FROM
EVIDENCE
TO ACTION

TURNING CITIZEN-GENERATED DATA INTO ACTIONABLE INFORMATION TO IMPROVE DECISION-MAKING

Danny Lämmerhirt, Shazade Jameson, Eko Prasetyo

DATASHIFT

OPEN KNOWLEDGE INTERNATIONAL

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.

FROM
EVIDENCE
TO ACTION

TURNING CITIZEN-GENERATED DATA INTO ACTIONABLE INFORMATION TO IMPROVE DECISION-MAKING

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

FOREWORD

Citizens and their organisations create data as a direct reflection of their issues. This citizen-generated data (CGD) yields the potential to tell counter-narratives of sustainable development, and truly make all voices count. Yet, critics argue that CGD lacks rigorousness and is not representative; as a result, the recurring question is what are the factors that can help or hinder the usability of CGD, increase its uptake and help drive the action that it aims to stimulate.

There are different paradigms of using data for monitoring or using data for action, different perceptions around similar topics, and different data quality requirements at different scale levels. For instance, monitoring is only one aspect in the larger cycle of human decision-making, including agenda setting, designing, and implementing solutions. These actions are equally important to drive sustainability as monitoring, but these issues get little attention in current debates. Therefore, it is important to understand which forms of action can be informed by CGD, and when and how monitoring of the higher level indicators such as the SDGs can be useful.

In order to properly zoom in on and untangle these differences, we present two research reports that work together in tandem. This piece, 'From Evidence to Action', focuses on factors that help make CGD relevant and actionable. It emphasises that in order for data to inform decisions around sustainable development, the data must be catered to different stakeholders in different forms with different aspects of data quality. The tandem piece, 'Acting Locally, Monitoring Globally?', explains how CGD can help to monitor the SDGs, discussing the challenges and opportunities that arise as the data moves from being used for action at the local level to being used for monitoring at a higher-scale level.

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.[1]

———

1   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?

EXECUTIVE SUMMARY

This report demonstrates how citizen-generated data (in the following: CGD) can support decision-making and trigger action. CGD is a representation of the issues that are most important to citizens. If evidence provided by CGD shall trigger action, the issue and its stakeholders need to be well understood. Stakeholders have different priorities, values, or responsibilities, and are affected differently by an issue. Stakeholders have certain capacities to engage with an issue, and are prepared differently to act upon it. Some actors may lack the literacy, knowledge, time, or interest to engage with complicated data. The task is for CGD projects to understand these nuances, and to translate their data into **digestible**, **easily understandable**, and **relevant messages**.

The **qualities of CGD need to match with the action that is planned**. Long-term monitoring needs reliable, accurate, and standardised data. Setting the agenda for a formerly unknown issue may require a CGD project to build trust, and to ensure credibility. Some projects might need to produce highly detailed data, other tasks only require rough indications of trends. **The engagement strategies should fit with the desired change, too.** To change policies, perceptions, or behaviour, a targeted engagement strategy should be used. Such a strategy includes various forms of engagement, from data portals, over public hearings, to community work.

In detail, CGD can inform four distinct types of action:

■ **Agenda setting:** Did an issue receive attention before the CGD project started? Agenda setting raises awareness for a problem. It is about altering the perceptions of stakeholders and to mobilise them.

■ **Designing solutions:** How could an issue be solved? CGD can be used to envision or plan alternative ways of managing an issue.

■ **Implementing solutions:** CGD can also directly steer behaviour and enable better actions by giving stakeholders relevant information to enable actions. CGD can also steer behaviour by helping taking decisions, or rewarding certain actions as performance indicators do. A caveat is that CGD will be 'gamed'. Thus, every effort to design CGD that steers behaviour must be carefully thought through.

■ **Monitoring and evaluating solutions:** CGD can also inform performance monitoring of all kinds—from process efficiency to satisfaction with service outcomes. Monitoring is based on pre-set criteria, compares performance against goals, and involves judgement. This stage serves to reflect upon solutions and can be supported by in-depth contextual information.

RECOMMENDATIONS

On the basis of our case studies, we suggest that CGD projects can better influence decision-making by assessing:

- **The audiences** they want to reach: Different audiences have different interests in the CGD project and can perform different actions to solve the issue. Which level of government is responsible for the issue? Who are the stakeholders that can be mobilised?

- **The power and interest** stakeholders have in an issue. Power can be understood in many ways, such as the power to legislate or manage an issue, or the power of building confidence within communities to engage with decision-making processes. Depending on power and interest, different engagement strategies should be applied.

- **The message** data should convey to these audiences. What is the relevant data that is needed to engage with the stakeholders being targeted. Issues should be framed so that they resonate with the knowledge, perceptions, and lived realities of stakeholders. Different engagement strategies are important to ensure that the data are listened to.

- **The engagement strategy** to connect with different audiences. CGD projects should design outreach and engagement strategies that are relevant and suitable for the context. Furthermore, targeted engagement is most likely to change behaviour and drive action.

Good quality data must be understood holistically, as its validity and usefulness will vary according to the issue and the stakeholders invested in it. This requires a thorough, integrated project design and a careful methodology. We recommend that CGD projects consider the following methodological issues during data production and processing:

- **Validity and reliability** are generally important for CGD projects. Only **accurate data can be credibly used to make claims about an issue**, to aggregate or compare data or to calculate trends and correlations.

- Yet, **data quality is** largely **determined by the intended use**. CGD projects should think about how 'complete' and timely data has to be, in order to become useful for a task. It should be asked: how is the accuracy of my data affected, if some data is not included in a dataset? Timeliness must not be confused with 'real-time data': instead data is timely if it is provided in appropriate and useful rhythms.

- **Data aggregation** relies on categorisation and standardising data. Both operations can be done during the data capture phase (in the form of standardised data capture methods) or by cleaning and classifying data afterwards. A **major challenge is to define common categories that are meaningful and relevant** for data producers (those who want to describe the issue) as well as for data users (those who need to understand the issue).

- **Data visualisations** help communicate information and patterns in complex data but demand data literacy and graphicacy. Often, readers also need topical knowledge to interpret the sometimes complex underlying information of data visualisations.

The usefulness and relevance of CGD can be leveraged by:

- **Designing targeted engagement strategies:** Research around evidence-based politics highlights partnerships as an important means to transfer knowledge, establish trust, and make key messages graspable. Targeted engagement strategies do not end with publishing CGD reports or visualising data online. Instead, the engagement methods need to be suitable for individual stakeholders. Examples are public hearings, education meetings with local decision-makers, on-site visits with decision makers, hackathons, or others.

- **Choosing the right degree of participation for stakeholders throughout the project:** Successful projects manage whom they engage in different phases of the project. The degree of participation is a crucial element of each CGD project. For instance, should citizens or policy-makers be engaged in the definition of data? How does this affect the credibility of data and buy-in? Who should be engaged in the dissemination of findings. Does the project benefit to collaborate with a 'knowledge broker' like an experienced advocacy group, a university, or a newspaper?

- **Acting like a 'knowledge broker', crafting targeted messages:** Data should be translated in a way that is understandable and relevant to stakeholders. Long, detailed reports might interest researchers, while 'killer charts', and concise information might appeal to busy decision-makers.

- **Granting open access to raw data:** Several CGD projects grant access to their data as long as these do not contain personal information. Is my audience a group of researchers, a journalist unit, or some other knowledge broker who can translate and analyse the data? Open access helps gathering expertise from outside and increases the relevance of raw data.

- **Explaining raw data:** Raw data is often produced in a messy process. Data values can be incomprehensible for both, humans and machines. Metadata and other documentation can help to understand what the data means, how it was created, as well as methodological strengths and weaknesses of the data.

INTRODUCTION

Human decision-making is complex. Our daily routines, perceptions, values, and lived realities all influence how we make choices. Data is seen as a panacea to inform better decision-making, be it to tackle corrupt and value-laden policy processes with evidence, or to remove opinionated bias from (individual) decisions. Yet, there is no straight-forward answer as to how data turns into actionable, relevant evidence that informs decision-making and behavioural change.[2] The term 'evidence' is commonly associated with neutrality and objective facts. However, the data underlying evidence is often a matter of concern, rather than a matter of fact. Especially in policy contexts ideologies, political programs, values, and beliefs influence which evidence is used for which decisions. Research states that evidence-informed policy-making is a "power-infused, non-linear process"[3]. What counts as actionable and high-quality evidence lies in the eyes of the beholder.[4]

Citizen-generated data (in the following: CGD) is increasingly used to provide evidence for decision-making. Examples repeatedly show that CGD can change how issues are perceived, interest groups are mobilised, policies are designed, and issues are tackled.[5] CGD is a means to voice the concerns of individual citizens or civil society at large. It flags all types of issues–ranging from environmental damages, to labour conditions, or perceived corruption. But democratising the means of data production is not faced without resistance. Often data generated by citizens is refuted as not being statistically sound, as lacking representativity and accuracy, or as not meeting other features of 'good quality data'.[6] Data is never raw, but always 'cooked' and born out of accepted routines to produce data.

---

2   There is a significant overlap between what is considered 'good data' and what is considered 'good evidence'. For a detailed discussion, see: Sutcliffe, S., Court, J. (2005): Evidence-based Policymaking: What is it? How does it work? What relevance for developing countries? Available at: **https://www.odi.org/publications/2804-evidence-based-policymaking-work-relevance-developing-countries** as well as Wang, R. Y. and Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, 12 (4), 5-33. Available at: **http://courses.washington.edu/geog482/resource/14_Beyond_Accuracy.pdf**.

3   See also: Shucksmith, M. (2016): InterAction. How Can Academics And The Third Sector Work Together To Influence Policy And Practice? Available at: **http://www.carnegieuktrust.org.uk/carnegieuktrust/wp-content/uploads/sites/64/2016/04/LOW-RES-2578-Carnegie-Interaction.pdf**

4   See also: Poel, M. et al. (2015): Data for Policy: A study of big data and other innovative data-driven approaches for evidence-informed policymaking. Report about the State-of-the-art. Available at: **http://media.wix.com/ugd/c04ef4_20afdcc09aa14df38fb646a33e624b75.pdf**

5   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**

6   See for example Lämmerhirt, D. Jameson, S., Prasetyo (2016): Making Citizen-Generated Data Work. Towards a framework strengthening collaborations between citizens, civil society organisations, and others. See also Piovesan, F. (forthcoming): Statistical Perspectives on Citizen-Generated Data.

If CGD shall voice the concerns of citizens, it is necessary to understand under which circumstances it becomes a trusted source of information, used in consensus, relevant, and fit-for-use.[7] Each project working with CGD should consider two elements relevant to assess when its data is 'good enough' for action: a human element and the data element.

The human element

Using data for decision-making and other actions ultimately remains a human issue. Former research has discussed data's role as evidence to influence behaviour and policy processes.[8] Actors involved in policy-making seem to prefer 'hard' evidence (including quantitative data collected by researchers and government agencies) over 'soft' evidence (including data such as narrative texts, written reports, personal perceptions, or autobiographical material).[9] Research criticises that soft evidence is often neglected, in favour for numbers which become a main argumentative device. A fixation to numbers could lower the quality of policy-making which is why personal, qualitative stories (including from marginalized groups) should be more often considered in policy decisions.[10]

The data element

Data quality is not only a statistical issue. Beyond representativity and accuracy, data quality may be regarded as whether it is 'fit-for-purpose'.[11] Whether data is 'fit-for-use', depends on the users and the tasks at hand. Does the data contain a sufficient amount of information? How often, and how quickly, should data be available in order to count as relevant and actionable? In which form should the data be presented to be understood by data users?

This report argues that CGD projects need to understand the issue and the stakeholders invested in an issue, in order to collect relevant data, and to communicate it effectively. The report acknowledges the myriad ways how data informs decision-making and action, and starts off stating that there is no general recipe to create good data. Instead the report wants to spark the imagination of citizens, civil society groups, policy-makers, donors and others on how they may wish to employ CGD for fostering sustainable development.

---

7   See also: Lagoze, C. (2014): Big Data, data integrity, and the fracturing of the control zone. Available at: **http://thirdworld.nl/big-data-data-integrity-and-the-fracturing-of-the-control-zone**

8   These studies define evidence as systematically gathered knowledge which can be presented in any form–be it as quantitative data, or qualitative narrative texts. See also: Sutcliffe, S., Court, J. (2005): Evidence-based Policymaking: What is it? How does it work? What relevance for developing countries? Available at: **https://www.odi.org/publications/2804-evidence-based-policymaking-work-relevance-developing-countries**

9   There are several observations how data and other information provide evidence and inform decisions.

10 Evidence is less important to legitimize political or legal. CSOs mainly struggle to use evidence in order to claim expertise, knowledge, information or competence that justifies its actions and its influence on authoritative decisions.

11 See also: Shucksmith, M. (2016): Interaction. How can academics and the third sector work together to influence policy and practice? Available at: **http://www.carnegieuktrust.org.uk/carnegieuktrust/wp-content/uploads/sites/64/2016/04/LOW-RES-2578-Carnegie-Interaction.pdf**

The report addresses the following research questions:

- What qualities does data need to have in order to be relevant and actionable for stakeholders to drive sustainable development?

- Which engagement strategies are applied to involve stakeholders in using data? Which other factors enable these actions?

To do so, the report discusses three elements to understand good quality data: i) the stakeholders and potential users of CGD, ii) the different criteria of data quality, iii) the different communication methods turning data into actionable evidence. After describing these elements, the report sheds light onto different forms of action that result from using data. Thereby the report makes the point that it is necessary to reimagine what counts as 'good data': data that is not only statistically representative, valid, and reliable, but that is able to address an issue and become meaningful for stakeholders.

UNDERSTANDING STAKEHOLDERS

As highlighted throughout another report by the authors, CGD needs to accommodate concerns among different stakeholders.[12] This report understands stakeholders as individuals, organisations, groups, associations, or networks who are likely to affect or be affected by the implementation of CGD projects. Each stakeholder may perceive and value information differently, necessitating a user-centric design for CGD. Such a design puts the issue, the intended message, and its stakeholders before the data.[13] Hence CGD projects should identify stakeholders early. A stakeholder mapping helps to understand who is potentially affected by an issue, what their interest in the issue is, and which power the stakeholder yields to tackle the issue. These questions also affect which data will be relevant for which actor. Depending on the level of power and interest, each stakeholder requires different engagement strategies.[14]

Power is a complex idea. Relating to the context of CGD, it may be described as the degree of influence stakeholders will have on the CGD projects. This report proposes a more nuanced model of power based on empirically observed cases[15] and inspired by Robert Chamber's model of citizen empowerment[16]. For instance, governments and administrative bodies can have power over a phenomenon. This power can be very nuanced and be defined by sovereignties: For instance, national government can be responsible for allocating money into water, sanitation, and hygiene (WASH) infrastructure, while the maintenance of pipes, wells, boreholes, and other

---

12 Lämmerhirt, D. Jameson, S., Prasetyo (2016): Making Citizen-Generated Data Work. Towards a framework strengthening collaborations between citizens, civil society organisations, and others.

13 Wand, Y., and Wang, R. Y. (1996) Anchoring Data Quality Dimensions in Ontological Foundations. Communications of the ACM, 39 (11), 86-95. Available at: http://www.thecre.com/pdf/MIT-wandwang.pdf. Wang, R. Y. and Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, 12 (4), 5-33.
Available at: http://courses.washington.edu/geog482/resource/14_Beyond_Accuracy.pdf.

14 The Overseas Development Institute, (2009). Planning Toolkit: Stakeholder Analysis.
Available at: https://www.odi.org/publications/5257-stakeholder-analysis

15 See 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?; as well as Lämmerhirt, D., Jameson, S. and Prasetyo, E. (2016). Making Citizen-Generated Data Work. Towards a framework strengthening collaborations between citizens, civil society organisations, and others.
Available at: http://civicus.org/thedatashift/wp-content/uploads/2015/07/Making-citizen-generated-data-work.pdf

16 Chambers, R. (2012) Provocations for Development, Practical Action.

infrastructure is the duty of local government. In this case community groups need to understand which data is useful for which government body, in which process of the water management. Community groups can have the power to engage with politics, for instance through laws enshrining demonstrations, or public consultations as accepted means for citizens to engage with government actors. 'Power with' means the power to mobilise collective action across organisations, individuals and networks. 'Power within' is the self-confidence to do something. This is often a side-effect of community monitoring strategies where communities feel empowered by gaining knowledge about how to hold governments to account. Who holds power over what depends on the governance context.[17]

Using the case study of Humanitarian OpenStreetMap in Jakarta (HOT), a simple method for stakeholder analysis is presented in figure 1 as a power-interest-matrix.

Figure 1: Example of a power-interest matrix (Humanitarian OpenStreetMap)

HIGH

Media
UN agencies

Indonesian National Disaster Management Agency (BNBP)

Australia Department of Foreign Affairs and Trade (DFAT)

The World Bank

Parner universities

Local Communities

Other universities
Other CSOs

Partner CSOs
Online volunteers

POWER

LOW | **HIGH**

INTEREST

17 See also: Lämmerhirt, D., Jameson, S. and Prasetyo, E. (2016). Making Citizen-Generated Data Work. Towards a framework strengthening collaborations between citizens, civil society organisations, and others. Available at: http://civicus.org/thedatashift/wp-content/uploads/2015/07/Making-citizen-generated-data-work.pdf

Stakeholders with high-power and interests aligned with CGD projects are important, they need to be fully engaged and be brought on board. The HOT team perceived government, donors, local communities, and partner universities as stakeholders who have both high-interest and high-power (as evidenced, among others, by a bilateral agreement between the governments of Australia and Indonesia to develop methods of disaster risk reduction). The team engaged with highly relevant actors through trainings (of government officials), mapathons in communities, and collaborations with partner universities.[18]

Stakeholders with high-interest but low-power need to be informed and mobilised. They may become an interest group or coalition which can contribute toward change. CSOs and online volunteers are stakeholders with high-interest (for instance in improvements of public services) but with lower-power. On-field volunteers are mobilised, for example, to map territory in the aftermath of the Aceh earthquake.[19] Stakeholders with high-power but low-interest should be brought around as patrons or supporters. These stakeholders may be critics who reject CGD for lack of credibility or other quality issues (see Section Rethinking What Counts As 'Good' Data). Whilst not working directly with UN agencies, HOT collaborate with them for knowledge sharing events. And lastly, stakeholders with low-power and low-interest need to be monitored but with minimum effort.

ENGAGING STAKEHOLDERS & THE LIFECYCLE OF CITIZEN-GENERATED DATA

CGD projects use a data value chain. The following graphic shows such a value chain. It is inspired by the idea that data is the raw material for actionable information (which is data put into context and a pre-existing stock of knowledge).

Graphic 1: Data value chain

Issue definition → Data definition → Developing data capture methods → Data collection phase → Analysis → Product of information → Action

---

18 HOT selected 4 universities out of 13 (in 2013) for the current project implementation based on the commitments and outcomes of previous project phase.

19 See more at: https://hotosm.org/updates/2016-12-13_hot_indonesia_launches_a_tasking_manager_and_pidie_mapathon_in_response_to_aceh

Each CGD project can have different degrees of participation and variances in the way it assigns tasks to stakeholders along the data value chain. By definition each CGD project engages citizens in the data collection phase. Some projects also involve citizens or decision makers in the data design phase to increase buy-in and legitimacy among those groups. The stakeholder mapping may inform the degree of participation during the data value chain. Lead questions could be: Is it necessary to engage citizens in the beginning of a project? How does this influence the acceptance of my project for citizens, and its credibility for government? How does it shift power within a project to engage with several actors and how will the data design be affected. Should I seek advice from government when defining my data capture methods? Which audiences should be addressed with which information? The choices of whom to engage with, how, and at what stage of the CGD project, all affect how the quality of data is perceived (see Section Rethinking What Counts As 'Good' Data).

KEY TAKE-AWAYS:

- **The audiences** they want to reach: Different audiences have different interests in the CGD project and can perform different actions to solve the issue. Which level of government is responsible for the issue? Who are the stakeholders that can be mobilised?

- **The power and interest** stakeholders have in an issue. Power can be understood in many ways, such as the power to legislate or manage an issue, or the power of building confidence within communities to engage with decision-making processes. Depending on power and interest, different engagement strategies should be applied.

- It is important to **understand the data value chain of CGD**, and which stakeholder may be engaged throughout producing data. Each stage has its own methodological pitfalls and opportunities to team up with others. Whether and how to partner with an organisation depends on several questions (see point below).

- **Choosing the right degree of participation for stakeholders throughout the project:** Successful projects manage whom they engage in different phases of the project. The degree of participation is a crucial element of each CGD project. For instance, should citizens or policy-makers be engaged in the definition of data? How does this affect the credibility of data and buy-in? Who should be engaged in the dissemination of findings. Does the project benefit to collaborate with a 'knowledge broker' like an experienced advocacy group, a university, or a newspaper?

RETHINKING WHAT COUNTS AS 'GOOD' DATA

Once the relevance of particular CGD has been understood for various actors, the next question is what qualities does data need to have in order to be actionable for them? There are myriads of definitions for data quality.[20] In quantitative social sciences data quality is understood as objective and accurate (reliable and valid). It is acknowledged that good data should always be i) accurate and reliable, ii) objective and non-biased.[21] But data quality also has to match the task at hand and can be defined from a user perspective. Table 1 shows seven dimensions of data quality. These dimensions are derived from Louise Shaxson's work on robust evidence for policy-making. Even though focussing on the policy context, we see her framework as a useful entry point to understand the robustness and quality of information for broader use cases. The list is complemented by empirical observations taken from other reports written by the authors.[22]

Table 1: Dimensions of data quality

EXPLANATION	QUESTIONS TO ASK YOURSELF
CREDIBILITY	
Credible evidence relies on a strong and clear line of argument, tried and tested analytical methods, analytical rigour throughout the processes of data collection and analysis, and on clear presentation of the conclusions.	Would others see the same issues in my data? What reputation do I have? Does it affect the credibility of the results, and for whom? Do my methods to capture and analyse the data limit my findings? Does the information I present make sense to the people I engage with? How to improve my credibility by engaging stakeholders during data definition, capture, and analysis?

20 For an overview of data quality, we propose to read Borgman (2011), Wand and Wang (1998) as well as Shaxson (2005).

21 Some projects like Africa's Voices embrace the biases of subjective data because they want to foreground specific perceptions of citizens in very specific contexts. Africa's Voices for example seeks to read social norms in 'biased' responses dos sometimes controversial topics.

22 These reports are available at: http://civicus.org/thedatashift/learning-zone/research/

EXPLANATION	QUESTIONS TO ASK YOURSELF
ACCURACY	
Accuracy describes whether a project is measuring what it actually intends to measure.	Do we come to similar findings when replicating our study? If not, why?
	Does the data form a sound basis for monitoring, or similar purposes for which repeated data collection is necessary?
COMPLETENESS	
Completeness does not mean that data is all-encompassing.	How much detail do I need to gather my evidence?
	How much detail and coverage do stakeholders need to act upon my information?
Completeness is better understood as data that contains enough information to become meaningful for a task.	Do I need to aggregate my data (for instance from individual perceptions of health services to numbers of dissatisfied people)?
	How much disaggregation is needed? Is it hard to access very detailed data? Which level of detail is harmful for individuals or violates their privacy?
	Do I limit my accuracy through the data sample? Do I need to capture more information to present and understand my issue more accurately?
	In which time intervals do I have to provide data? Does it suffice to measure data over a short period? Or do I need to observe an issue over a longer time span?
OBJECTIVITY	
The information CSOs collect are bound by the assumptions that are made during data capture and analysis. Objective data is data that seeks to eliminate subjective bias as much as possible.	Does my data collection allow for bias through subjective assessments? For instance when running surveys, are my participants invited to give their opinion?
	Do I value subjective assessments as part of my evidence? Or does it distort my findings?
	Which methods should I apply to reduce every possible bias?
	How can I understand and communicate the bias in my data?
TECHNICAL ACCESSIBILITY	
The data needs to be accessible in formats that make the information it contains usable.	Can I stimulate uptake of my data when making it openly accessible to other audiences (without limitations in re-use)?
	Which pieces of information do I have to remove from my data before making it publicly available? Could my data do harm to someone (e.g. violating privacy rights) by publishing data openly?
	Do I gain credibility when making data and the collection methodology accessible?

EXPLANATION	QUESTIONS TO ASK YOURSELF

UNDERSTANDABILITY

Data is understandable when humans and machines can readily make sense of it. Understandability is needed not only to convey messages to audiences, but is also necessary to make data comparable to each other (i.e. interoperable)

How do I need to document my data so that others can understand and use it?

What is the most appropriate format to convey key messages?

Do I need metadata, narrative text or visualisations to make my data understandable?

Do I need to adhere to data standards so that my data can be integrated into other datasets ?

GENERALISABILITY

Generalisability implies that the findings of a CGD project can be applied to other contexts.

Can I take the information and evidence I created and apply it to another context or another location?

How can I scale the findings of my project across other settings?

Is my evidence for an issue context specific because of my data sample (biased by a set of specific people, limited to some regions)? Because my questions foreground very specific facts?

What is the nature of the issue I want to describe? Which bits of context make my data less general? Why?

PRODUCING RELEVANT DATA

The following section offers some examples how to cater data to different audiences. A commonly presented critique states that CGD is not representative, only partial (low-coverage), or not comparable over time (inaccurate). The section will demonstrate that the value of CGD is more nuanced–and that often data representativity, comparability, or completeness depend on the use case.

WHEN IS DATA COMPLETE ENOUGH? SCALING THE DETAILS

Which level of detail data should have (how complete they should be) depends on the audience, and how it should be engaged to act upon a problem. The level of detail is known as level of aggregation. An example of highly aggregated data is the number of inhabitants in a country. Disaggregated (more detailed) data would be, for instance, how many women and men there are within this population, or how many live in a specific rural area. Often CGD affords the physical presence of citizens who collect data on the spot, thereby enabling the collection of detailed data.

A common question is how to scale this data across regions.[23] However, the first question should be why data should be scaled in the first place? Which information is gained, which is lost? Who can use this information to do what? Governments have 'power over' a territory through legislation, or public service provision. Depending on their sovereignty and responsibilities the CGD projects should interact with them using different types of information.

RETHINKING DATA COMPLETENESS: THE EXAMPLE OF VULNERABILITY MAPPING

As discussed above, complete data needs to offer sufficient information to become relevant for a task. Environmental risk and vulnerability mapping measures the risk of a hazard such as flooding. In this example the most common practice is to measure elevation and where water will flow, which can produce better flood models. However, measurement of hazards does neither capture socioeconomic vulnerability to the floods, nor how those vulnerabilities are distributed across regions. It is often the poor who live on floodplains. Not only are they in the path of the hazard, but they have weaker shelter and fewer economic resources to deal with the aftermath of the disaster.[24] If data should represent the marginalised in this case and inform strategies to alleviate their exposure to hazards, integrated vulnerability mapping is required. This is possible with using a base map (which may be provided coming from a cadastral office) and overlaying multiple layers of data showing different forms of vulnerability–such as poverty levels or housing conditions.[25] In this sense, CGD can significantly contribute to the complexity and granularity of data sets, pointing to blank spots that would otherwise be missing on maps.

THE TRADE-OFF BETWEEN DETAIL AND GENERAL INFORMATION

The patterns detected in vulnerability maps may not always be comparable or generalisable across regions. Socio-economic factors such as poverty, housing conditions and others, may only apply to a local context, may be due to specific historical factors, or (missing) governance mechanisms. The value of such maps may lie in laying foundations for location-specific interventions, such as regional policies to invest in these regions. If you want to map the underrepresented it may mean that your data (an integrated flood map for example) are by default not comparable. Yet, if repurposed, the data can become generalisable. As the next point shows making data generalisable has its very own purposes.
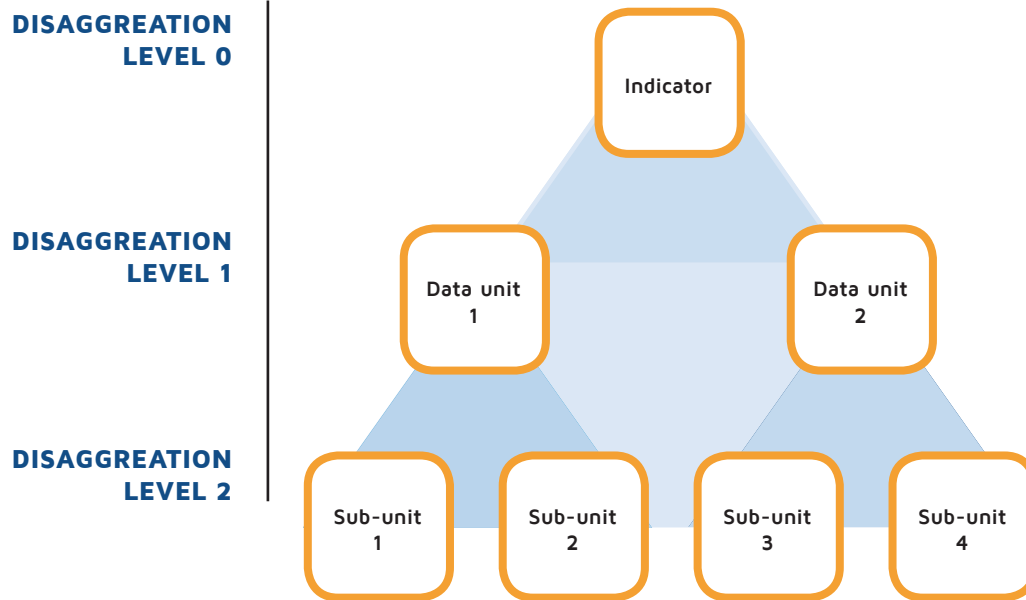
23 See Piovesan (forthcoming): Statistical Perspectives on Citizen-Generated Data.

24 Sara, L. M., Jameson, S., Pfeffer, K., & Baud, I. (2016). Risk perception: The social construction of spatial knowledge around climate change-related scenarios in Lima. Habitat International, 54, 136-149.

25 Baud, I. S., Pfeffer, K., Sridharan, N., & Nainan, N. (2009). Matching deprivation mapping to urban governance in three Indian mega-cities. Habitat International, 33(4), 365-377.

To think through the level of detail data should have, we propose a data aggregation model (figure 2). The model shows a pyramid of information. The bottom shows the most detailed data (sub-unit 1 and 2). By combining this information, CGD projects can have a more general information (data unit 1). More general information can be combined into indicators, for instance for national level monitoring.

Figure 2: Data aggregation model

DISAGGREATION
LEVEL 0

DISAGGREATION
LEVEL 1

DISAGGREATION
LEVEL 2

Indicator

Data unit 1

Data unit 2

Sub-unit 1

Sub-unit 2

Sub-unit 3

Sub-unit 4

A working example: A CGD project wants to understand if a non-formal settlement has enough public water and sanitation facilities. It can map different sanitary facilities like toilets or boreholes per region (Sub-Units 1 and 2) and create a total number of facilities (Data Unit 1). The project can furthermore count the number of people with and without access to these facilities in a given region (Sub-Units 3 and 4) and calculate a total number (Data Unit 2). Dividing the amount of persons with access by the number of accessible facilities can show whether there are enough toilets provided in a region (Indicator). The pyramid can be expanded at the top. For instance, several indicators can be combined to a 'composite indicator'. Prominent examples are the Gini index, the World Happiness Index or indices such as the Press Freedom Index. All of them are based on individual numeric indicators, whose importance is weighted against each other through a scoring that is assigned to each.[26]

---

26 The Organisation for Economic Co-Operation and Development (OECD) provides a useful introduction how to create composite indicators. See also: OECD (2008): Handbook on Constructing Composite Indicators. Available at: **http://www.oecd.org/std/leading-indicators/42495745.pdf**

Other CGD can foreground why some people do not have access to facilities. Is it because facilities are broken, non-existent, or because the travel distance is too long? Regional indicators of accessible sanitary infrastructure per person can be used to compare the provision with toilets and other services across regions or to understand the rough patterns, where provision is especially bad. **Indicators therefore often provide a bird's eye view on issues to see the big picture.** This can be useful if a nation state is responsible to allocate budgets to regions and to develop their infrastructure. Using a comprehensive indicator offers a bird's eye view on the national territory and to inform budget allocation. **More detailed information provides perspectives on the ground.** Why is access in some regions worse than in others? This information can be useful to understand an issue on the ground and to enable local government to improve governance to better maintain services.[27] Once again, which level of detail is needed depends on the actions that are needed, and the responsibilities, interest and power of stakeholders to tackle an issue. For a further discussion on the usefulness of indicators see also the Section 'For Whom Is An Indicator Useful' in our paper 'Acting Locally, Monitoring Globally?'. Ultimately, the data aggregation pyramid enables us to understand how to make qualitative data comparable. For instance, an initiative can code unstructured, qualitative data such as personal perceptions of violence into categories such as 'physical violence' or 'psychological violence'. Thus individual experiences are rendered equal, and become calculable, at the expense of evening out the nuances between individual experiences.

METHODS TO COLLECT DETAILED OR GENERAL DATA

The production of comparable information can be supported in the following ways: by collecting data according to agreed upon conventions, by standardising data during production or analysis, as well as through documentation of what the data means (metadata).

DATA CONVENTIONS
Data conventions and other agreed upon methods to collect, document and analyse data can reinforce credibility and acceptance. Data conventions refer to the data items collected, as well as to the methods to produce them. For instance, the organisation Twaweza collaborated with National Statistics Offices to collect data that reflect the educational curriculum and measures the quality of education according to standards relevant for government. The Louisiana Bucket Brigade consulted the Environmental Protection Agency (EPA) of the United States who deemed the project's air pollution sampling techniques as reliable.

———

27 This is exemplified by the work of the Social Justice Coalition and Ndifuna Ukwazi to monitor janitor services in Cape Town's slums.

METADATA

Metadata is useful to develop a shared language between CGD projects and audiences. Metadata, that is data about data, makes it easier to retrieve, use, or manage information.[28] Metadata can contain descriptions and keywords to interpret the data, including the topic the data describes, last update of the data, frequency of the updates, the capture method, explanations of values and others.[29] This is also important if a CGD project wants to mix its data with other data or wants others to reuse the data.

An example: If a country would like to understand the stability of an existing energy grid it could start by counting the incidents of electricity outage. Different CGD initiatives collect data about electricity issues and name it unclearly without describing what each data means (one project may collect its data in a spreadsheet, naming the data column 'issue: electricity', another may call the issue 'electricity shortage'). If someone would like to use this data it becomes practically impossible to add up the number of incidences without manually checking each report. Therefore metadata can help to describe what data named like 'electricity shortage' exactly means to make it comparable. Thus, metadata reduces ambiguity and makes others understand what data wants to represent.

TRIANGULATION

One method to increase the accuracy and reliability of the data is triangulation, which is using multiple information sources to verify data describing a similar phenomenon. Often used in areas like intelligence or the estimation of casualties in conflicts triangulation is also applicable to CGD.[30] For example, the Living Lots NYC project seeks to understand whether publicly owned lots are currently used. The problem: cadastral datasets of New York City are openly available, but they provide partial information on tenure and land use. The project therefore combined data on public tenure with data of existing community gardens published by the organisation GrowNYC as well as data about recent ownership transfers to see whether land was still city-owned.[31] Using geo-locations as common denominators enabled the project to overlap several map layers and identify already existing community gardens that were falsely classified as 'vacant' by the City.

28 National Information Standards Organization (2004): Understanding Metadata.
Available at: http://www.niso.org/publications/press/UnderstandingMetadata.pdf (accessed: 12 December 2016)
29 See also: World Bank (n. y.): Education Statistics. Available at: http://data.worldbank.org/data-catalog/ed-stats
30 The Human Rights Data Analysis Group uses a method called 'multiple systems estimation' to estimate casualties.
This method uses several available lists indicating the names of casualties.
The differences between these lists allow to infer the number of casualties.
See also: https://hrdag.org/2016/10/30/using-mse-to-estimate-unobserved-events/.
31 These plans indicate how the City intends to re-use publicly owned lots.

DATA AGGREGATION AND PRIVACY

The 'Million Dollar Blocks' project conducted at Columbia University, mapped the provenance of inmates in order to identify whether incarcerated people came from distinct neighbourhoods. To protect the identities of inmates the raw data of each inmate's address needed to be aggregated at block level, before they could be used and published to a broader audience.[32] It is important to note that with the rise of big data practices, aggregation can also lead to new challenges for privacy at a group level.[33]

KEY TAKE-AWAYS:

■ **Validity and reliability** are generally important for CGD projects. Only **accurate data can be credibly used to make claims about an issue**, to aggregate or compare data or to calculate trends and correlations.

■ Yet, **data quality is** largely **determined by the intended use**. CGD projects should think about how 'complete' and timely data has to be, in order to become useful for a task. It should be asked: how is the accuracy of my data affected, if some data is not included in a dataset? Timeliness must not be confused with 'real-time data': instead data is timely if it is provided in appropriate and useful rhythms.

■ A **major challenge is to define common categories that are meaningful and relevant** for data producers (those who want to describe the issue) as well as for data users (those who need to understand the issue). For example, citizens can produce data about local environmental damages containing location specific information useful for local decision-making. This data can be grouped in categories, be plotted on a regional map, and guide large-scale investments in environmental risk reduction strategies. Both types of information have different use cases for different purposes.

■ CGD projects can use **data standards and conventions** to improve reliability

■ **CGD does not have to abide by all quality criteria.** Often some qualities are more important than others. For instance, if the data is not fully reliable and shall be used for a first investigation, credibility gains importance.

■ **Comparing multiple data sources (triangulation) is a commonly used practice to verify CGD, or to increase accuracy of data.**

■ **Using metadata and standardised data structures increases data interoperability.**

---

32 Kurgan, Laura (2013): Close Up At A Distance. Mapping, Technology, And Politics. MIT, Cambridge.

33 In big data, algorithmic groups can be created during processing which do not necessarily have a connection to 'natural' groups (e.g. ethnicity), which can then be discriminated against without the people about whom the data is about having insight. See Taylor, Floridi, & van der Sloot (2017).

TELLING STORIES WITH CITIZEN-GENERATED DATA

CGD can be turned into a narrative in different ways. Which communication method to choose depends on the message, the audience to be reached, and their data literacy and 'graphicacy' (the ability to read and understand graphics). This section gives an overview of how CGD projects turn data into meaningful stories as well as their communicative strengths and weaknesses.

DATA VISUALISATION

Data visualisation is described as "the visual representation of statistical and other types of numeric and non-numeric data through the use of static or interactive pictures and graphics. Data visualisation does not replace narrative, but is often used in combination with it to improve understanding"[34]. Data visualisation is useful to find patterns and gaps in complex data. To design visualisations well, data needs to adhere to a couple of quality criteria: In order to tell 'the right' stories through visualisations, the underlying data has to be compatible and comparable, valid, and reliable.[35] Furthermore, visualisations are helpful for "preparing visuals for **specific audiences** [emphasis added by the authors], identifying [the relevant] 'story' and the appropriate chart to use, displaying relationships that the brain can process more quickly, and uncluttering so as to not detract from the main story".[36]

Data visualisations can be divided into four broad categories: comparison (such as bar charts or tables), distribution (such as histograms), relationships (such as scatter or line charts), and composition (such as pie charts and stacked column charts).[37] Before visualising facts, it is important to understand the key messages the data tells us: For instance a CGD project might want to compare the total deaths due to car accidents between countries with huge differences in population sizes.

---

34 See also: Gatto, M. (2015): Making Research Useful: Current Challenges and Good Practices in Data Visualisation.

35 In this context, high quality data is understood as compatible, adequately aggregated, valid and reliable. See also Robinson, I. (2016): "Excel sheets aren't everyone's friend": How data visualization can assist research uptake. Available at: **http://datadrivenjournalism.net/news_and_analysis/excel_sheets_arent_everyones_friend_how_data_visualization_can_assist_**

36 Gatto, M. (2015): Making Research Useful: Current Challenges and Good Practices in Data Visualisation.

37 Andrew Abela compiled a 'chart chooser' exemplifying different styles of data visualization. It builds on the work of Gene Zelazny's book 'Saying it with Charts'. The 'chart chooser' is available at: **http://www.verstaresearch.com/types-of-charts.jpg**. For projects wondering about which chart type to use, Juice Analytics have created an interactive tool with filters to guide them. See: **http://labs.juiceanalytics.com/chartchooser/index.html**.

In this case, the number of car accidents should be adjusted *per capita* and not be compared in absolute numbers, because the resulting large differences can stem from the differences in population size rather than number of accidents.

THE VALUE OF A QUALITATIVE INDICATOR

Hence data visualisations should be used carefully in order not to distort information. An example is the CIVICUS monitor analysing the status of 'civic space' around the world. It combines a heat map visualisation with narrative reports (see Graphic 2). The monitor measures whether a nation state "respects and facilitates (citizen's) fundamental rights to associate, assemble peacefully and freely express views and opinions"[38], as well as the degree to which it protects civil society. Civic space is categorised as open, narrowed, obstructed, repressed, and closed civic space. These categories are plotted in different colours onto a world map.

Graphic 2: Example of a heat map used by the CIVICUS Monitor (Source: monitor.civicus.org)
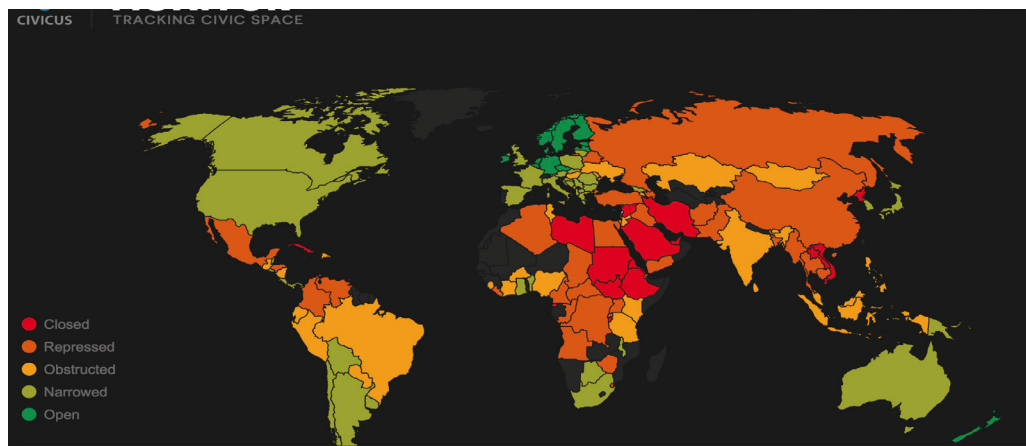
The CIVICUS Monitor heat map does not rank countries according to numerical scores. This stems from the fact that the nuances in civic space are context-specific and not standardisable without reducing the meaning of what is measured as civic space. The heat map enables users to get an overall impression of civic space around the world. Users can select single countries on the map and read their country profiles. A quantitative ranking would narrow the notion of civic space to mechanical descriptions such as 'number of allowed demonstrations per year'. These numbers divert attention away from the political context that brings these numbers into being. Using broader categories such as open or closed civic space helps users to envision broad differences of 'civic spaces', while acknowledging local differences.

38 See also: https://monitor.civicus.org/whatiscivicspace/

Therefore the heat map context-specific nature of civic space that makes a qualitative assessment more sensible.[39] Country profiles providing more nuanced information on local situations which are by default not countable.

DATA DASHBOARDS

Originally used in cars to indicate the most important information about the engine, data dashboards are nowadays increasingly used in the context of data visualisation. Dashboards represent the most important information as graphics in a visual display so they can be digested and monitored at a glance.[40] As such, dashboards allow to rank, order, and emphasise the most important information for decision-making which makes them a prominent tool for performance measurement in different societal areas, including business management, security management, or urban governance.[41] While dashboards simplify flows of information, they are criticised for potentially being overly reductionist.

> *"Although dashboards are increasingly our analytical window into the world of data, they are not necessarily neutral purveyors of that data. They invariably shape and prioritise the information that is presented (...). Which metrics are privileged? Who decides when a particular indicator moves into the red? How regular is the refresh rate, that is, what kind of temporality is built into the dashboard and how does that move us to act? Which metrics are not available, or deliberately left out?"[42]*

These general definitions cannot prevent that there seems to be confusion about what may count as a dashboard and that there are several design decisions to choose from.[43] The requirements of the data (timely coverage, standardisation, interoperability, etc.) depend on the data visualisations used. **Box 1 describes two different dashboards, discusses their communicative strengths and weaknesses, and to whom they are most useful.**

---

39 The choice whether to use a quantitative or a qualitative indicator therefore depends on the use case and what the data should be used for.
40 See also: Kitchin, R., Lauriault, T., McArdle, (2015).
41 See also Bartlett, J., Tkacz, N. (2014).
42 See also Bartlett, J., Tkacz, N. (2014).
43 For some discussions, see also: Chambers, L. (2016): Keep Calm and Make a Dashboard.
Available at: **http://techtohuman.com/dashboards**, as well as Akkerman et al. (2014): Dashboard of Dashboards. A Visual Provocation to Provide a Dashboard Critique.
Available at: **https://digitalmethods.net/Dmi/DashboardsOfDashboards**

**BOX 1: TWO EXAMPLES OF DASHBOARDS AND THEIR USE CASES**

Public service delivery–The Open311 dashboard

This dashboard shows how city data can be aggregated and related to each other. The Open311 dashboard presents public service issues such as potholes, noise nuisance, and others. **The dashboard ranks, compares and calculates quantitative data** including the total number of issues in a city, the number of issues in a given location or neighborhood (geo-coded issues), the most recent issues, or the most often reported issues. The dashboard indicators are designed to show public service performance counting and comparing issues reported and handled. To build a reliable indicator it is necessary that the dashboard uses data which is interoperable and comparable. It is for instance possible that someone would want to compare the number of two different issues in different neighbourhoods. One neighbourhood names an issue 'street damage', the other names it 'damages on street and sidewalk'. In order to reliably compare both, it must be clear that the issue 'street damage' includes damages of street signs. The **Open311 dashboard is a tool to measure performance** (response time, response rate, etc.). An interviewee stated that such information is **usually relevant for the heads of public works departments**. Contractors handling issues on the ground however were more interested in locating the issues and getting detailed descriptions of the issue (in form of text-based descriptions) in order to handle the issue more efficiently. Both user groups need different data and different visualisations to work with effectively.

Graphic 3: Dashboard indicating city performance indicators
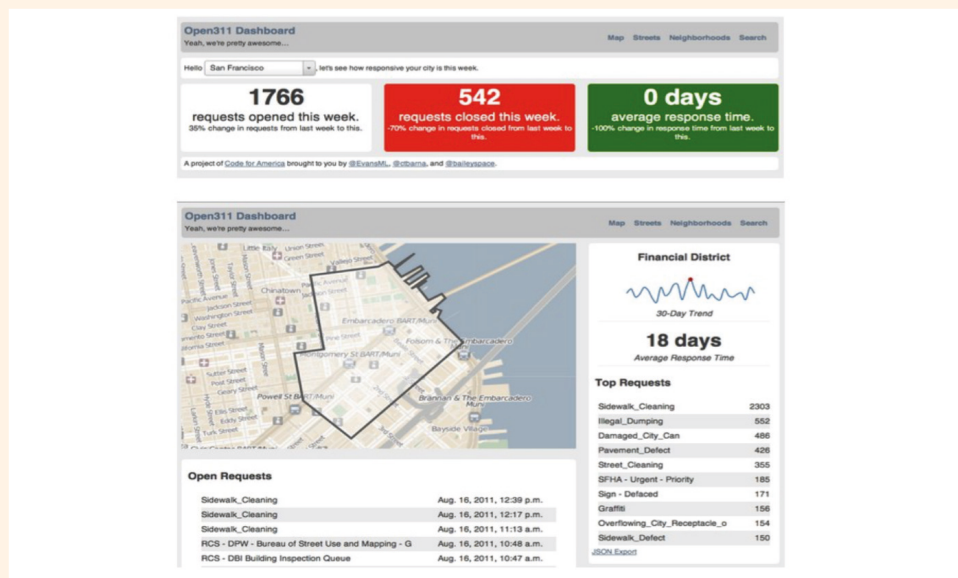
Health-related SDGs

The Institute for Health Metrics and Evaluation (IHME) developed a website with interactive visualisations of health-related statistics available for several countries from 1990 until 2015. The website allows, among others, to compare single indicator scores (See Graphic 4, sun diagram to the left), and to understand how one single indicator developed over time (see Graphic 4, line diagram to the right).

Graphic 4: Interactive dashboard: (Source: Institute for Health Metrics and Evaluation)
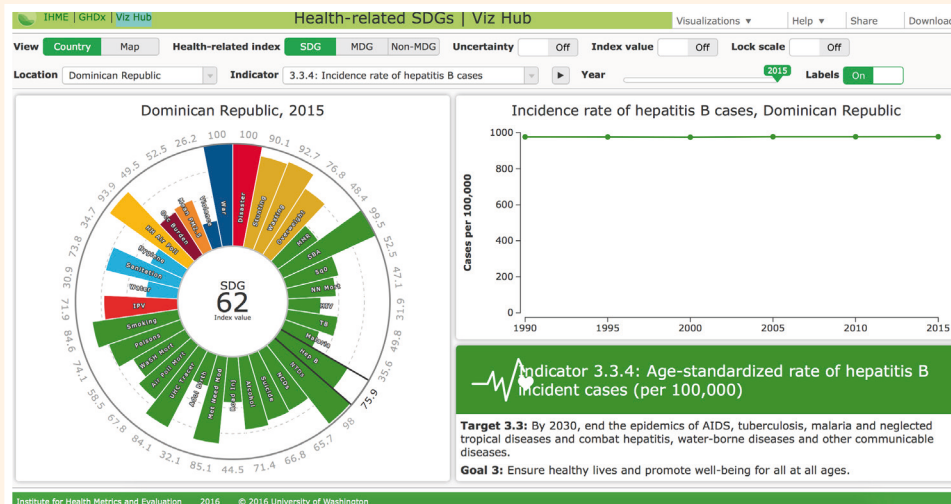
The graphical representations offer different insights–such as how indicators compare to one another. In order to do so, the platform mainly uses **relative indicators**, such as hepatitis B cases per 100,000 persons (right image). The indicators to the left in graphic 4are made comparable by **adjusting their scores** to a common scale.

QUALITATIVE DATA STORIES

There are several ways of using qualitative data for storytelling. Besides aggregating qualitative data through coding schemes (see section on data processing), qualitative data can be embedded into maps as added information which links human stories to their place. The North West Bushwick Community Map emerged from a citizen initiative to fight displacement and other effects of gentrification in New York City by "focusing on human stories behind data"[44].

---

44 Segal, P. (2015): From Open Data to Open Space: Translating Public Information Into Collective Action, Cities and the Environment (CATE), 8 (2). Available at: **http://digitalcommons.lmu.edu/cate/vol8/iss2/14**

It combines the city's open data of land use, rent stability, and others with background information of the issue, and human stories of gentrification. Personal stories are collected by housing organisers and through the project's own investigations. A project member argues that highlighting personal experiences puts social and political issues on the map which rent price maps do not tell on their own.[45] The map allowed to make the effects of rezoning, gentrification, and displacement tangible and helped the project becoming part of several coalitions with non-profit organisations and government officials.

Graphic 5: Visual representation of the Bushwick Neighbourhood, geo-locating qualitative stories in the map (left image), and patterns of land usage (right image) (Source: North West Bushwick Community project).
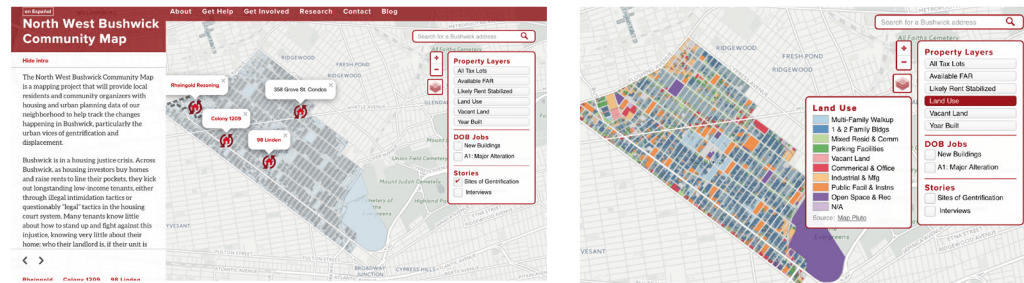
ENGAGING BEYOND MEDIA: TARGETED ENGAGEMENT

CGD projects may foster engagement by employing multiple communication channels to reach different audiences.[46] To do so, CGD projects engage team members covering a broad range of skills, including data analysts, designers, or communications experts. In some cases, CGD projects may need to collaborate with external figures such as journalists, advocates, and public figures. The InfoAmazonia is a good example where several organisations and journalists work together to report the environmental damage in the Amazon region.[47] Targeted engagement strategies are relevant across sectors and issues: Follow the Money Nigeria engages with different audiences such as beneficiaries, policy-makers, public sector officials, communities and news media to understand whether and how money travels from the public purse to recipients. Each stakeholder is addressed with different data acknowledging that information has different relevance to them.

45 Interview with a project member of the North West Bushwick Community Map.
   See also: **https://www.bushwickcommunitymap.org/html/about.html?language=en**
46 Lämmerhirt, D., Jameson, S. and Prasetyo, E. (2016). How to Make Citizen-Generated Data Work.
   Towards a framework strengthening collaborations between citizens, civil society organisations, and others
47 See also: **https://infoamazonia.org**

Graphic 6: Information material of the Living Lots NYC project in New York City, installed on fences (left) and printable maps (right) (Source: Segal, P. 2015).
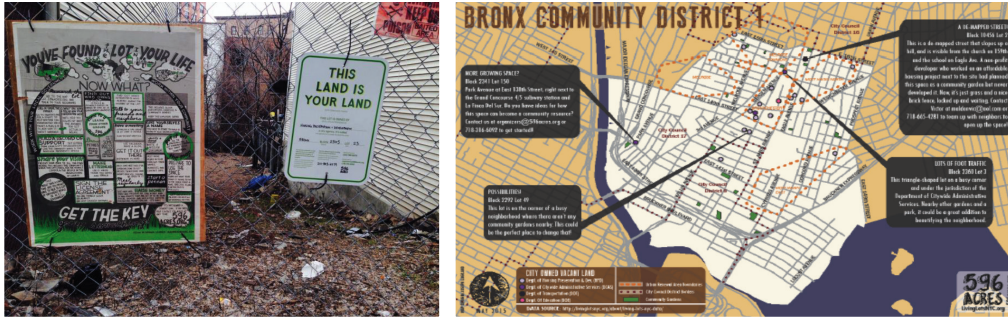
Another example is Living Lots NYC, a project strengthening the confidence of communities to reclaim publicly owned land in New York City. To engage with different audiences such as communities and urban planners, the project members use an online platform, a newsletter, and face-to-face communication. Particularly interestingly, the project is

> *'putting information about the city's vacant land portfolio where people most impacted by vacant lots will find it–on the fences that surround [them]. [see Graphic 4] The signs announce clearly that the land is public and that neighbours, together, may be able to get permission to transform the vacant lot into a garden, a park or a farm".[48]*

By diversifying the communication channels, the project is able to be heard by many stakeholders, including those who have an interest in reusing vacant land and are immediately affected by it in their everyday lives, but who would normally not consult a webpage to learn about it. Similar forms of mixed-media are public hearings bringing together different stakeholders.

CONSIDERING THE UNINTENDED EFFECTS OF DATA

Data not only represents, but also creates the worlds we live in–shifting our attention and shaping the realities that matter to us. CGD projects should be mindful which details it wants to use to convey which message. Performance indicators, rankings, and other classifications for instance are not only designed to measure activities; they also intend to improve behaviour. School 'brandings' and rankings like the school diversity index want to highlight which schools perform best in providing racially diverse classes.

---

48 See also: Segal, P. (2015): From Open Data to Open Space: Translating Public Information Into Collective Action, Cities and the Environment (CATE), 8 (2). Available at: **http://digitalcommons.lmu.edu/cate/vol8/iss2/14**

They penalise 'black schools' which are much more diverse in the way they teach and organise education. How data classify and rank therefore influences whether the problems they seek to tackle are alleviated or exacerbated.[49]

KEY TAKE-AWAYS

- **The interpretation of CGD should be performed with much care**. Data can easily be misinterpreted, and can lead to wrong conclusions. CGD projects therefore need to understand what the key messages of the data are, as well as sources of error. If necessary partnerships with trusted organisations such as universities or methodologically advanced civic groups can help.

- **CGD projects should document clearly what the data tells**, how it was analysed, and what its strengths and weaknesses are.

- **CGD projects craft messages: that resonate with the needs of stakeholders** Data should be translated in a way that is understandable and relevant to stakeholders. Long, detailed reports might interest researchers, while 'killer charts', data visualisations, and concise information might appeal to busy decision-makers.

- **Data visualisations** help communicate information and patterns in complex data but demand data literacy and graphicacy. Often, readers also need topical knowledge to interpret the sometimes complex underlying information of data visualisations.

- **Targeted engagement strategies do not end with publishing CGD reports or visualising data online.** Instead, the engagement methods need to be suitable for individual stakeholders. Examples are public hearings, education meetings with local decision-makers, on-site visits with decision-makers, hackathons, or others.

- **Partnerships are an important means to transfer knowledge, establish trust, and make key messages graspable**. This can include partnerships with trusted or experienced 'knowledge brokers' such as universities and media outlets, or close collaborations with local decision-makers.

---

49 Espeland, W., Sauder, M. (2009): Rankings and Diversity.
   Available at: http://lawweb.usc.edu/why/students/orgs/rlsj/assets/docs/issue_18/Rankings_and_Diversity.pdf

TURNING EVIDENCE INTO ACTION

Ultimately, CGD aims to drive change around an issue. How this change is envisioned firstly depends on the issue and on the stakeholders able to bring about change. Based on our case studies and a review of literature we propose **five broad forms of action forming part of an Issue Lifecycle (see Graphic 7).**

These forms of action imply that actions can be taken at different stages of an issue, be it at the very beginning when an issue is unknown, or to review existing solutions to deal with an issue. We suggest this model to understand how data can inform decisions and other forms of action. Our model is adapted from the Policy Cycle[50] which is used to understand the process of evidence-based policymaking and more narrowly focused on decisions within government.

The stages of the issue cycle are not linear but interwoven. For example, a CGD project might detect new issues when monitoring or reviewing another issue. In practice the differences between monitoring, review, and identifying new issues are not clear-cut. This, however, does not delimit the use value of the cycle. We propose to use it to inspire our thinking about possible interventions into issues.

For each stage CGD can have different functions. Table 2 demonstrates how example projects employ CGD in different stages of an issue. The projects often not only tackle one phase of an issue but still have a main focus, to which they are assigned in the table below. The table demonstrates that different forms of data quality can become important to stimulate different forms of action depending on the audience. It also shows that there are other forms of action which may evolve from the main activities of a project.

---

50 See also Sutcliffe, S., Court, J. (2005): Evidence-based Policymaking:
What is it? How does it work? What relevance for developing countries? Available at:
https://www.odi.org/publications/2804-evidence-based-policymaking-work-relevance-developing-countries
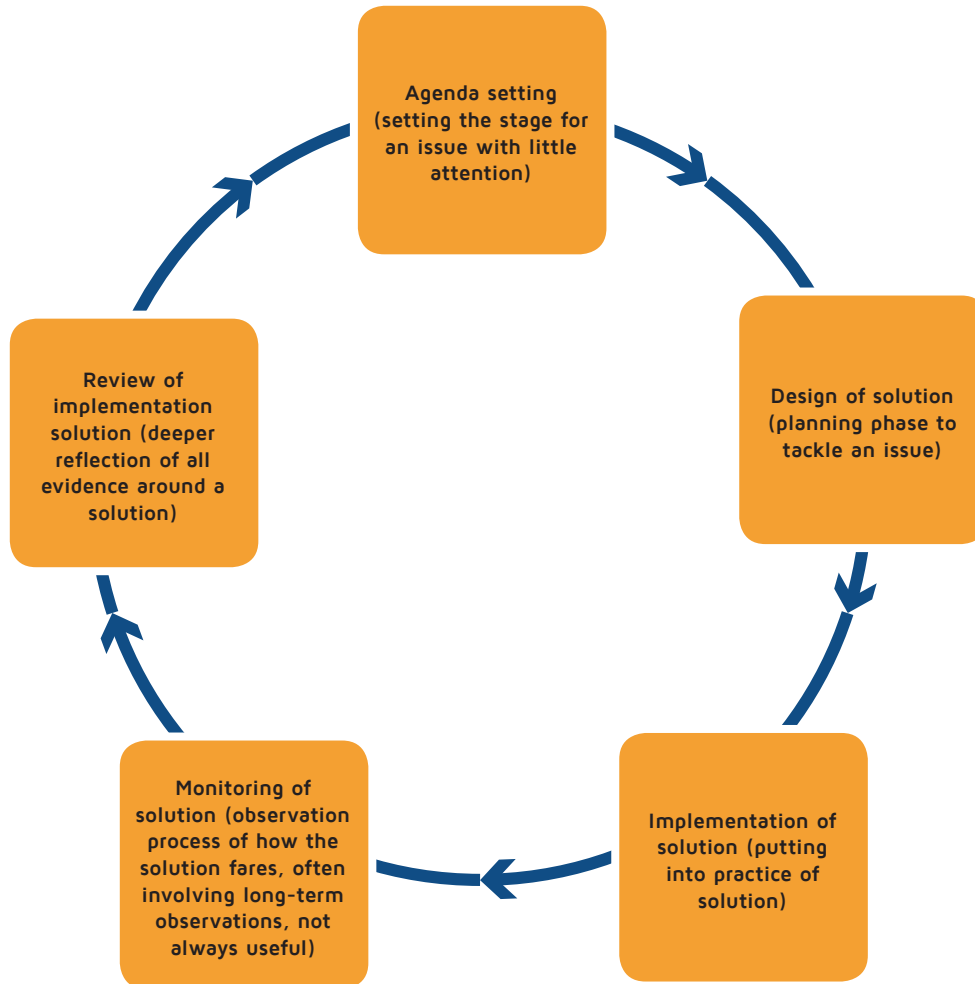
Graphic 7: The Issue Cycle

Agenda setting (setting the stage for an issue with little attention)

Design of solution (planning phase to tackle an issue)

Implementation of solution (putting into practice of solution)

Monitoring of solution (observation process of how the solution fares, often involving long-term observations, not always useful)

Review of implementation solution (deeper reflection of all evidence around a solution)

Table 2: Types of action and relevant data qualities

PROJECT AND PURPOSE	DATA USED
AGENDA SETTING / ISSUE DISCOVERY	
Project name: Harassmap **Purpose:** The project aims to create a platform to show the magnitude of sexual abuse and harassment. **Stakeholders:** local citizens, students, public service providers	The project uses reports submitted by citizens through an online platform, text message, and social media accounts. The data are presented on an online map and in research reports.
DESIGN OF SOLUTION	
Project name: Living Lots NYC **Purpose:** The project uses spatial information on vacant city-owned land to enable urban dwellers in poorer neighborhoods to turn them into community-owned areas such as parks and gardens. **Stakeholders:** neighborhood communities, urban planning department	New York City's open data on land ownership, urban renewal plans (accessed through freedom of information requests), complemented with data held by a local NGO registering urban gardens in NYC.
IMPLEMENTATION OF SOLUTION	
Project name: WeFarm **Purpose:** It uses SMS services to enable farmers to share questions they face with their crop cycles. Other farmers give them advice. **Stakeholders:** smallholder farmers	Unstructured texts sent via SMS
MONITORING OF SOLUTION	
Organisation name: Social Justice Coalition; Ndifuna Ukwazi **Purpose:** Both organisations run a project to monitor the provision of janitor services in informal settlements. **Stakeholders:** local communities, municipal government, janitors	The project uses standardised surveys to compose process and outcome indicators. The standardised surveys enable it to monitor i) the number of janitors employed, ii) how they are equipped, iii) whether toilets are broken, iv) how citizens perceive of service quality.
EVALUATION OF IMPLEMENTED SOLUTION	
Organisation name: Africa's Voices Foundation **Purpose:** Gaining understanding in perceptions and collective norms of citizens **Stakeholders:** citizens as beneficiaries of development projects, development actors	Perceptions to understand why development projects are rejected.

QUALITY CRITERIA FOR UPTAKE	OTHER ACTIONS EVOLVING FROM PROJECT
The project provides data on the location, time, and type of sexual abuse. It shows the magnitude of sexual harassment synthesised from large amounts of quantitative data (which are not comprehensive). The forms of sexual abuse are evaluated through an in-depth reading of qualitative data. Both types of data are **based on surveys with randomly selected participants**.	Harassmap exceeds mere agenda setting by actively crafting and implementing solutions together with other partners who have different degrees of influence in mitigating harassment. Actions include: i) guiding police presence by visualising harassment hotspots, ii) creating harassment free zones in collaboration with shop owners iii) create policy guidelines for sexual harassment reporting in schools and universities
Since the project wants citizens to reimagine and repurpose urban spaces, data needs to be **understandable** to citizens. This is achieved through a mixed-media strategy (see Section: Telling Stories With Citizen-Generated Data).	Living Lots NYC provides legal advice and technical assistance in order to inform residents about possible political interventions such as i) applying for approval of community spaces from the local Community Board ii) winning endorsement from locally elected officials iii) negotiating with the responsible agency holding titles to a lot
Main enabler is **trust** among farmers. The information exchanged between farmers is also **relevant** in local contexts. Farmers have local, tacit knowledge that can be shared and immediately applied.	Data can be aggregated into issue types and catered to other audiences like agribusiness. Thereby it informs reviews of value chains, or the design of new solutions supporting smallholder farmers.
Accuracy is assured through training that sets assessment criteria (for instance, when does a toilet count as broken). Impact achieved because the **data indicated deficiencies in this public service.** The janitor service improved partly due to **public attention and rising political costs,** even though local government initially rejected the findings as neither representative nor credible.	CGD projects enable local community members to 'read' and understand how bureaucratic procedures work. Being involved in the data design process i) teaches citizens how policies are designed, ii) renders policies tangible, iii) gives communities an argumentative basis within which to ground their evidence for public service deficiencies (and gain confidence to engage with government).
Accurate data about socio-demographics (sex, location,). Not representative for total population, but displaying sub-populations. Embracing biased answers as possibility to foregrounding perceptions.	Citizens are 'heard' and have a feeling of acknowledgement for their problems.

5 CONCLUSION

This paper demonstrates that CGD builds evidence to inform diverse types of human decision-making, from agenda setting, to the design, implementation and monitoring of solutions. It is thereby much more than a mere device for agenda setting. CGD can inspire citizens to design their own solutions. It can also give citizens the literacy to 'read' and understand governance issues and thereby provide confidence to engage with politics. Sometimes data can be used to directly implement a solution to an issue, or it can be a monitoring device. The value of CGD is thus very broad and depends largely on the issue it is used for and the individuals, groups, organisations, and networks using it. These actions are important drivers to promote progress on sustainable development issues–and CGD often gets little attention in current debates.

Yet CGD is not a panacea. It should be noted that actions can be the result of engagement over a long time and might need political 'heavy lifting' and 'working the system'. CGD projects focusing on improving public services for instance, need to provide meaningful information to support their claims, gain trust from stakeholders (including government), and offer practical solutions to problems. These actions are beyond the mere act of collecting data or publishing it in a data visualisation. In order to drive action, CGD projects should be seen as socio-technical systems composed of people, perceptions, tasks, information and technologies to capture and process data.[51] Therefore the way evidence turns into action often remains a very human issue.

The report thereby shows that the usual understanding of 'good data quality' is more nuanced and not only a matter of rigorousness, validity or representativity. It does not mean that methodological rigour is irrelevant for CGD. The opposite is the case. Data should be thoughtfully and holistically designed in order to address specific tasks and to respond to the human elements of data quality (Is data credible and trustworthy? Who defined the data methodology? etc.). A human-centric understanding of data quality also acknowledges that data is never 'raw' but always 'cooked', meaning that decisions have to be made about which parts of reality to capture and how.

---

51 See also: Heeks, R.B. (2001) 'Reinventing government in the information age', in: Reinventing Government in the Information Age, R.B. Heeks (ed.), Routledge: London, 1-21

This holds true for all types of data, including numbers, which are often seen as sterile facts born out of standardised data creation procedures. Hence, numbers and other quantitative data is not more valuable or more reliable than other data. What matters is that citizens collect data in a systematic way that demonstrates how the data was collected, and processed in the first place.

Importantly, different types of data have different usefulness. The term data itself seems to suggest a very narrow notion of numbers, figures, and statistics. **Debates around the data revolution or sustainable development data should not gloss over the fact that narrative texts, individual perceptions, interviews, and images all count as 'data'–which might be best understood broadly as a building block of human knowledge, decision-making, and action.** Referring to the Sustainable Development Goals (SDGs), CGD can help to overcome silo thinking, and to understand the sustainability issues more contextually. Much focus has been paid to monitoring progress around the SDGs via national statistics. On the contrary, CGD can actively enable to drive progress around sustainability on the ground. As such, a broader vision of data is needed, which puts issues and humans in its center. Therefore the report suggests that decision-makers, government, local communities, civil society organisations first put the issue before the data and then consider which type of data is best suited to address it. Such an approach would acknowledge that different data can have a diverse, but equally important value for decision-making than official statistics.

FURTHER READING

AN INTRODUCTION TO CITIZEN-GENERATED DATA

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**

Datashift (2016). Making Use of Citizen-Generated Data.
    Available at: **http://www.data4sdgs.org/guide-making-use-of-citizen-generated-data/**

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**

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**

Lämmerhirt, D., Jameson, S. and Prasetyo, E. (2016). How to Make Citizen-Generated Data Work. Towards a
    framework strengthening collaborations between citizens, civil society organisations, and others. Available at:
    http://civicus.org/thedatashift/wp-content/uploads/2015/07/Making-citizen-generated-data-work.pdf

UNDERSTANDING DATA AND DATA QUALITY

Borgman, C. (2011): The Conundrum Of Sharing Research Data.
    Available at: **http://onlinelibrary.wiley.com/doi/10.1002/asi.22634/full**

Wand, Y., and Wang, R. Y. (1996) Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of
    the ACM,* 39 (11), 86-95. Available at: **http://www.thecre.com/pdf/MIT-wandwang.pdf**

Wang, R. Y. and Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of
    Management Information Systems, 12 (4)*, 5-33.
    Available at: **http://courses.washington.edu/geog482/resource/14_Beyond_Accuracy.pdf**

TURNING EVIDENCE INTO ACTION

Pollard, A., Court, J. (2005): How Civil Societies Use Evidence to Influence Policy Processes: A literature review.
    Available at: **https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/164.pdf**

Shaxson, L. (2005): Is Your Evidence Robust Enough? Questions For Policy Makers And Practitioners.
    **http://www.cepa.lk/content_images/publications/documents/131-S-Shaxson-Evidence%20&%20Policy-Is%20
    your%20evidence%20robust%20enough.pdf**

Shepherd, J. (2014): How To Achieve More Effective Services: The Evidence Ecosystem.
    Available at: **http://orca.cf.ac.uk/69077/**

Shucksmith, M. (2016): InterAction. How Can Academics And The Third Sector Work Together To Influence Policy
    And Practice? Available at: **http://www.carnegieuktrust.org.uk/carnegieuktrust/wp-content/uploads/
    sites/64/2016/04/LOW-RES-2578-Carnegie-Interaction.pdf**

Sutcliffe, S., Court, J. (2005): Evidence-based Policymaking: What is it? How does it work? What relevance for
    developing countries? Available at:
    https://www.odi.org/publications/2804-evidence-based-policymaking-work-relevance-developing-countries

The Overseas Development Institute, (2009). *Planning Toolkit: Stakeholder Analysis*.
    Available at: **https://www.odi.org/publications/5257-stakeholder-analysis**

DATA VISUALISATION: BACKGROUND AND BEST PRACTICES

Gatto, M. (2015): Making Research Useful: Current Challenges and Good Practices in Data Visualisation.
    Available at: **https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Making%20Research%20Useful%20
    -%20Current%20Challenges%20and%20Good%20Practices%20in%20Data%20Visualisation.pdf**

Data-Driven Journalism: Various articles available at: **http://datadrivenjournalism.net**

NOTES

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 **@SDGdatashift**

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.