Do the numbers add up?

How data inform decisions at the sub-national level
Introduction

As part of Tanzania’s open government efforts, the government’s open data portal (opendata.go.tz) was launched in 2014. The site hosts an ever-growing collection of data, available “to be used and re-used by anybody,” including extensive data on health, education and water services.

This brief presents the findings of a study looking at potential and actual users of data in Tanzania, covering civil society and local government. It asks what data they are currently accessing, where they are accessing it from, and what they are using it for. And it explores the challenges and opportunities they have found in using data, including data from the open data portal.

Introducing Open Data and opendata.go.tz

The Tanzanian open data portal describes its purpose as follows:

“By making this data publicly available, a wide range of actors can be brought into the policy process and debate, bringing valuable new ideas and new thinking to policy making, and stronger public participation in monitoring and citizen feedback.”

The data published on the site meets the definition of open data, which includes the following two elements:

i. The data is legally open – published in the public domain with no (or very minimal) restrictions on how it may be used

ii. The data is technically open – published in electronic formats that are readable by a machine, i.e. a spreadsheet or similar file, not in pdf format

A core element of the open data philosophy is that the data is available to anybody. In practice, in addition to the obvious categories of users such as academics, the technology community, the media and civil society, the data can also be very useful to others within government. It is often simpler for a civil servant to access data from an online source, for example, rather than requesting it from a different department through formal procedures.

At the time of writing, the Tanzanian government’s open data portal hosts 163 datasets, the majority covering priority sectors of education, health and water.

Table 1: Examples of data available from opendata.go.tz
<table>
<thead>
<tr>
<th>Sector</th>
<th>Dataset</th>
<th>Details</th>
<th>url</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>Pupil to teacher ratio for government primary schools, 2016</td>
<td>Includes every government primary school in Tanzania (over 16,000 schools), the number of enrolled pupils and teachers and pupil-teacher ratios.</td>
<td><a href="http://opendata.go.tz/dataset/uwiano-wamwalimu-kwaswanafunzi-kwa-shule-zammsingi-za-serikali-2016">url</a></td>
</tr>
<tr>
<td>Education</td>
<td>National form four examinations schools ranking with location, 2012-14</td>
<td>Showing form four pass rates for each government and private Secondary School (4,407 schools in total), for each year from 2012 to 2014, includes geographical coordinates of each school.</td>
<td><a href="http://opendata.go.tz/dataset/national-form-four-examinations-schools-ranking-with-location">url</a></td>
</tr>
<tr>
<td>Water</td>
<td>Rural waterpoints, Tanzania mainland</td>
<td>Includes every public waterpoint in rural areas of mainland Tanzania, the location, type of source, and details of whether each waterpoint is functional.</td>
<td><a href="http://opendata.go.tz/dataset/water-points-in-rural-water-suppy">url</a></td>
</tr>
<tr>
<td>Health</td>
<td>List of health facilities with geographical location, 2014</td>
<td>Details of over 7,000 health facilities nationwide, with the geographical location of each. Includes details of the type and ownership of each facility.</td>
<td><a href="http://opendata.go.tz/dataset/list-of-health-facilities-with-geographical-location">url</a></td>
</tr>
<tr>
<td>Health</td>
<td>Number and causes of death occurred by region</td>
<td>Showing how many people died in each region of mainland Tanzania for each of 50 possible causes of death, including malaria, tuberculosis, heart failure and HIV/AIDS, 2012 to 2014.</td>
<td><a href="http://opendata.go.tz/dataset/number-and-causes-of-death-occured-by-region">url</a></td>
</tr>
</tbody>
</table>

Further, the site provides a number of tools and dashboards designed to present key data in more easy-to-understand visual formats, such as maps and charts, thus broadening the range of users who can benefit from the site beyond those with data handling skills.
Methodology and sampling

The survey was designed as a quantitative assessment, administered through CATI (Computer Aided Telephonic Interview). Data collection took place from February 2 to 24, 2017, achieving a sample size of 170 respondents across three categories of data users.

Table 2: Breakdown of survey respondents

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGA officials</td>
<td>Civil servants from Local Government Authorities (LGAs), primarily from the water, education and health sectors</td>
<td>50</td>
</tr>
<tr>
<td>Local civil society</td>
<td>Representatives of Twaweza’s partners on the Uwezo initiative and grantees of the Foundation for Civil Society (FCS), randomly selected from the full list of Uwezo partners and FCS grantees</td>
<td>100</td>
</tr>
<tr>
<td>National civil society</td>
<td>Representatives of all national-level civil society organisations engaged in the Open Government Partnership</td>
<td>20</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>170</strong></td>
</tr>
</tbody>
</table>

This relatively small and non-random sample presents some challenges when it comes to representativeness. The data should therefore be interpreted with some caution. However, for an exploratory study looking at a small and narrowly defined population (civil society representatives and LGA officials), a sample of this nature can still offer very valuable insights.

Table 3: Breakdown of survey respondents by sectors of operation

<table>
<thead>
<tr>
<th>CSO respondents</th>
<th>LGA respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>61% (73)</td>
</tr>
<tr>
<td>Health</td>
<td>60% (72)</td>
</tr>
<tr>
<td>Agriculture</td>
<td>26% (31)</td>
</tr>
<tr>
<td>Water</td>
<td>15% (18)</td>
</tr>
<tr>
<td>Others¹</td>
<td>50% (60)</td>
</tr>
<tr>
<td>Education</td>
<td>36% (18)</td>
</tr>
<tr>
<td>Health</td>
<td>32% (16)</td>
</tr>
<tr>
<td>Water</td>
<td>28% (14)</td>
</tr>
<tr>
<td>Land / engineering</td>
<td>4% (2)</td>
</tr>
</tbody>
</table>

¹ Including environment, law, women’s rights, community development, etc.
Findings

User and uses of government data

The vast majority of respondents (94%) report using data collected by government in their regular work. This includes all respondents from LGAs and national civil society.

![Figure 1: In your regular work do you use / refer to data collected by the government?](chart1)

Among those who report using government data, one in four (25%) reporting having looked for such data in the past week. A further four in ten (39%) report doing so in the last month. Local civil society representatives were less likely to have sought government data in the past month (47%) than other groups: 92% of LGA staff and 75% of national civil society respondents, for example.

![Figure 2: When did you last look for government data on public services?](chart2)
When looking for such government data, respondents report using a variety of different sources. One in four (23%) report turning first to government websites. A similar number (24%) start with an informal request to a government official. Formal requests are less common (9%), and were not cited at all by any LGA officials. A small number report gathering data from lower level government institutions such as schools and villages. Reported use of websites was considerably higher among national civil society (90%) than among other groups.

The vast majority of those who sought data reported that they found it (97%) and that it was useful (95%).

Among all respondents who report using government data in their work, the most common uses of the data were to advocate for improvements in public services (41%), mobilising the community to take action (34%) and ensuring their own work meets community needs (33%).

Uses of data varied between different groups. LGA officials, for example, were more likely to report using data for planning and budgeting purposes (38%) than local CSO representatives (9%). Conversely, half of the local CSO representatives (49%) report using data to mobilise community action, compared to one in ten (8%) LGA officials.

Among national civil society, accountability purposes were the most commonly cited use of data (55%), followed by research (45%). Use of data for these purposes was reported to be much lower among other groups.
Figure 5: For what purposes do you most generally use data from government? 
(all, n=160)

- To advocate for improvements in public services: 41%
- To mobilise the community to take action: 34%
- To ensure your own work meets community needs: 33%
- To conduct research: 26%
- For monitoring and/or evaluation of own work: 19%
- For accountability purposes: 18%
- For planning / budgeting: 17%
- To advocate for changes in government policy: 9%
- For monitoring / evaluation of another’s work: 5%
- To inform the public: 2%
- Other: 11%

Figure 6: For what purposes do you most generally use data from government? 
(LGAs, n=50)

- To advocate for improvements in public services: 40%
- For planning / budgeting: 38%
- To ensure your own work meets community needs: 30%
- For monitoring and/or evaluation of own work: 22%
- For accountability purposes: 10%
- To mobilise the community to take action: 8%
- To inform the public: 6%
- To conduct research: 4%
- To advocate for changes in government policy: 0%
- For monitoring / evaluation of another’s work: 0%

Figure 7: For what purposes do you most generally use data from government? 
(local CSOs, n=90)

- To mobilise the community to take action: 49%
- To advocate for improvements in public services: 42%
- To ensure your own work meets community needs: 35%
- To conduct research: 35%
- For monitoring and/or evaluation of own work: 17%
- For accountability purposes: 14%
- To advocate for changes in government policy: 10%
- For planning / budgeting: 9%
- For monitoring / evaluation of another’s work: 8%
- To inform the public: 0%
- Other: 11%
The open data portal

Four in ten respondents (40%) said they were aware of the open data portal, though some (2%) only recognised it after a short explanation had been given to them. Recognition was much higher among national civil society (90%) – not very surprising given that these respondents represent organisations working on open government.

Among the 67 respondents who were aware of the portal, just under half (45%) said they had ever visited it. As such, out of all 170 respondents, just 30 (18%) said they had ever visited the site.

Of the 30 respondents who had visited the site, 19 (63%) said they had looked for data on the site and 15 said it was either very easy or somewhat easy to find the data they were looking for. However, only 10 said they were completely successful at finding the data, with the other 9 saying they were partly successful.
Of the 19 respondents who had looked for data on the site, 10 (53%) said they had done so within the last month.

The majority (73%) of the portal’s visitors reported using a personal laptop. The remainder all reported using a work computer. Nobody reporting visiting the site from an internet café or public place, or on a mobile phone.

Among those who knew of the website but had not visited or used it, two in three either didn’t know enough about the site (35%) or preferred to use a source they were more familiar with (33%).
Combining the data above into a single image, we can see that several hurdles will need to be overcome in order to convert potential users of the portal into successful users. Of the 170 respondents, 67 had heard of the portal, 30 had visited it, 19 had sought data, and just 10 (6%) had successfully found the data they were seeking.
Perceived relevance of example datasets

Those who had not visited the portal were given a brief explanation of one dataset as an example of what is available on the site. Of these, around half (46%) said the example they were given sounded either very or somewhat relevant to their work.

Data handling skills

Respondents were asked to rate their own personal skills for working with spreadsheets, as a key component of data handling skills. Among LGA staff, one in five (20%) rated their own personal skill level as “very advanced”, but over half (54%) rated their skills as basic. Local civil society rated themselves a little higher. However, national civil society representatives were much more confident in their skill levels, with seven in ten (70%) saying they had either some advanced skills or very advanced skills.

In all these cases, it must be noted that respondents were simply reporting skill levels – it was not possible to actually test skill levels as part of this exercise.

Asked to rate the spreadsheet skills of others in their department, respondents in local civil society reported more confidence in the rest of their team’s skills than they had in their own expertise. Over half of this group said others in their team had “very advanced” skills.
Figure 20: What level of skills do you feel others in your department/organisation have for using spreadsheets, such as Excel, in your work?

<table>
<thead>
<tr>
<th></th>
<th>Very advanced skills</th>
<th>Some advanced skills</th>
<th>Basic skills only</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (n=170)</td>
<td>45%</td>
<td>20%</td>
<td>32%</td>
<td>3%</td>
</tr>
<tr>
<td>LGA (n=50)</td>
<td>20%</td>
<td>24%</td>
<td>52%</td>
<td>4%</td>
</tr>
<tr>
<td>Local CS (n=100)</td>
<td>53%</td>
<td>18%</td>
<td>27%</td>
<td>2%</td>
</tr>
<tr>
<td>National CS (n=20)</td>
<td>70%</td>
<td>20%</td>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Legend:
- Very advanced skills
- Some advanced skills
- Basic skills only
- None
Conclusions

The findings of this exercise provide a useful guide for anyone with an interest in increasing and improving the use of data among employees of local civil society and local government authorities. Many of the conclusions are relevant well beyond the open data portal that was the specific focus of some questions. While the detailed analysis above provides a number of insights, three overarching conclusions stand out:

First, respondents report a high level of readiness to use government data. 94% reported that they are already doing so, with over half (64%) having done this in the past month, rising to 92% among LGA officials. Almost all of those who had sought data said they had been able to find it (97%), and almost all said they had found it useful (95%). Further, half (49%) reported having at least some advanced skills in working with spreadsheets, and two thirds (65%) that someone in their department or organisation had some advanced skills.

This exercise did not, however, include an independent test of data handling skills or of the quality of work being done with data, and as such, this conclusion should be treated with some caution. However, at the very least, we can state with some confidence that representatives of local government and local civil society organisations feel that they are already able to access data and use data.

Second, government websites do not currently represent the main source of data for these users. Less than one in four reported that the first place they looked when in need of data was a government website. This figure was substantially lower among LGA officials, with just 8% of this group turning first to government websites. Informal requests from government officials were a more common approach when seeking data. Interestingly, not one government official in the sample reported having made a formal request for data: within government, informal requests are how data moves around.

The exception here is national civil society. Among this group, 90% cited government websites as the first place they looked for government data.

Third, it is clear that several hurdles will need to be overcome in order to convert potential users of the open data portal into successful users. Well under half our respondents (40%) reported having ever heard of the portal, and less than one in five (18%) reported having ever visited the site. Indeed, only 19 out of 170 potential users selected from within the site’s key target audiences had ever sought data on the portal, and just 10 had successfully found the data they were looking for. Among those who knew of the site but hadn’t visited it, two thirds said the reason was either that they didn’t know enough about the site or preferred to use sources they were more familiar with. Addressing these challenges will therefore require further concerted efforts at awareness raising, along with continuing to add more datasets and more detailed datasets to the portal.