Land Governance in an Interconnected World
Annual World Bank Conference on Land and Poverty
Washington DC, March 19-23, 2018

THE DATA IS NOT ENOUGH:
SOME HURDLES WE MUST OVERCOME IN THE DEMOCRATIZATION OF REMOTE SENSING AND GIS TECHNOLOGY

DEAN MC CORMICK
Hexagon Geospatial, South Africa
dean.mccormick@hexagongeospatial.com

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

Abstract

For many years the data collection for the census in South Africa was a manual process. Field workers used to receive paper maps to orientate themselves to their enumeration areas. This has been a very tedious and complicated way of collecting data which required extra knowledge of map interpretation.

With the improvement and democratization of technology, Statistics South Africa, the largest and arguably the most advanced national statistical office in Africa, now benefits from the Smart.Census solution.

The Smart.Census solution enables the use of imagery base maps in a web-based smart GIS application with predefined workflows that control and limit each user (including fieldworkers) to their allocated geographical areas and tasks. A mobile application, intelligent caching, data storage and backups make it possible for users, after only a limited amount of training, to have all the functionality required to do data capturing in the field without internet access.

Key Words

census, GIS, Hexagon, M.App Enterprise, statistics

Introduction

Technology is supposed to be a great equalizer. It is supposed to take the power out of the hands of the few and make it accessible to the many. It has made information available to millions, but in many cases the ability to collect, process, analyze and present this data has remained locked up with the few who have the domain and technology knowledge to understand it.

Remote Sensing and GIS are two fields that have been locked away. While the ability to use satellite imagery to analyze landcover and land use half a world away is fairly standardized and well-documented, access to the data has been restricted to only those few who can afford it. And processing the data is a technological barrier to entry: understanding the complex nature of satellite/aerial imagery capture and processing, GIS analysis, and geospatial analytics restricts the pool of available users even further, including only those who have sufficient training and education. Added to this is the complexities of interpreting the data and understanding what is being communicated. Mapping and cartography are complex studies, and clear communication – instead of propaganda or obfuscation of the information – is difficult.

We have made some vast improvements to this field. The democratization of data through generous programs like Sentinel and Copernicus makes satellite imagery and radar data freely available to everyone. The Internet of Things and explosion of sensors (drones, webcams, video feeds, connected traffic sensors, etc.) have ensured that we have access to more data than ever before.

While the data itself is freely available, the technology to process the information is not. We can perform automated change detection to see where buildings or landcover have been altered and then address those changes. While we can get free radar coverage of every point on the earth every two weeks, processing that data is still complicated. We can unwrap radar data to find minute changes in the earth's surface (down to a millimeter), but the knowledge of those algorithms and how to use them remain a black box – and there is no foreseeable end to that.

We can perform change detection on satellite imagery, but it still requires a human being to interact with the imagery and make a map. Frequently, this process can take weeks to months, and by the time the imagery is processed and a map is created, the information if out-of-date. Because making traditional maps takes so long and is so expensive, we try to make them do too much. Every map has to perform multiple purposes: land cover, land use, roadway mapping, and topography to name a few. This compromises the intent of the map, if not the accuracy.

The use of the data collected by these sensors has great potential. What we need is a way to allow the domain experts to build sophisticated, reusable algorithms that can plug in to streams of sensor data. We need a platform that allows data to be plugged in to the platform as soon as it is collected, and then have the system pull the data through the processing steps so that the analysis runs automatically and generates an updated map that not only allows the end-users to see the current state of the land, but to see the entire time series so they can understand the patterns behind the change and begin to formulate predictions. They need to see not only what was, but what is, and what can be.

This technology exists today. It is not a map, it is a smart map – a dynamic information service. By moving from the static map model, which collects data, analyzes it, and then produces a static map – whether that map be printed, digital, or web-based – to a dynamic information service described above, we can not only automate the process, but we can build job-specific and use-case-specific maps. Instead of multiple departments sharing a single map, they can all access the data and access their own specific view of the map. Because these Smart Maps are lightweight and quickly produced, they are easy to prototype. Domain experts can build the map, build in the feedback from the users, and then stand up the map for the land use departments. As new data comes in, it is fed into the system and the map updates, including all of the relevant analytics.

The final step – the real transformative step – is transparency. Even if the data is freely available, and the map is automated to include the data in real time and automatically perform the analysis, it is of no real use to the inclusive land reform unless the affected people have access to the information and can see it with their own eyes. Public portals are vital to true change in land use discrepancies and empower not just citizens, but watchdog organizations and entities to keep an eye on the ground truth, verifying that everyone is working from a level playing field and that change is happening.

It's time to stop using maps to communicate and to start using dynamic information systems that overcome the technological hurdle that keeps people from using these powerful tools to analyze spatial data.

Case study: Smart.Census in South Africa

Census

A population census is the total process of collecting, compiling, evaluating, analyzing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country. A census is the most important and costly statistical data collection exercise conducted by a national statistical office (NSO). It is usually conducted every ten years and the results provide a detailed, small geographic area, snapshot of the demographic, socio-demographic and housing status of a country. It also provides the basis for a wide range of sample surveys. In preparation for a population and housing census the entire country is divided into small areas of land, each one small enough to be handled by one interviewer during the time of the census. This is referred to as pre-enumeration census mapping or demarcation and the resultant demarcated areas are called Enumeration Areas (EAs). In some countries, such as Sweden and Austria, door-to-door canvassing to collect census data has been replaced by a registration-based census. However, in some parts of the world, such as Africa, civil registration and housing registers are not complete and current. Therefore, a conventional census is an important source of information and will remain relevant, if not critical, for many years to come. In order to conduct a door-to-door census the census cartographer needs to provide the census enumeration team with a set of unique maps covering the entire country that accurately defines the boundaries within which each interviewer (enumerator) has to work during the enumeration phase of the census. Furthermore, today's user community demands statistics to be provided within a spatial context. To facilitate this, desktop and server-based GIS solutions have become an important part of census data products and dissemination.

Smart.Census

Statistics South Africa, the largest and arguably the most advanced national statistical office in Africa, now benefits from the Smart.Census solution.

Smart.Census has been developed on Hexagon Geospatial's M.App Enterprise platform. It combines traditional GIS functionality with a powerful workflow and workforce management tool to provide a total solution that covers all the phases of a census: pre-enumeration mapping; digital enumeration (including logistics, workforce training and management, CAPI and dynamic progress reporting); as well as the dissemination of census results through dynamic Smart M.Apps.

Statistics South Africa is experiencing a number of benefits from Smart.Census.

- *Costs:* Censuses are by far the costliest statistical data collection project in a country. There is constant pressure on NSOs to improve efficiency whilst cutting the costs associated with censuses. Traditional desktop GIS software in combination with mobile GIS software is the current standard in pre-enumeration census mapping. Since the costs associated with desktop- and mobile GIS are determined by the numbers of users/licenses it is usually significant – especially in countries where large numbers of temporary GIS operators and fieldworkers are deployed to do the work. Smart.Census provides one license which allows for unlimited users - this results in a significant cost-saving. Although it does not have all the functionality of a high-end desktop- and mobile GIS, it has all the GIS functionality required to do pre-enumeration census mapping and more. Raster and vector data is not calculated/processed on the server side but on the client side. This enables fairly sophisticated GIS functionality, including vector data capturing, attributing, redlining, measuring and querying on the client side. Smart.Census therefore eliminates the need for numerous sophisticated and costly desktop- and mobile GIS licenses.

- *GIS Infrastructure:* Smart.Census implies fewer desktop GIS licenses but it does not imply the replacement of the entire current GIS Infrastructure that may exist at an NSO. Since it accesses any established spatial database server, such as Oracle Spatial or SQL Server, it can be used together with any desktop software e.g. GeoMedia or ArcGIS. This is a huge advantage to NSOs where there might already be an established GIS infrastructure and expertise on a particular platform. The status quo can either be retained with the addition of Smart Client to provide the increased software capacity required for the large temporary census workforce or some of the desktop GIS licenses that are usually no longer required can be replaced by Smart Client for Census.

- *Effective implementation of census workflows into the GIS environment*: Although desktop- and mobile GIS software has all the functionality to do census mapping the respective software (desktop GIS in particular) have limitations when it comes to the handling and managing project-specific workflows. This is a huge challenge because it implies the use of a range of (usually unrelated) tools to establish and manage workflows. This leads to problems such as variances in the interpretation of the methodology by different users and data integration issues that emerge as a direct result of using a set of disparate tools. Ultimately, it introduces unnecessary complexity

that has a negative impact on data quality whilst increasing the overall risks associated with the project. At its core, Smart.Census has highly configurable rules and powerful workflow engine. This is a key attribute of the software. This enables the Subject Matter Expert to implement census-mapping workflows and feature-level access control that is managed by a single fully integrated system where all the parties, each with associated roles, access the same database(s) through a single web server. Each user (role) has a specific set of tasks relevant to his/her role. Furthermore, access to the system is through standard internet/intranet security protocols involving a username and password. The access control system allows a supervisor to allocate a specific unit of work, referred to as a production unit (PU), to a particular user. Only the assigned user can do the work in the PU allocated to him/her and he/she can only execute the tasks relevant to his/her role at that particular part of the overall workflow. This eliminates duplication of efforts since it is impossible for two teams to work on the same PU at the same time and it is also impossible for a user to do anything other than the tasks related to his/her role with regards to the allocated PU at that particular step in the overall workflow. Once the tasks for a particular PU are completed, it is submitted and the Supervisor receives a notification. This then triggers a set of quality control/quality assurance (QC/QA) steps, after which the work is either accepted or rejected and sent back for correction.

- *Project Management:* A census-mapping project is often a huge undertaking, with a large workforce ranging from about 50 to more than 1000 persons, depending on the size of the country, the methodology, and the project timeframe. Project management is therefore a huge challenge. Smart.Census enables the GIS Manager to manage the entire census mapping project using a single tool. Since work allocation and task execution, as described in the previous paragraph, happens on the same system, the GIS Manager knows exactly who is doing what at any particular moment in time and also, importantly, the exact overall status and progress of each phase of the project. And since Smart.Census has a strong spatial component, work scheduling and progress tracking can be done using a combination of maps and tables and lists. This key attribute of Smart.Census effectively manages, if not eliminates, most of the inefficiencies common to census mapping projects. This includes: less than optimal scheduling of fieldwork that requires field teams to drive unnecessary distances between PUs; duplication of effort where the same PU is allocated to two teams or operators; obtaining of status reports from the field and

office and compiling of progress reports. A key attribute of Smart.Census is the ability to detect a problem swiftly - long before it turns into a major crisis.

- *Data quality:* Due to the fact that numerous teams are usually working simultaneously on a census-mapping project, it is a huge training and management challenge to ensure consistency in the interpretation and execution of the methodology. Smart.Census vastly improves the consistency of work, and this ultimately improves the overall quality of the data collected and processed by the respective parties in the workflow. Every user/role has a (sub-) workflow with specific tasks configured in the software as required by the project. The user can only execute the tasks associated with that particular (part of the) workflow – nothing else. The software literally guides the user through the steps. Only the software functionality required for the execution of that particular workflow is available to the user. This ensures that all users execute the tasks in the same way ensuring consistency in the execution of the overall methodology and ultimately ensuring overall data quality and consistency.

- *Training*: Since a specific user has only access to the software functionality required to execute the tasks for that particular part of the workflow, it is not a key requirement to have huge GIS skills and knowledge to do the work. This is a huge advantage in Africa where GIS skills are relatively limited. Furthermore, training can be much more focused and task-specific, resulting in significant savings in the training budget and quicker project implementation. It is also much easier to replace staff during the project because of the lower training requirements.

- *Fieldwork in remote areas*: A key requirement for census mapping software applications for fieldwork is the ability to work offline. Smart.Census uses advanced caching to enable off-line use in remote areas where data access is limited or non-existent. If required, the field data capturer can work off-line for days or even weeks. It is, however, advisable to sync with the server as often as possible because progress reports are based on data obtained from the server and it will also avoid or limit potential loss of valuable data due to malfunctioning, loss, or theft of field computers.

Case Study Conclusion

Smart.Census provides an exciting solution to census mapping in South Africa. It provides an excellent platform for easy scheduling of work, project tracking and reporting. The software is stable and robust. It is easy to use and requires minimal training compared to conventional desktop GIS. The flexibility to configure the software to ensure that each user knows exactly what he/she is supposed to do eliminates duplication of effort and potential differences in methodology interpretation. This improves data quality and consistency. The fact that no software installation is required (on devices) makes the implementation of software/application updates easy. All updates happen on the server side. Each time the user logs in any updates or changes are automatically reflected on the client side. The off-line functionality enables fieldwork to be done in remote areas. The advanced caching ensures that relatively small data packages are sent to and from the server keeping data costs and connection time to the minimum – this is important from a fieldwork and budget perspective.

The flexibility of the software enables it to be used for the planning and management of the subsequent phases of the census project. It can ultimately also be used as a dissemination tool. Smart.Census compliments any existing GIS infrastructure rather than replacing. It does however eliminate the need for mobile GIS software and it lowers the need for high-end desktop GIS software.

Conclusion

As we can see from the Smart.Census case study, the challenge facing the democratization of remote sensing and GIS workflows lies not in access to data, but in access to the processes that create the data and the information derived from the data. In this case, the democratization of the data is not achieved by free access to satellite data, but rather by empowering NSOs to scale up their collectors. By implementing a streamlined workflow system, the Smart.Census solution empowers workforces of any size to be quickly trained and to focus on their particular duties instead of navigating through a complex and perhaps overpowered GIS system.

Under the hood, of course, there is a lot of specialized processing going on. But that is ameliorated by keeping it under the hood and limited to the SMEs. They can set up the processes and workflows, and then step back and allow the data to flow into the system. They can focus on QA/QC tasks instead of building map after map after map.

In addition, the solution ensures adherence to standards, simplifies the planning process, and automates the report generation at the end of each cycle. Not only does this make it easier on the NSO, but it also makes this data more accurate and more accessible. This means that it is of higher quality, and more quickly disseminated to the stakeholders for their use in understanding the makeup of their country or region, and they can better understand what is to shape smart change in their world.