Catalyzing Innovation
ANNUAL WORLD BANK CONFERENCE ON LAND AND POVERTY
WASHINGTON DC, MARCH 25-29, 2019

CAPTURING DATA GAPS: COMPARATIVE STUDY ON AVAILABILITY OF LAND DATA IN AFRICA

MICHAEL OCHIENG ODHIAMBO, LISETTE MEY, LAURA MEGGIOLARO

lisette.mey@landportal.org

Paper prepared for presentation at the
"2019 WORLD BANK CONFERENCE ON LAND AND POVERTY"
The World Bank - Washington DC, March 25-29, 2019

Abstract

It is an often-repeated rhetoric that there is a lack of land data - whether it is lack of reliable or up-to-date data or a lack in the existence of any data. Collecting data is a time-consuming and costly process and one can only imagine the enormous impact new data capture technologies can have on the speed and volume of new data collection. With digitization of information, increased use of internet, and growing demand for more data, the risk is that we get swept up by the potential of the latest technology and only add to the wealth of data, without having analyzed or digested any existing data. This paper presents a scoping study in five countries in Africa to uncover the information landscape. We hope to trigger thoughts on use of the data ecosystem, rather than 'simply' adding to its continued growth.

Key Words: data ecosystem, open data, information landscape, land data, data gaps

1. Introduction

Having and using information has always been a powerful force for change, helping to fight corruption, enabling citizens to participate more fully in public life and allowing people from all walks of life to exercise their fundamental human rights. We are living in a time in which paradoxal topics such as 'fake news' and 'big data' are part of our everyday lives. These discussions are symptomatic and reflective of the fact that now, more than ever, we need governments and heads of state, policy makers and change makers, civil society, academia and the private sector alike, to make decisions that are evidence-based, track progress and clearly measure accountability.

It is often asserted that there is a lack of data in the land sector - whether it is lack of reliable or accurate data, up-to-date data or simply the absence of any data in certain areas or with regards to certain topics. Capturing and collecting data is a time-consuming and costly process and one can only imagine the enormous impact new data capture technologies can have on the speed and volume of new data gathering and how this adds to the wealth of existing data.

Data is collected and captured on a massive scale already, but research shows that of all existing data worldwide, less than 1% is actually analyzed and digested.[1] With increasing digitization of information, increased use of internet in all parts of the world, and continuously growing demand for more data, the risk is that existing data is either purposely cast aside (as the source may be from outside our trusted networks) or simply overlooked.

The current reality of land data is that in many parts of the world data remains inaccessible, fragmented, poorly managed or simply unusable. In this paper we want to call for more substantiation of the common rhetoric of the "lack of land data" and encourage actors to not immediately get swept up by the potential of the latest app or gadget, but to actually take a moment to scope, digest and understand the existing information and data landscape.

[1] John Gantz and David Reinsel, December 2012.

2. Scoping study to uncover the information landscape

The Land Portal Foundation and partners in Africa are undertaking a massive scoping study to understand the data ecosystem in five countries in Africa, namely Kenya, Tanzania, South Sudan, Uganda and South Africa. The aim of the research is to uncover the many different sources of land data and information at the country-level and help to identify actual data and information gaps, with a view to establishing a baseline for targeted 'information-based' interventions to improve the information ecosystem.

You will notice we use data and information almost interchangeably, purposely so. When we perform a scoping study on "what is known" or somehow documented about land in a country, it would be a major oversight if we only include raw data and statistical indicators. Much of what is known, particularly at the grassroots level, is not captured in an indicator, but rather in a publication or news article, for example. In this scoping exercise, we therefore very purposely talk about both data and information.

The report is useful as a tool for any land governance work that requires access to data and information. A researcher may use this report to identify gaps in information and identify research priorities accordingly. A land practitioner working at the global level may use the information sources as a basis to monitor land governance performance against international indicators. Private companies may find the report a useful starting point in due diligence processes prior to working or investing in a country. Local information providers may identify weak links in their data sharing practices and implement concrete recommendations. Governments can use the outcome to establish or strengthen their policies that aim to increase access to data and information by citizens. Ultimately, we hope the report will make data and information more visible and usable by any potential user and thus improve the local information ecosystem from the bottom-up. Ultimately, this overview will provide the general public with an entry point of any data- or information-related activity they may take part in - including applying new data capture technologies in those instances where data gaps truly exist.

3. Scoping study to uncover the information landscape

The scoping methodology has been carefully identified through thorough research but continuously enriched and improved through application and trial-and-error in previous scoping studies. Key research methods used include desk top research on existing data and information complemented by an open survey of stakeholders based in or working in the particular country, followed by a rigorous peer review process in a multi-stakeholder setting in-country.

The parameters for the scoping study were set on the basis of key land issues identified by the Land Portal. The mantra of "building on rather than duplicating" that underlies the entire effort of this study has also been applied to the process of identifying the key land issues. We drew from key land indicators and guidelines from several global and regional land monitoring initiatives, among them:

1. Sustainable Development Goals, "SDGs" (United Nations)[2];
2. Voluntary Guidelines on the Responsible Governance of Tenure, "VGGTs" (FAO)[3];
3. Land Governance Assessment Framework, "LGAF" (World Bank Group)[4];
4. Global Land Indicator Initiative, "GLII" (network facilitated by GLTN/UN-Habitat)[5];
5. Monitoring & Evaluation of Land in Africa, "MELA" (IFPRI & Land Policy Initiative)[6];
6. International Land Coalition Dashboard (facilitated by ILC)[7];
7. Africa Data Revolution Report (facilitated by Open Knowledge International)[8].

Based on the categories, indicators and principles included in these international land data monitoring and governance guidelines and frameworks, the Land Portal has grouped overlapping

[2] https://www.un.org/sustainabledevelopment/sustainable-development-goals/

[3] Food and Agriculture Organization of the United Nations, "*Voluntary Guidelines on the Responsible Governance of Tenure of Land, Fisheries and Forests in the Context of National Food Security*", Rome 2012.

[4] http://www.worldbank.org/en/programs/land-governance-assessment-framework

[5] https://gltn.net/global-land-indicators-initiative-glii/

[6] https://melafrica.wordpress.com/

[7] International Land Coalition, "*The Dashboard Indicators*", Rome May 2018.

[8] World Web Foundation, "*Africa Data Revolution Report 2018. Status and Emerging Impact of Open Data in Africa*", 2018.

indicators and principles into the following seven categories: Legal, Policy & Institutional Framework; Land Tenure data; Land Cover, Use and Management Data; Land Disputes; Human Settlements; Land Markets & Financing; and Land, Climate Change & Environment.

For each of those categories, associated key information has been identified, that also incorporate cross-cutting issues such as gender disaggregation of data and information or specific information on Indigenous Peoples, different land uses (agricultural land, forest land, urban land, etc.) and more. Specific questions about these key categories are identified to guide a scoping researcher to answer the questions and provide the dataset or information product, whatever format the information be presented in.

While our flexibility on the format of the data and information broadens the scope of this research immensely, we feel it is an essential element to our study. We use the words ‘data’ and ‘information’ almost interchangeably, and purposely so. We consider that when we perform a scoping study on “what is known” or somehow documented about land in a country, it would be a major oversight if we only include raw data and statistical indicators.

For example, much of what is known, particularly at the grassroots level, may not be captured in an indicator, but rather in a publication or news article. By including sources such as news articles, grey literature but also research articles that might use data but not publish the raw data along with their interpretation in the research article, we are truly looking at the entire information and data ecosystem, not just a small portion of it that is only useful to a small group of data scientists.

It is important to mention that although the scoping study has been as rigorous and as targeted as possible with the use of key issues around land, we do not claim to have captured all data and information ever captured or published about land in a particular country. In particular, offline resources are difficult to find and it is difficult to know all the possible sources that somehow have collected some form of data, information or knowledge about land issues. Moreover, new

perspectives are constantly being collected every day. The picture of the Information and Data ecosystems we are presenting in these reports is therefore not all-encompassing, but is, rather, a snapshot of a certain moment in time. That said, it is the most comprehensive and definitive resource for land data and information in Kenya, Tanzania, Uganda, and South Sudan and South Africa available and will be considered a "living" document that to be updated regularly.

What sets this research apart from any other (perhaps more traditional) monitoring initiatives is that the focus is on the database or dataset and its sources; with no judgment made on the value or content of the information. We intend to highlight simply that source A or B provides data about a particular key topic in dataset X or Y. This information is accompanied only by objective criteria, such as: where can you access this data (online or offline)? Do you need to log in or pay to access the information? When was this dataset last updated? Who is the source of this data?

We believe that data quality, accuracy and reliability lies in the judgement of the user, and therefore we do not feel it is useful at present to highlight which of the data sources are most credible in our opinion. We are thus presenting this report as basically "the state of land information" in a given country. It is purposely focused on the "state of land *information*", as our perception is that showing that there are two or more sources of data providing conflicting values for a particular key topic in a country -- is worthwhile.

For the very first time, we look at the entire landscape of a country to see trends and gaps when it comes to land data collection. The State of Land Information report concludes with -where necessary- concrete recommendations to data and information providers to improve their data sharing practices, so as to help establish a functioning, inclusive and democratized ecosystem of data.

4. Preliminary findings on Kenya

At the time of submitting this paper, the scoping research for Kenya has been completed, while we are finalizing the work for Uganda, Tanzania and South Sudan. For Kenya, we looked at the

entire landscape of data and information related to land in Kenya, assessing over 250 land resources from over 60 different sources, to see trends and gaps when it comes to data collection and to see how well the key categories were covered.

The statement that there is a lack of data is partially accurate: our scoping exercise shows that 80% of key land resources are available as documents, not statistical data. However, there was no key land category where no knowledge or information was found, suggesting that there is indeed knowledge generated and published, but not (yet) translated to statistical data (where possible). Our research also shows that the knowledge is published online (98%), it is available for free (98%) and largely publicly accessible without requiring registration or identification (94%). The rudimentary access to data and information seems to be in a very good state in the Kenyan Data and Information Ecosystem, but important to mention is that those resources that are online, free and accessible without barriers are also those most likely to have been identified in our scoping exercise.

Another important caveat to the statement that data was found for each key category of land, is that 43% of the resources identified were either dated from before 2010 or were published without a clear publishing date. This is a significant constraint for these resources to be useful or used. Another important aspect that defines the usability of a resource, is knowing the source of the data or information. The main source of key data and/or information identified in this scoping exercise was the government, accounting for the majority of the resources available. This is particularly the case with regards to the publication of policies and laws on land and land-based resources. Kenyalaw.org in particular, is a great resource for accessing laws in Kenya. A similar website for policies does not exist.

Research Institutions provided 25% of the total resources identified in the scoping exercise. A notable weak link in terms of information provision, were the (national) Civil Society Organizations (CSOs), which accounted for less than 15% of the total resources identified and provided little information for almost each key category. We recognize that this is not

necessarily because CSOs do not have data, information or knowledge to share. Rather, it may well reflect on their poor information sharing practices, thereby demonstrating the need to improve the discoverability of their perspectives online.

Availability of Data and Information Green = good practice; orange = practice can be improved; red = poor practice							
Key Category	Data available ?	Representation of Sources					Data up-to-date?
		Government	Research Institutions	National CSOs	Int. Organizations	Other	
Land Tenure Data							
Land Cover, Use & Management							
Land Disputes							
Human Settlements							
Land Markets & Financing							
Land, Climate Change & Environment							

5. Conclusion

With the 'State of Land Information' Methodology, we seek to provide an overview of existing data and information on key land issues and promote awareness of the data ecosystem in which we operate. Our aim is to uncover the many different sources of land data and information in several countries in Africa and thus provide a basis to substantiate, refute or nuance the oft-repeated assertion that there is a lack of land data.

To this end, we have developed a scoping methodology with concrete questions on availability and coverage of data and information based on key land categories, based on renowned international and regional land frameworks and indicators. Even with this open and inclusive process to scope the information sources, this research cannot claim to have captured everything that exists in land data. At most, it provides a snapshot of what exists at the moment of writing

the report and will become outdated the next day. It is therefore intended to be an open, “living” tool, which stakeholders are welcome and encouraged to keep updating and keep relevant.

Ultimately, we hope that we can trigger thoughts, invoke ideas and provide an easy tool to use and analyze data, rather than simply collecting more and more data based on a non-substantiated notion that no data exists - building on what is already done before we get swept away by challenging and exciting new technologies to ‘simply’ add to a data ecosystem without regard to whether and to what extent the data is being used.