but rather providing the data to allow others to use it as they see fit. They have become a trusted source of information which encourages the platform and organisation to run long after the initial panic of Fukushima subsided.

INTEROPERABILITY.

CGD in interoperable formats remains relevant in the long term. Interoperability allows data to be used for multiple purposes, including those which have not yet been thought of yet. This is important if the specific framing of issues and concerns changes over time. It also incentivises partnerships and buy-in from multiple interests, which helps CGD become more relevant to more people and build collaborations (see above). The classic counter-example is the relatively common practice of organisations sharing their data in pdf format, which makes it difficult for others to use.[30]
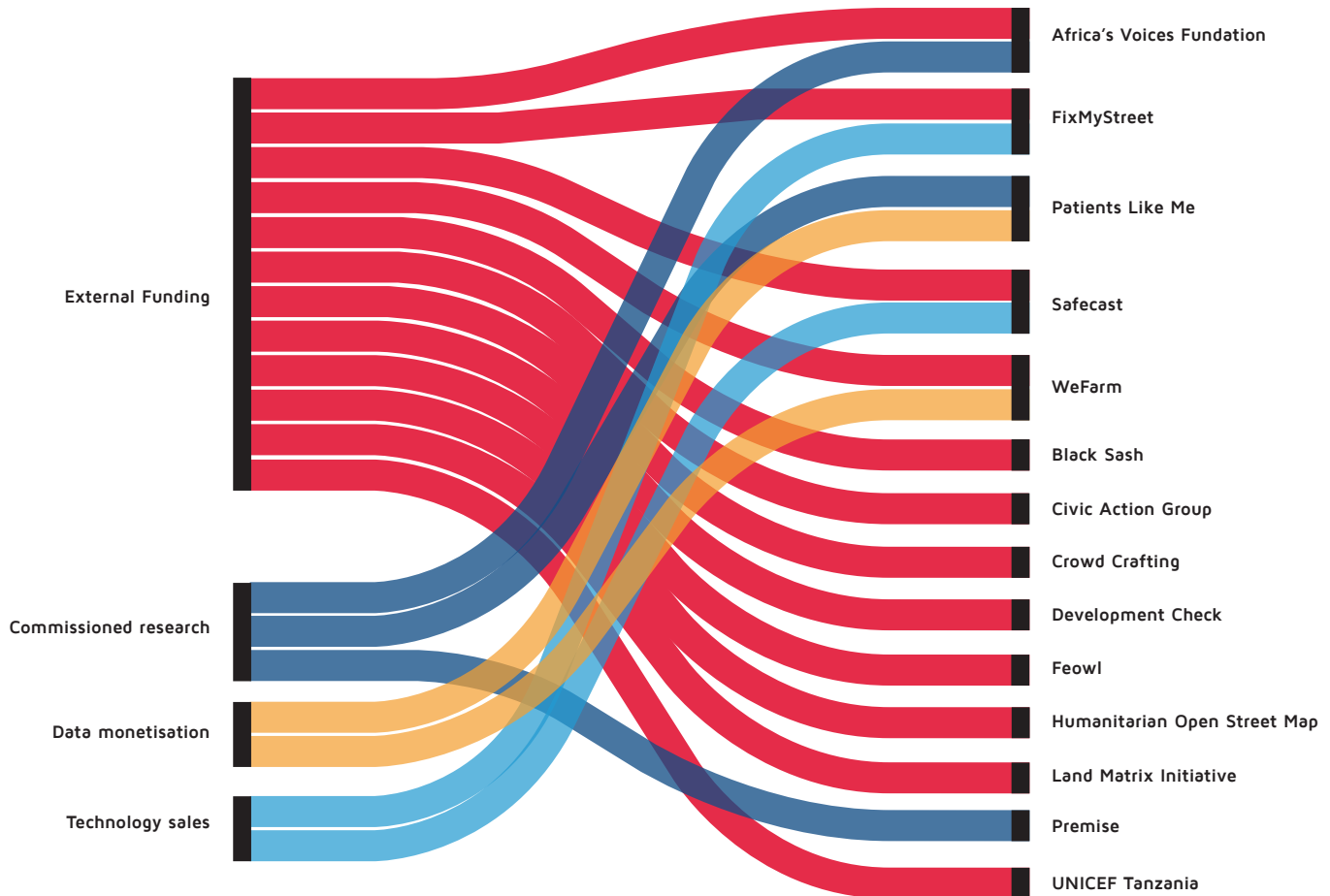
4.4.2 EXPLORING REVENUE STRATEGIES

Funding streams are a key consideration for the success of any project. Some projects are in a "proof-of-concept" phase and therefore depending on external funding to sustain their operations before they are able to generate sufficient income. Other projects are exploring different revenue strategies and diversifying their income streams. The report observed that some projects monetise CGD operations and generate revenue internally. Throughout the interviews **four main sources of financial support** for CGD projects emerged:

- **EXTERNAL FUNDING** describes any money allocated to a project through external parties–including grant funding through development agencies, foundations and governments (Land Matrix Initiative, Humanitarian OpenStreetMaps, Civic Action Group, FixMyStreet), as well as seed investment (Patients Like Me, WeFarm), or allocation of funds through public-private partnerships.

- **COMMISSIONED RESEARCH WORK** harnessing CGD (Premise, Africa's Voices Foundation).

- **MONETISATION OF DATA** through a mixed model of free network services and paid analytics (WeFarm, Patients Like Me).

- **SALES OF TECHNOLOGY**, measurement instruments and infrastructure (Safecast, FixMyStreet).

30 See Tabula, a tool to grab tables locked in PDF files: **http://tabula.technology**

Graphic 4 demonstrates the most often occurring income sources. *It does not compare the size of funding per source*, and neither does it represent a general distribution of funding sources across CGD projects. It's aim is to visualize the diversity of revenue sources and how these relate to individual projects.

Graphic 4 Sources of funding per project

There is a clear dependence on external funding sources such as grants to sustain project operations. Yet, other models are also emerging, particularly mixed funding models which include both sales, commissioned research and data monetisation. The for-profit companies Patients Like Me and Premise are the only projects generating their own revenue independent from grant funding. These mixed funding models tend to be open and closed platforms (see Section 3.2), one example of which is the social network model, as explored in Box 1 (*How WeFarm replaces external funding by monetising data: the social network platform model*).

BOX 1

HOW WEFARM REPLACES EXTERNAL FUNDING BY MONETISING DATA: THE SOCIAL NETWORK PLATFORM MODEL.

WEFARM describes its service as "a peer-to-peer learning platform that allows farmers to easily access agricultural information via SMS". As a social network, WeFarm fosters peer-to-peer interactions, similar to Patients Like Me. The main clients, smallholder farmers, see their peers as trusted information sources, contrary to agribusiness representatives and government officials. WeFarm captures and analyses the issues of farmers emerging through their SMS interactions. The analytics "help businesses make better decisions regarding their supply chain", and to develop support strategies for farmers. Agribusiness companies may gain insights in the problems of individual farmers in specific locations, to which they would normally have no access. WeFarm aims to close the feedback loop between farmers and agribusiness by enabling companies to offer targeted training and allocate support to farmers, and to adapt their resource allocation to the supply chain accordingly.

Like other platforms WeFarm seeks to increase the amount of data created by its users to increase value of both the supply and demand sides of the chain and be able to monetise the data. The goal of scaling the network is to cross-subsidise the service through revenue generated from data analytics. Scaling is forecasted to decrease the fixed costs of the service while generating increased revenue through more diversified analytics. So far, WeFarm depends on seed investment to finance its free SMS service for farmers.

DISCUSSION

DIFFERENT BENCHMARKS FOR CGD PROJECT SUCCESS

CGD serves a goal and is meant to 'speak to' somebody. A common understanding of success is whether a project has managed to communicate this goal clearly. Longevity can be another part of success, but not all projects aim to be long-lasting, and rather solve immediate or temporary issues. Longevity may be an important aspect of strategic uptake. This is especially important for projects including governments. Such projects achieve longevity through political buy-in, linked to expected efficiency gains for public sector operations. In this case projects transfer knowledge to government, building capacity within government to independently perform or replicate them. This knowledge transfer requires the government to invest, either in the form of training staff members, project funding, or technology. Projects monitoring government performance need to produce CGD over a long period of time in order to understand whether government programs (such as investments in services) meet their intended goals.

Some projects develop social networks around an issue citizens care about, enabling them to produce data as by-products of their (routine) interactions (including Patients Like Me, WeFarm). The benchmark for success is a network's relevance for different user groups. This is commonly achieved by increasing the number of users and interactions. The primary goal is to scale the network over time and increase the amount of data that are produced on the platform which can be catered to other users. Some projects explicitly gain their relevance from running many short-lived projects, targeted towards a specific need for data (commissioned research). Interviewees mentioned that long-term partnerships with clients enable to learn about the needs of a client facilitating the design of research but also increasing its impact.

CITIZEN-GENERATED DATA DOES NOT HAVE TO BE STANDARDIZED OR REPRESENTATIVE TO SERVE ITS PURPOSE

Some of the observed projects depend on CGD that is produced in a consistent way to monitor a phenomenon over time (such as community monitoring projects measuring changes of public service performance). Yet, some projects embrace the non-standardized and non-representative nature of CGD: for instance unstructured text messages sent via SMS allow to understand a broad variety of issues. When combined with an adequate trigger such as a survey, these messages can yield time-bound, targeted, and context-specific information enriching already existing quantitative indicators.

UNICEF for instance uses Africa's Voices Foundation's text analyses to understand collective perceptions of citizens preventing its development programs to succeed. The data sheds light on the perceptions of small and potentially marginalized sub-populations. Since it is highly granular and interpretative information the CGD is not aimed towards providing a representative picture of an entire population. In other cases CGD does not provide comprehensive data but shows trends and gives partner organisations pointers to identify an issue and focus their own data collection efforts to gain more insights into the issue. Yet, even though CGD projects sometimes deliver only patchy data they establish data verification methods - including the triangulation of data to detect outlier data, or rigid design procedures for data collection frameworks.

TECHNOLOGY USED IN CGD PROJECTS MAY EXACERBATE CITIZENS' ISSUES IF THE CONTEXT IS NOT UNDERSTOOD
CGD projects should always take into account existing power structures and understand the political, social, and legal dynamics of the context in which they operate. Otherwise technology may exacerbate existing power asymmetries, particularly in a weak governance contexts. The mere use of technology to collect customary tenure information in indigenous communities can for instance exacerbate situations for marginalized populations when this data is not accompanied by legal procedures ensuring their responsible use.

CITIZENS PRODUCE THE DATA - BUT NOT ALWAYS HAVE OWNERSHIP OVER THEM
The legal rules for data storage and data ownership have implications for to what extent data intermediaries can monetize, or sell the data. Especially as CGD projects search to diversify their revenue streams, it is important that income strategies remain ethical and in line with the promotion of inclusive human well-being. Data ownership laws are also important for privacy concerns, especially when CGD collects personal or sensitive information. Some CGD projects may position themselves as ethical middlemen highlighting their role as neutral and transparent intermediaries.

THERE ARE DIFFERING DEGREES OF PARTICIPATION AND ISSUE OWNERSHIP
Participation is a defining element of citizen-generated data, and projects vary widely in how much they let citizens participate. Some projects engage citizens only in the data production, while others let them take broader ownership, from issue definition to data analysis and use.[31] Particularly within a multi-

---

31 For detailed information on citizen-generated data and participation, see also: Wiggins, A., Crowston, K. (2012). Goals and Tasks: Two Typologies of Citizen Science Projects. Syracuse University. Available at: **https://crowston. syr.edu/sites/crowston.syr.edu/files/hicss-45-final.pdf**

stakeholder partnership this question gains importance. The different roles actors play throughout the project (who is the producer of data, who defines the data and the issue, who benefits in what way from the data?) has implications for how much ownership they feel over the issue, and when. For instance, community organisations, international NGOs, or donors might identify an issue, sensitise citizens for the issue, and train them to collect the data - with the result of giving them a feeling of empowerment.

RISKS AND POTENTIAL OF PLATFORM MODELS FOR CGD

Since platforms are two-sided markets catering citizen-generated data to other organisations there are often significant financial incentives to sell citizens' personal data, either as a good in and of itself, as known from Facebook, or as a basis for research, such as with Patients Like Me. There is the risk of using citizens as mere sensors and disconnecting them from how the data are eventually used and by whom. These risks do not have to outweigh the benefits of platforms however. For example, Safecast's open communication enables citizens to learn, create, and reappropriate radiation data for their own campaigning efforts

METHODOLOGICAL REFLECTION AND LIMITATIONS

Though our case studies were selected on the basis of representing a wide range of issues and project setups, some initially selected CGD projects were excluded due to being unsuitable with the research objectives, difficulties in reaching team members, and time constraints. In light of this limited sample, the research process paid particular attention to verification and triangulation of our findings by iterative reflection and feeding our findings back into future interviews with a variety of actors with relevant expertise in CGD projects, such as academics, NSOs, volunteers, CSOs, and donors. Nevertheless, especially as a qualitative research piece, the findings are designed to inspire questions for other CGD projects, not be a generalisable blueprint to replicate exactly.

REFERENCES AND FURTHER READING

Badiee, S. et al.(2016): The State of Development Data Funding 2016. Available at: http://opendatawatch. com/wp-content/uploads/2016/09/development-data-funding-2016.pdf

Brabham, D. 2013. Using Crowdsourcing in Government. Collaboration Across Boundaries. IBM Center for The Business of Government: Washington, DC. Available at: http://www.businessofgovernment.org/report/using- crowdsourcing-government

Breu. T. etl al. (2016): Large-Scale Land Acquisition and Its Effects on the Water Balance in Investor and Host Countries. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150901

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

Cohen, J. (2012) Configuring the Networked Self: Law, Code, and the Play of Everyday Practice. Georgetown Law Faculty Publications and Other Works.

Datashift (2016). Using Citizen Generated Data to monitor the SDGs: A Tool for the GPSDD Data Revolution Roadmaps Toolkit. Available at: http://www.data4sdgs.org/s/Data4SDGs_Toolbox-Citizen_Generated_Data_for_SDGs.pdf

Doherty, J. (2016): Community Matters. Scifabric shares the secret to successful citizen science projects. Available at: https://scifabric.com/blog/2016/10/28/Community-matters.html

Eisenmann T., Parker, G., Van Alstyne, Marshall (2006): Strategies for Two-Sided Markets. Harvard Business Review. Available at: https://hbr.org/2006/10/strategies-for-two-sided-markets#

Folke, C., Hahn, T., Olsson, P. and Norberg, J. (2005) Adaptive Governance of Social-Ecological Systems. Annual Review of Environment & Resources 30, no. 1 (2005): 441–73.

Gray, J. (2015). Democratising the Data Revolution: A Discussion Paper. Open Knowledge. Available at: http://blog.okfn.org/2015/07/09/democratising-the-data-revolution/,

Gray, J., Lämmerhirt, D. (2015): Changing What Counts: How Can Citizen-Generated and Civil Society Data Be Used as an Advocacy Tool to Change Official Data Collection. Available at: http://civicus.org/thedatashift/wp-content/uploads/2016/03/changing-what-counts-2.pdf

Gray, J. and Lämmerhirt, D. (forthcoming): Data And The City. How Can Public Data Infrastructures Change Lives in Urban Regions?

Haklay, M., et al. (2014). Crowdsourced Geographic Information Use In Government. Report to GFDRR (World Bank). Available at: https://www.gfdrr.org/sites/gfdrr/files/publication/Crowdsourced%20Geographic%20Information%20Use%20in%20Government.pdf

Harvuy, E., Wu, F., and Chakravarty, S., (2005) Incentives for Developers' Contributions and Product Performance Metrics in Open Source Development: An Empirical Exploration. IIMA Working Paper, Indian Institute of Management Ahmedabad. Available at: http://www.iimahd.ernet.in/publications/data/2005-03-04sujoy.pdf.

Heeks, R. (2002). Failure, Success and Improvisation of Information Systems Projects in Developing Countries. Development Informatics Working Paper Series, No.11/2002. Manchester: Institute for Development Policy and Management. Available at: http://ictlogy.net/bibliography/reports/projects.php?idp=199

Higgins. K., Cornforth, J. (2015): Civil society and SDG monitoring: Harnessing Civil Society and Citizen-Generated Data. Available at: http://civicus.org/images/SDG%20Monitoring%20-%20DataShift%20Background%20Note%20-%20Final.pdf

International Council for Science (2015): Review Of Targets For The Sustainable Development Goals: The Science Perspective. Available at: http://www.icsu.org/publications/reports-and-reviews/review-of-targets-for-the-sustainable-development-goals-the-science-perspective-2015/SDG-Report.pdf

Irby, E. (2015): Kenya Learning Paper 2015–Closing the Loop: Combining Community Integrity Building with Integrity Education. Available at: http://integrityaction.org/publication/kenya-learning-paper---2015-closing-loop-and-combining-community-integrity-building

Kennedy, H. et al. (2016). Introduction: Data & Agency. Big Data & Society. Available at: http://bds.sagepub.com/content/data-agency

Knight, R. (n.y.): Leveraging Community Land Documentation Processes to Strengthen Intra-Community Land and Natural Resources Governance. Available at: http://www.landandpoverty.com/agenda/pdfs/paper/knight_full_paper.pdf

Partnership for Statistics in Development in the 21st century (2016): Support to statistics remains low, putting SDG monitoring at risk. Available at: http://www.paris21.org/sites/default/files/PRESS-2016-web-final.pdf

Peixoto, T., Fox, J. (2016). When Does ICT-Enabled Citizen Voice Lead To Government Responsiveness?. World Bank Group. Available at: https://openknowledge.worldbank.org/bitstream/handle/10986/23650/WDR16-BP-When-Does-ICT-Enabled-Citizen-Voice-Peixoto-Fox.pdf

Prosterman, R. (n.y.). Cost-Effective Methods for Strengthening Land Rights in Africa. Available at: http://www.landesa.org/cost-effective-methods-for-strengthening-land-rights-in-africa/

Purdam, K. 2014: Citizen social science and citizen data? Methodological and ethical challenges for social research. Available at: http://csi.sagepub.com/content/62/3/374.abstract

Rhodes, W. (1996): The New Governance: Governing Without Government. Available at: https://www.researchgate.net/profile/R_A_W_Rhodes/publication/227979762_The_New_Governance_Governing_Without_Government/links/5405e8150cf2bba34c1ded30.pdf

Salamon, L., Haddock, M. (2015): SDGs and NPIs: Private Non-Profit Institutions–The Foot Soldiers for the UN Sustainable Development Goals. Available at: http://ccss.jhu.edu/wp-content/uploads/downloads/2015/11/SDGs-NPIs_Salamon-Haddock_5.2015.pdf

Scholz, T. (forthcoming): Platform Cooperativism. Challenging the Corporate Sharing Economy. http://www.rosalux-nyc.org/wp-content/files_mf/scholz_platformcoop_5.9.2016.pdf

Sjoberg, F. et al (2015): The Effect of Government Responsiveness on Future Political Participation. World Bank Group. Available at: https://www.mysociety.org/files/2015/03/SSRN-id2570898.pdf

United Nations (2015): Ensuring no one is left behind: A human rights-based approach to data. http://www.ohchr.org/Documents/Issues/MDGs/Post2015/DataDisaggregation.pdf

United Nations Economic and Social Commission for Asia and the Pacific (n.y.): What Is Good Governance?, Available at: http://www.unescap.org/sites/default/files/good-governance.pdf

Wiggins, A., Crowston, K. (2012). Goals and Tasks: Two Typologies of Citizen Science Projects. Syracuse University. Available at: https://crowston.syr.edu/sites/crowston.syr.edu/files/hicss-45-final.pdf

APPENDIX

QUESTIONNAIRE

DIMENSIONS	QUESTIONS
Organisations	Which organisation(s) is involved with the project? How did the project start?
Issue definition	How is the issue defined and by whom? Was the project triggered by a specific need or event? How does the data address this issue?
Data production	Who produces or collects the data? What kind of data is produced, using what tools?
Data use	How is data used and by whom? What different purposes is the data used for?
Methodology	What elements of methodology are paid specific attention to and why? How does the project address methodological challenges of representivity, coverage, comparability? To what extent do individuals participate in the definition of data and methods to capture data?
Collaboration	How do different actors collaborate, or not? Which common goals do they have? What resources and benefits do these collaborations bring? Why do people want to engage with the project? Has this changed over time?
Communication style	Is there a collaborative or confrontational style of communication? Is one actor addressed with the data (e.g. government), or are there two-way interactions? Is feedback incorporated and at what stage?
Contextual drivers	What contextual drivers significantly affected why the project works as it does? What economic, organisational, technological, socio-political, cultural, environmental factors are important?
Opportunities & Blockages	What specific opportunities helped the project's development? What specific blockages hindered it? How were challenges overcome?
Vision	What is the ideal goal? What opportunities do they see coming in the future?
Benefits	What short-term benefits does citizens and actors receive from creating and using the data? What long-term benefits are observed or expected?

Danny Lämmerhirt, Shazade Jameson,
Eko Prasetyo, *Making citizen generated data
work: Towards a framework strengthening
collaborations between citizens, civil society
organisations, and others.*

Printed on recycled paper/FSC.
Graphic design by Federico Pinci

FSC

Join the **DataShift Community** of civil society organisations, campaigners and citizen-generated data and technology practitioners by signing up at **www.thedatashift.org** and follow us on Twitter via **#datashift**

DataShift is an initiative of **CIVICUS**, in partnership with **Wingu**, **The Engine Room** and the **Open Institute**. We are part of a growing global community of campaigners, researchers and technology experts that is using citizen-generated data to create change.

CIVICUS

Open Institute

THE ENGINE ROOM

Wingu
Tecnología sin fines de lucro