Land Governance in an Interconnected World
ANNUAL WORLD BANK CONFERENCE ON LAND AND POVERTY
WASHINGTON DC, MARCH 19-23, 2018

Taking Land Governance Performance Monitoring To The Next Level: Towards a Uniform Vision For Data Standards

LISETTE MEY, **NEIL SORENSEN, LAURA MEGGIOLARO**
Land Portal Foundation, the Netherlands
lisette.mey@landportal.info

Paper prepared for presentation at the
"2018 WORLD BANK CONFERENCE ON LAND AND POVERTY"
The World Bank - Washington DC, March 19-23, 2018

Abstract

There is an increasing drive to collect data that helps illustrate the land governance situation globally. However, a uniform vision towards standardizing the use of geospatial and land governance information systems is lacking. In order to reach the full potential of land governance information systems, gathering data about specific indicators, producing community maps or making databases accessible is not enough. While this is an essential starting point, these efforts must go beyond the mere collection of data to ensure long-term sustainability on investments. The data that is gathered needs to be shared, analyzed and challenged. In this regard, the land sector can benefit from data standards on how to publish information in a way that ensures discoverability and interoperability. There is a need to agree on a uniform vision through which the land sector benefits from these data technologies, and also incorporates the controversial reality that is land governance data.

Key Words:

data monitoring, data technologies, open data, sustainable development goals

1. Introduction

Recognition that land governance is a crucial component of development is receiving increasing recognition on a global level. This is complemented by the inclusion of several land-related indicators as part of the Sustainable Development Goals (SDGs). Cadastral administration, community land mapping and enhancing land tenure frameworks represent substantial opportunities for sustainable development and economic investments. Land governance performance monitoring is a vital part of this process, not only for the SDGs, but also other initiatives such as the Voluntary Guidelines on the Responsible Governance of Tenure (VGGTs) and the Land Governance Assessment Framework (LGAF).

There is an increasing drive to collect data that helps illustrate the land governance situation in countries all over the world, which provides the land sector with a unique opportunity to monitor land governance globally and work together across all sectors - governments, academia, practitioners - to make the improvement of land governance a global and national priority. However, a uniform vision towards standardizing the publishing and use of land governance data is lacking. While local and national governments as well as international institutions and organizations invest heavily in geospatial systems, community land mapping and land tenure information systems, among other initiatives, surprisingly little thought is given to the interoperability of these platforms and systems, and harmonizing their employment and use, which would enable greater visibility and reduce duplication, while ensuring the responsible use of financial resources and promoting transparency.

With this growing momentum and increasing drive to collect and publish data around land governance, it is of vital importance that this is done in a coherent way. This paper aims to highlight the lack of coherence and harmonization in the many different data collection efforts, and provide recommendations for the land sector to publish and use land governance data in a way that ensures it can be used, reused and built upon, and ultimately - enable positive change to land governance systems and secure land rights for vulnerable people.

2. Using Data Momentum To Assess Our Data Sharing Practices

An enormous amount of data is being collected every day across the world. Over a million satellite images from the earth's surface are taken on a daily basis, showing land cover and its changes over time.[1] Google Scholar lists over 43,000 research articles and books that relate to "land rights", published in 2017 alone.[2] D-Portal - an open aid monitoring portal - records over 6,000 development projects with a relation to land globally, with a cumulative worth of land rights-related projects of well over 700 million USD in

[1] "Every day, this Satellite Company Takes A Snapshot Of The Entire Planet" *Fast Company*, consulted d.d. February 7th 2018. https://www.fastcompany.com/40498033/every-day-this-satellite-company-takes-a-snapshot-of-the-entire-planet,
[2] Google Scholar, consulted d.d. February 7th 2018. https://scholar.google.nl/scholar?as_ylo=2017&q=land+rights&hl=nl&as_sdt=0,5.

2017 alone.[3] And that accounts for *open* aid-related data. One can only imagine the sheer volume of knowledge acquired and data produced with such numbers in only one year's time.

Considering the unprecedented momentum of data collection and monitoring in light of the SDGs, this body of data is going to grow exponentially in the upcoming years. We need to ensure that this data is not collected for the sake of the momentum, but so that the data will benefit the right people and used in the context where it can be most valuable. In the current hype and enthusiasm around data collection and monitoring, we should take a moment to step back and assess: for what purpose are we collecting data and for whom?

2.1 Look beyond your usual networks

The ultimate aim we are all working towards is securing tenure rights for all: farmers, urban settlers, women, indigenous peoples, pastoralists, forest dwellers - anyone. The data and information we are gathering - whether it is within academic circles, within civil society or other networks - is gathered for that purpose. The data we collaboratively collect every day is of a different value in each different context. Raw data such as the millions of photos by satellites might mean very little to a local farmer, but should be analyzed and digested by specialists to be able to inexplicably conclude land cover changes, or even identify community settlements over time. Advocacy institutions need stories from the ground to use as convincing messages to put the land rights issue under global attention and political agendas. Sometimes, information should reach farmers in the field, so they have the right information to be empowered against outside threats.

The land sector's task is to make sure data is delivered in the right hands and in the right context, where it can be useful. Even in this digital age where we can share anything with anyone on the web, we still seem to share the data mostly within our usual networks and the stakeholders that we are used to deal with. The fact that the internet *also* contains a lot of false information from incredible sources, does not mean we should deny it as an extremely powerful dissemination tool for valuable and credible information. Sharing information needs to happen outside the usual silos to enable more people to learn from the data and bring their perspectives. Data-sharing should not be restricted to academic circles, or only shared within groups that speak the same language, or only in specific urban or rural networks. Insecurity of tenure is not an issue that is purely academic, or only applicable in a certain region or country, and it influences many different disciplines. This means that collaboration is a necessity, not a luxury.

2.2 Consider user of information

When we share information on the internet, we need to ensure it is available and valuable for its users. Users of the data need to have the permission of the publisher to use the data (through an open license), but also have a guarantee of the quality and provenance of the information. To provide that, every dataset, every publication and every single piece of data needs to be published with solid and consistent metadata.

[3] D-Portal (704,635,478 USD), consulted d.d. February 7th 2018: http://www.d-portal.org/ctrack.html?search=land%20RIGHTS#view=sectors.

2.2.1 Licensing

The current information landscape is far from ideal to accommodate the access to, and use and re-use of land data. Peer-reviewed publications are still dominated by the academic publishers that keep the publications behind payment barriers, that often restrict even universities based in the global South from accessing them. If they do publish the articles and publications based on data gathering efforts, too often, the full, raw data is not published at all. A lot of data still is not publicly accessible.

Even the information that are available online, openly accessible without payment or log in barriers, that information is often not licensed openly. We will take the Land Portal website as a (limited) reference for open access information. As of February 15th 2018, we count 46.357 publications and media items in the Library. Only 2,144 of those items have some form of *open* license that allow users to re-use and build on the information. This is a meager 4% of the total of open access publications that are available on the Land Portal's Library. The All Rights Reserved license, which means that any and all use is restricted and requires the explicit permission of the publisher every single time, accounts for 15,513 of the open access publications, 34% of the total. The remaining publications (62%) do not have any license specified at all - which in effect means the same as an All Rights Reserved license. In any case, that strikingly high number of information without any license at all, proves that this is not an element that is high on the list of information providers to consider when they make their information available.

2.2.2 Quality & Provenance

In order to build and maintain trust in information, it is necessary to provide the user with as much information as they need to be able to determine whether this piece of information is of high enough quality and a reliable enough a source for them to use it. Knowing that data is from a reliable source and that is was collected using reliable methods (or via methods with known constraints) is extremely important.[4] Concretely, this means that when sharing information, we need to consider who uses it and what might be of value for them to know and make sure this is accessible for them to access. Providing this valuable background information that allow a user to judge the quality and provenance of the resource in the metadata is extremely important.

There are many standards relating to the minimal metadata model and fields for all the different data types that are produced. For bibliographic information, this is the Dublin Core Metadata model. However, in its many years as a knowledge broker in the land sector, the Land Portal Foundation has found that the information shared in the land sector are much below par in adopting these standards. Many information providers simply do not structurally apply a metadata model when they publish information. When information providers do have a metadata model, the model often not cover these standard core fields that are required for a user to determine the quality and provenance of a source. Precisely how many times an unfamiliar acronym is indicated as the source of a piece of information, an acronym that is used by hundreds of different institutions on the web, we cannot say. But it is definitely not an exception where

[4] GODAN Action, Open Data Management in Agriculture and Nutrition, Using Open Data, Quality and Provenance, p. 5.

the Land Portal's data curation team, was simply unable to retrieve which person or institution is behind a certain type of data. Not having this kind of information, has a very simple effect on a user: this dataset or publication will not be used.

3. Working Towards Common Data Standards and a Shared Information Ecosystem

The previous chapter established that the way we publish data and information in the land sector is far from where it needs to be. Information remains within siloes, with too little consideration for licensing and there is no uniform vision on data standards and providing a user with the tools and information they need to determine quality and provenance of a piece of information. A lot can be improved. And this international momentum is the perfect opportunity for the land sector to join hands and establish a collaborative approach towards land governance monitoring and sharing.

3.1 Adopting Common Data Models

A relatively easy first step to this uniform vision is to agree that everyone publishes their data and makes it available on the web, and ensure that every dataset, every publication and every other kind of contribution to the web, has at least the core agreed metadata fields that users need to publish data - a core metadata model.

Metadata is information about the data or information that you are publishing. It explains the dataset and allows a user to find important information to find the data later, to understand what the data is later, and to share the data.[5] Having a documented and identifiable set of metadata for all datasets not only provides a user with the information they need to be able to use the information, it also helps avoiding collecting the same data more than once.

As a sector, we should strive towards a discovery-level metadata for each dataset, in a form that is suitable for users to reference. The exact fields for a metadata model may differ per data type - statistical indicators or geographically referenced data will be different from bibliographic publications - but for each data type, we need to ensure that we all have the same *core* metadata model that we adhere to. There is no need to reinvent the wheel, many core metadata models have already been identified for different types of metadata, such as Dublin Core, Darwin Core, Federal Geographic Data Committee, Data Documentation Initiative, Content Standard for Digital Geospatial Metadata, and more.[6]

3.2 Making Information Discoverable to Users

If all information providers in the land sector open up their information for anyone to access and adhere to commonly agreed metadata models, we will have made significant process in comparison to the status quo. Use of information will be facilitated because users will be able to determine the quality of the resource, the reliability and will be allowed to use it according to the open license. An important

[5] GODAN Action, Open Data Management in Agriculture and Nutrition, Making Data Open, Managing Datasets, p. 8.
[6] GODAN Action, Open Data Management in Agriculture and Nutrition, Making Data Open, Managing Datasets, p. 8.

constraint that needs to be addressed, is that the user should actually be able to find the data the sector publishes.

3.2.1 Increasing discoverability through interoperability

There is a wealth of information available online. But it is, understandably, dispersed across different websites on the world wide web. Staying within usual networks and working with known stakeholders is not only an issue on the information providers' end. Users, in their turn, also tend to stick to the information providers and sources they know and trust. As an information provider, it is important to make sure that your information is indexed and easily discoverable for a user.

However, imagining that all the data sources are indexed and are easily discoverable for a user to find - it is not possible for one person to digest all the information that can be found online, not to mention the enormous body of work that is not even published on the web. In this regard, the land sector can and should benefit from technologies that are already being applied on a large scale in other sectors. Technologies that make the data machine-readable, and therefore allowing machines to help humans digest this enormous body of data that we are discussing.

Help by computers does not simply mean that information is findable through a search engine; search engines can be very advanced but also have very serious limitations. The way the search results are ordered is arranged through a complicated algorithm. One important aspect of that algorithm is that a website with more visits, generally weighs more and ends up higher in the search results. You can even pay search engines to get higher on the result list. Larger organizations with big budgets hire 'Search Engine Optimization' experts that know how to play into these algorithms and use it to their advantage. This all results in the fact that the 'big players', the larger and usually international information providers will always have greater visibility through search engines, whereas grassroots knowledge usually only gets picked up a few pages in the search results. Besides that, a search engine can only provide limited guidance on language differences, searching for synonyms and elements like that.

3.2.2. Machine readable data formats

Machine readability means several things. The metadata and data should be published in formats that machines can read, and specifically: that machines do not need specific software (such as Microsoft Excel or Adobe Acrobat) to be able to read the data. Common data formats that are non-proprietary are XML, RDF and DDL for databases.

3.2.2. Making use of machine readable standards to classify information

Another important element for machines to become smarter and more efficient to digest the wealth of information of users, is when the metadata is consistently structured by making use of standards. It is well known that a machine reads only in 0s and 1s. Where a human will be able to understand that when a resource is tagged with "Laos", or "Lao PDR", or "Laos People's Democratic Republic", or even when it is indexed with a slight typo, say "Lao Poeple's Democratic Republic" - a machine sees these as four

completely separate inputs. It does not have the same understanding as a human does that these inputs are similar. These technical constraints are easily overcome by making use of a standard - for country names this can be the ISO3 standard, where all countries are indexed according to 3-letter codes.[7] These codes are integrated in databases, and do not need to be known by users. The computer will simply associate the known names for the country, with the code - and thereby connect them all together.

Such a machine readable standard also exists for topical keywords that relate to land governance. The Land Portal Foundation facilitates the establishment and development of LandVoc, a standard vocabulary for land governance concepts. This is extremely powerful, as each land concept is tied to a machine readable code - and that machine readable code can be associated with an endless number of synonyms, translations, and relationships with other terms. For example, a concept 'slums' would be associated with 'information settlements', 'kampungs', 'townships', 'favelas', 'ghetto's and more. This gives the machine a lot of tools to be extremely smart when querying a database, and providing the user with more valuable outputs.

3.2.3. Aligning and connecting databases

If all information providers use a common data model for the different data types, use machine readable formats and standards to publish their data, the different databases that contain information about land governance can be aligned and connected. A technical infrastructure needs to be in place that allows for the flow of information to and from the database. Once that technical infrastructure is in place, the improved accessibility and discoverability of information is endless.

Once databases have this infrastructure in place and are connecting to and from each other, users do not need to visit several websites to query for the same things - but one search in one database will allow the user to search across all the other databases this one database is connected to. And it does not have to be the same database, this is possible in each database that is a part of these connections. Enabling these connections will establish a linked information ecosystem, with information that is findable, accessible and usable for every possible type of user on the internet.

4. Ensuring Data Enables Meaningful Change

Establishing this information ecosystem has incredible and immense potential, but of course it does not directly make people's land rights more secure. An important element of sharing data, is to actually learn from and to apply data. As mentioned, data is at its most valuable when it is communicated to the right people and used in the right context. When we talk about establishing an information ecosystem on the world wide web, it is important to note that internet access is not a given in the rural areas of the global South. In between the information providers and the beneficiaries of the information, are the intermediaries. They might sometimes be the information providers, but not always. These intermediaries have the incredible important task to fine-tune data delivery to a particular target audience.

[7] https://unstats.un.org/unsd/tradekb/knowledgebase/country-code.

4.1 Learning From Data & Information

Data in itself does not drive change; we can solicit the help of computers to digest the information but at the end of the day, the analytics, understanding and defining actions based upon this data - is something that humans should do. This is an important step in the data momentum: we need to ensure the data will have impact, that it will drive meaningful change.

Research in light of an academic publication or a development project should not end when an article or a report on the findings is produced - this should only be the beginning. The quota-centered approach to data collection and publishing has the very serious risk that researchers and practitioners barely scratch the surface of their data before moving onto the next publication or project output. To drive meaningful change, this culture needs to shift from simply sharing information for the sake of sharing - but actually learning from it, identifying new opportunities for research, collaboration or interventions on the ground.

4.2 Ethical and Responsible Use of Information

Data has the potential to empower new voices and approaches, but it can also expose the vulnerable and marginalized.[8] This is crucial and what the open data society sometimes tends to forget: not all information should be unequivocally open or re-used without any restrictions. When publishing and using politically or personally sensitive data, responsible data ethics are key.[9]

These challenges place an even bigger responsibility in the hands of intermediaries, who need to be aware of the sensitivity of the data they are working with, and ensure that they handle the data with the responsibility and sensitivity it requires. An element to consider is future-proofing: what might seem as unproblematic data at the moment of gathering or publishing, may turn out to be sensitive in the future. It is extremely important to be mindful of the particular vulnerabilities and circumstances of the people and communities the data covers.[10]

The Responsible Data Forum has identified key principles of responsible data reuse that always should be considered to avoid doing harm rather than good for the communities the data concerns. Key principles of which are the right to consent and the concept of privacy; privacy concern with control over information, who can access it and how it is used. The Open Data Institute has developed its Data Ethics Canvas, that helps organizations identify and manage potential data ethics considerations, including principles, policies and processes. The Land Portal Foundation encourages the users of its platform and the whole land sector - information providers, intermediaries and users - to consider these initiatives and identify the ethical risks in any information-related project they enter into.

[8] GODAN Action, Open Data Management in Agriculture and Nutrition, Open Data Principles, Ethics in Open Data Lifecycle, p. 4.
[9] Responsible Data forum. "Handbook of the Modern Development Specialist." 2016. *Responsible Data Forum*. The Engine Room. November 2017.
[10] GODAN Action, Open Data Management in Agriculture and Nutrition, Open Data Principles, Ethics in Open Data Lifecycle, p. 4.

5. Making The Most of the Data Momentum: Next steps

To make the most of the momentum around data gathering in light of the SDGs, the land sector needs to come together and take information sharing and learning to the next level. Each type of stakeholder has an important role to fulfil and make the information ecosystem a reality - and providing users of information the tools they need to ensure this data momentum impacts meaningful change and actually secures people's land rights.

Information collectors and providers: have a responsibility to make the data that they collect publicly available. To ensure ethical use and not to invoke harm on any vulnerable people, they should adopt and publish clear data management plans that address these ethical issues. The data that they do publish, should be openly licensed that allows for use and re-use without explicit permission; the information should be structured according to a common data model and the metadata should also be openly available; information providers should make use of machine readable standards when they publish their information, to allow alignment of databases and enable interoperability between databases; they should invest in the technical infrastructure to make the information ecosystem technically possible and allow for greater discoverability of their data.

Intermediaries: have an important responsibility to fine-tune delivery of data and information through means and in a way that is tailored to the target audience they want to reach. This should all be done with careful consideration of (publicly available) ethical standards when using information and delivering this to target groups.

Donors: are the stakeholders that have not been mentioned much in this article, but play an essential role. They are a major source of the funding for the information and data collection efforts - during this SDG momentum more than ever. Donors have the power to require their grantees to open up any information that they gather, donors have the power to demand robust and ethical data management plans for any initiatives that they fund. Furthermore, they should not penalize its grantees if they do publish 'bad' practices - in the end, there is a lot to learn from the things that went wrong - not only of success stories.

The Land Portal Foundation encourages every individual that works to improve land governance and secure vulnerable people's land rights, to enable the flow of information and to make sure that every piece of data that is gathered globally, is delivered in the right hands and to the right context. Let's collectively ensure that the data gathering and monitoring initiatives are not done purely for the sake of the momentum and result in publications that only a few people read - but that it all enables meaningful change and secure land rights for all.