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 INTERNATIONAL

INTRODUCTION

Citizen-generated data (CGD) needs partnerships to thrive. Even though produced by citizens, their data production is often supported by civil society organisations, governments, business partners, or community-based organisations. These actors unites their often necessary role in providing resources, support, and knowledge to citizens to produce data. In return they can harness data created by citizens to support their own mission. Thus, citizens and other actors often gain mutual benefit from a CGD project.

But if CGD projects rely on partnerships, who has to be engaged, and through which incentives, to enable CGD projects to achieve their goals? How are such multi-stakeholder projects organised, and which resources and expertise do partners bring into a project? What can other projects learn to support and benefit their own citizen-generated data initiatives?

This report calls for stronger collaborations to tap into CGD. Once put into practice these projects can become long-lasting, effective means to tackling issues affecting both citizens and other actors involved in the project. It presents overarching and context-specific factors ensuring that projects achieve their goals and make collaborations work. The report concludes with a brief discussion reflecting on some concerns of using citizen-generated data, as well as recommendations for citizens, community groups, civil society organisations, policy-makers, donors, and businesses. It is based on a broader scoping study commissioned by the DataShift and conducted by Open Knowledge International.

---

1 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

OVERARCHING FACTORS ENABLING CGD PROJECTS

WHICH BENEFITS WILL CITIZENS GET FROM THE PROJECT?

Each CGD project should have a clear answer to the question: 'What does an individual citizen get in exchange for contributing data?'. Several forms of benefits are imaginable. The most immediate is the perception that data helps addressing an issue that is relevant to citizens. Another benefit is the feeling of belonging to a community and the possibility to learn from others through data. For example, Patients Like Me relies on building a sense of community, allowing patients to compare information and experiences with other patients with similar, often rare, medical conditions. Data may also align with personal values and interests. Thus it is important to explain what a project wants to achieve and to demonstrate how citizens can be useful in achieving the goal. For instance, projects aiming to collect and analyse information on hydraulic fracturing (fracking) may attract citizens from various backgrounds and with different skills - if the project's framing makes information on fracking relevant to their living context. On a very practical level, there is the question of ease of use of technology employed to collect and analyse data. Is this technology already embedded in the daily routines of citizens and project partners? Or do these routines need to adapt in order to produce and use CGD?

GENERATING DATA - VOLUNTEERED OR PAID?

A particularly complex topic is the choice whether and how to remunerate citizen's participation within a project. This report found three choices on how to deal with money as a financial incentive for individuals.

There are different forms of remunerating citizens for data collection. Smaller stipends cover occurring costs during the data collection of volunteers.

Payments incentivise individuals to collect data which is otherwise of little concern for them. In some cases payments may jeopardise citizens' motivation to participate or create undesired dependencies. Payments may undermine a citizen's intrinsic motivation and identity of being trusted authorities - posing obstacles to create a culture of accountability in communities. Yet, the relationship between payments and intrinsic motivation is not clear-cut. Payments can be an initial incentive before being replaced by intrinsic motivations.

When people donate data, they feel they are contributing to a cause. Social values and feelings of community-belonging become important. This is also why some projects monetising data do not want to pay users. It 'changes the dynamics' of the community, such in the case of Patients Like Me. In some cases people invest in a CGD project, and care about it from the get-go. This implies that the people who are collecting data care about data collection. On the contrary, requiring citizens to invest will limit who is able to participate - inhibiting the coverage the data will have. Also only citizens from a certain sociodemographic may be represented potentially leaving out marginalised groups.

INVEST IN HUMAN CAPACITY

Investing in individuals' abilities to learn not only keeps them curious, engaged and inspired by the issue, but it also increases the human capacity needed for a project to become an institutionalised practice. This is important for all actors involved - from the citizens providing the data, and the organisations running the project, to those actors using the data and putting their information into practice.

Effective projects often don't need significant amounts of people but a small, dedicated, and sometimes well-trained team. Particularly in virtual projects where already existing data is classified, scraped and analysed, a small team can accomplish significant successes. Nevertheless, especially in more traditional survey-style CGD projects, investments in methodological training are necessary - both to ensure that surveys are conducted in a sound manner, but also to familiarise citizens with the tools to capture data. Training also facilitates dialogue and communication e.g. through workshops around the particular

shared interests and values, which can spur new ideas among communities and individual citizens.

Investments in human capacities are also needed in organisations participating in a CGD project. This is especially true for CGD seeking to remove information asymmetries between individuals, groups, and organisations,shall stimulate knowledge transfer, or generally be adopted and institutionalised by external organisations (including government bodies).

THE ROLE OF PARTNER ORGANISATIONS

Partner organisations are an important element between citizens producing data and the eventual uptake of data. Certain organisations do not steer data production directly, but facilitate interactions among citizens by providing technical infrastructure. Other organisations actively mobilise citizen to produce and use CGD as part of a larger strategy. These types of projects can serve either internal operations and policies in government and the private sector or as accountability strategies for civil society. The drivers to produce CGD are manifold, and citizens have varying degrees of control over the data they produce and the means to produce it.

CGD: ALWAYS CENTERED AROUND CITIZENS' ISSUES, BUT CITIZENS DO NOT NECESSARILY HAVE TO "OWN" THE ISSUE

By definition, citizen-generated data is actively created by citizens, often to directly monitor, demand or drive change on issues affecting them. Yet citizens' needs and issues do not necessarily have to overlap with needs of collaboration partners. Our report suggests that citizens use CGD to gain benefits around rather immediate issues (with the exception of community-based monitoring which seeks to make citizens adopt an identity of auditors, aware of their rights and vigilant to hold governments to account). Collaboration partners such as international non-government organisations or governments gain structural, long-term-oriented benefits using CGD for instance to improve policies to the benefit of citizens. Other partners such as commercial infrastructure providers gain a commercial benefit in providing technology.

The issues and data can be defined from the bottom-up by citizens. Citizens then seek to encourage other actors to take ownership of the problem. Alternatively, issues can be defined from the top-down by non-civic actors taking ownership of the problem. In this case CGD projects have to pay attention which transactions / trade-offs may activate citizens to contribute data. 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. Thus, effective projects are **issue-driven and not tech-centred**. Technology is a support for the value-driven problem that the CGD project provides or helps find solutions for.

STIMULATING SHARED INTERESTS

Shared interests and values are the foundation of effective partnerships. By finding key points of overlap in shared interests, actors recognise the value they get from a collaboration. Shared value may for instance stem from:

- Highlighting how citizen-generated data can **complement existing organisational processes and tasks**.

- **Framing multiple issues by using the same data.** For instance, an Indian CGD project associated its flood risk maps with the issue of drinking water scarcity. This inclusive framing caught the attention of many interest groups who saw water scarcity as an urgent issue, hence mobilising them to become active around an issue.

- **Communicating goals and benefits clearly:** Patients Like Me defined a policy to inform patients how their data is being treated and that it serves to advance medical research. Contributing patients not only get information about their conditions, but perceive it as an active long-term benefit to contribute to research.

USE A FLEXIBLE PROJECT ARCHITECTURE

In order for data to be produced over a long period of time, the project must become resilient and be able to adapt to changing contexts and incentives. This can be achieved through i) diversified revenue sources, ii) an open infrastructure, open data, and transparency of the project methodology, iii) data interoperability, both on the technical and social level.

DIVERSIFY REVENUE STREAMS

Our report suggests that most of the 14 observed CGD projects depend on external funding sources to sustain their operations. **External funding** describes any money allocated to a project through external parties - including grant funding through development agencies, foundations and governments as well as seed investment or allocation of funds through public-private partnerships. Other revenue models include mixed funding through sales of technology and infrastructure, as well as commissioned research, and data monetisation (a mixed model of free network/platform services and paid analytics). If undertaken ethically and responsibly, these have the potential to enable CGD projects to be more self-sustaining.

COMMIT TO OPENNESS AND TRANSPARENCY OF DATA PRODUCTION AND USE

An open communication of the applied methodologies, weaknesses of the data, and how the data can be used breeds trust from different actors, especially from citizens contributing their data. Some projects openly communicate that their data do not aim towards being comprehensive or representative. Instead, this data can indicate tendencies and trends, offer guidance for other organisations to run their own data collection for verification, or provide new contextual knowledge for government, business or academic research.

Examples are location-specific, text-based surveys foregrounding the reasons why citizens object to policies; or patchy information around large-scale land purchases. These practices foster trust which is key for effective collaboration. This applies to contexts of weak governance, opposition between government and civil society, or when private actors are involved. Openness and transparency also help the CGD project to remain neutral around politically or

personally sensitive topics, and to avoid unwanted associations with negative political topics.

Make data interoperable

CGD which is flexible enough in its data formats, and access remains relevant in the long term. Interoperability can both be read from a technical and social perspective. It allows data to be used for multiple purposes including mixing data with additional data elements. This is important if the data shall remain applicable to issues and concerns that change over time.

CONTEXTUAL DRIVERS MAKING CGD WORK

The context in which a CGD project operates defines additional factors making a project work. The graphic below describes four distinct approaches how CGD projects seek to tackle context-specific issues.

		GOVERNANCE	
		Strong	**Weak**
LINK TO GOVERNMENT	Direct	E-government projects, improvement of government processes: Institutionalising CGD projects into government processes by building on and restructuring existing methods	Citizen monitoring: Building government and community capacities to create new processes for CGD take-up
	Indirect	Knowledge platforms: Maximising economies of scale of intermediary CGD platforms	Intermediaries building data infrastructures: Projects are short lived and focused on specific issues but the project architecture is flexible and taps into existing media systems, and infrastructure

Graphic: Approaches to CGD project success in different contexts

STRONG GOVERNANCE AND 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.

STRONG GOVERNANCE AND INDIRECT LINK TO GOVERNMENT

Case studies falling within this category operate in a context with strong legal regulations and a tendency to have a significant amount of information available. Projects aim to fill in data gaps. The observed projects are platforms using economies of scale by facilitating interactions between citizens to collect CGD. 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 analyses of the issues the platform addresses. 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 monetise these platform models. The choice of whether to do so or not depends on the goals of the project.

WEAK GOVERNANCE AND DIRECT LINK TO GOVERNMENT

Issues in this context are a lack of information flows and organisational processes put into place to capture information. The goal is to improve government processes based on multi-partnership collaborations to strengthen governance. Projects rely on human capital such as through mobilising community members for monitoring public services, or in form of government recruitment for government-assisted data collection. This poses challenges for increasing the geographic coverage of these projects and fostering institutionalisation of their uptake into government processes: Because of the nature of monitoring local issues, projects are highly context sensitive, start locally as pilots to be replicated across regions, and therefore challenging to scale.

WEAK GOVERNANCE AND 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. The goal is to provide infrastructure that can connect project partners such as international NGOs and donors with citizens on the ground by building on resources available to citizens and project partners. Projects thereby answer immediate strategic needs or policy issues of partnering organisations, including the reasons for defective development programs. Indirect links to government may exist through government as a funding source, a project organisation not primarily involving government actors but open to collaboration, and data yielding value for actors both within and outside of the public sector.

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

As mentioned above some projects depend on CGD that is produced in a consistent way to monitor a phenomenon over time (in this case how public service performance changes). Yet, some projects embrace the non-standardised 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 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 marginalised 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 marginalised populations when this data is not accompanied by legal procedures ensuring their responsible use.

DATA OWNERSHIP

The legal rules for data storage and data ownership have implications for to what extent data intermediaries can monetise, 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

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:

1. **Align interests among key stakeholders to encourage partnerships.** Successful CGD projects bring together actors with different interests in the same data. Data serve as common ground for actors and is the focal point of collaborations. There is often a difference between production, use, and uptake, and the benefits associated with each stage can be different. Different actors can value different aspects of the data; understanding how actors perceive this value is key to building multi-stakeholder partnerships.

2. **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.

3. **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.

4. **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.

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

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.