A DPRU Policy Brief*: **PB 14/42**
Commissioned for the

AN OPEN AFRICAN DATA APPROACH TO IMPROVING DATA QUALITY

Lynn Woolfrey
DataFirst Data Service, University of Cape Town

December 2014

***Disclaimer:** The Policy Brief series is intended to catalyse policy debate. Views expressed in these papers are those of their respective authors and not necessarily those of the Development Policy Research Unit, the Jobs Knowledge Platform, or any associated organisation/s.

INTRODUCTION

African governments collect primary data on their populations through regular censuses and sample surveys. They also collect primary data as administrative records in the course of governing. This primary data, or microdata, is detailed enough to give government decision-makers an accurate picture of their economies and societies (Lane, 2003: 12). Microdata can inform planning processes for more targeted, and possibly more effective, national policies. Primary data collection is advocated by Donor Organisations: The United Nation's agenda for the Post-2015 Millennium Development Goals looks towards future MDG initiatives and links microdata needs directly to development needs. The 2013 MDG report calls for "a data revolution" to open up disaggregated data as a resource for growth. The 2013 review makes the point that microdata are required for in-depth policy analysis because highly aggregated data can hide local and national inequalities (United Nations, 2013:23-24). Donor Organisations also part-fund the data collection projects of African governments. They provide technical support to African National Statistics Agencies, as the main collectors of primary data. This gives Donors access to the data they need for project initiation and monitoring. However, this type of support can also assist fact-based national planning in the region.

BARRIERS TO MICRODATA-INFORMED POLICIES

Political complexities

Collecting low level data as a planning resource makes sense. However, what the data reveals has historically not always fed directly into policy. Research by the African Economic Research Consortium (AERC) measured the impact of research on economic policymaking in seven African countries (Cameroon, Côte D'Ivoire, Ghana, Senegal, Tanzania, Uganda and Zambia). The study showed that the findings from survey research had little direct impact on decisions taken by the governments of these countries (Hoffman, J. 1995 http://dspace.africaportal.org/jspui/bitstream/123456789/32227/1/SP20.pdf?1). There are many reasons for these breaks in the collection to policy chain. Most relate to the complex nature of government planning: Policy-making is not a linear process of seeking information and utilising this information for problem solving. Rather, the many role players and scenarios involved result in a delay in the uptake of new data (Porter & Hicks, 2007:5).

Data Quality Constraints

However, data quality constraints also thwart data-intensive planning, including in many African countries. The quality of data relates to its utility (US Office of Management and Budget, 2006:9).To be optimally usable, microdata should be acceptable in a number of standard quality dimensions: Data should be *accessible*, that is, easy to find and use. Data also needs to be *relevant* to users' information needs, and *accurate,* with minimal difference between official estimates and true values. Quality data is expected to be *timely*, without long delays between its reference period and its availability. There is often a trade-off between accuracy and timeliness (Statistics Canada 2009:73). Data should be *comparable* across time and type to allow for the merging of disparate datasets (Cooper, 2005:2). Finally, data should be readily *interpretable*, which is a product of good data documentation (Statistics Canada, 2009:73). In the case of microdata, data *security* should be added to these traditional quality attributes. This in terms of protection from environmental hazards, as well as anonymised to protect respondent confidentiality.

AN OPEN AFRICAN MICRODATA APPROACH

Improving microdata accessibility may be a way to smooth the collection to policy path. Evidence suggests that data and its policy implications have to gain a certain amount of consensus to proceed to the policy stage (Porter & Hicks, 2007:16). This means that the more widely data is distributed for reanalysis and secondary analysis, the more consensus there will eventually be with regard to its value as a policy tool. An Open Data approach can therefore be an aid to fact-based governance. Open data is described in the G8 Open Data Charter as data that is available free and online, and which comes with enough information to be optimally usable. Open Microdata should fulfil the requirements in the G8 Charter. However, with microdata there is the additional requirement to anonymise the data to create Public Use Files suitable for sharing.

It follows from the above finding that the failure of African government microdata to reach a critical mass of researchers and be circulated, judged, and assimilated widely contributes to its inconsistent application for national planning. In many African countries the challenges of access are twofold: Firstly, decision-makers in government ministries are not provided with microdata that they need for planning. Secondly, large swathes of government microdata do not reach academia. National Statistics Agencies are the main producers of African microdata and currently most do not have policies for giving researchers access to their microdata. This was confirmed by surveys in 2009 and 2012 which investigated ease of access to official microdata from African countries. Findings were that research access to official African microdata was ad hoc: Requests for microdata were either declined, or handled on a case-by-case basis, with confidentiality concerns used to justify both approaches (Woolfrey, 2013).

The former stance wastes national resources: Such an approach prevents academics from using real-life data as a research and teaching tool to build skills and contribute to policy analysis. The latter position can lead to onerous requirements for data access: This is data sharing, which is not the same as Open Data. To be Open, data should be free and easy to source, obtain, and use. Some African government microdata are available for a fee. This often puts them out of reach of researchers at African research institutions. This and other restrictions on microdata access can limit the benefits of knowledge utilisation for African countries.

OPEN MICRODATA AND DATA QUALITY

Access is the cornerstone of data quality. Better access may boost other aspects of data quality because proper data curation becomes a by-product of sharing. Opening up their microdata compels official data producers to confront the challenge of how to handle their data. The digital revolution has allowed governments to produce more data more quickly. At the same time governments have failed to take advantage of the new technologies to manage this data effectively to ensure its ongoing reuse for planning. Challenges even face countries with well-established government information infrastructures: A recent report on national data in the US found that “no one knows how much born-digital U.S. Federal government information has been created or where it all is (Jacobs, 2014:3)”.

In less well-resourced countries, management of official digital data is even more incoherent: Many African countries lack the curation infrastructure to allow ongoing usage of appropriately disaggregated data for useful policy input. Curation refers to managing data for reuse (Digital Curation Centre, 2015). This is necessary for all types of data, but is essential for the handling of microdata because disaggregated data requires more careful management to prevent disclosure.

There is some evidence for the quality enhancing benefits of opening up microdata. The 2009 survey was combined with an assessment of a donor-funded technical support programme, the Accelerated Data Program (ADP). The ADP provides microdata management tools and training to NSAs. The survey examined the success of the project at promoting access to African government microdata. A follow-up to the 2009 Microdata Access survey was undertaken in 2012. The 2009 study judged microdata access in African countries. The 2012 research examined the ADP's impact on *all* quality aspects of official microdata. Can establishing mechanisms for data sharing help NSAs produce more accurate and usable data?

Microdata Preservation

Open data practices mean data must be preserved on an on-going basis. Few African NSAs have budgets for ongoing stewardship of microdata. The consequence is the loss of long-run microdata series that could have revealed much about their economies. Even where microdata are held, this is seldom curated methodically. Accelerated Data Program (ADP) training inevitably involves initial searches for microdata files and documentation across NSA departments. This shows that long-term microdata curation is not prioritised at these NSAs. The Project trains metadata teams to prepare their data for long-term curation. However, the 2012 survey revealed that there has been growing awareness of preservation issues among participating African NSAs, although this hasn't necessarily translated into access (Woolfrey, 2013).

Microdata Discovery

Open data also requires data discovery mechanisms. ADP support seems to have aided microdata discovery. NSA websites would likely be the first port of call for researchers wishing to access African government microdata. Research undertaken in 2006 showed that African NSAs did not provide information on their websites about the microdata they held (Woolfrey, 2010). By 2012, most participating NSAs had microdata descriptions (metadata) on portals linked to their homepages. The data descriptions were also being linked to by other useful discovery tools such as the World Bank's Central Microdata Catalog. Ease of discovery encourages data demand, which can also improve access. This has been the case in South Africa, where quantitative research at universities has increased as a result of access to government microdata. At the same time official data producers have adapted to increased demand by improving on access systems (Woolfrey, 2010).

Microdata Comparability, Relevance, and Timeliness

With regard to data comparability, ADP NSAs use uniform software and metadata and microdata standards which are expected to support comparability across regional datasets. The surveys in 2009 and 2012 exposed the weak communication channels between producers of official data and policy researchers. ADP training experience reveals that government functionaries also struggle to obtain official microdata. This situation can make official statistics supply- rather than demand-driven, which affects both their relevance and timeliness. The 2012 survey showed that only six of the 23 NSAs participating in the ADP had recent datasets listed on their site (less than two years old). Data timeliness is also affected by resource shortages at official data producers. In many African countries funding constraints mean surveys are conducted irregularly and periods between population censuses can be unacceptably long (Kiregyera, 2005: 70-72).

Can a more Open Data approach make official data products more timely and relevant? The history of microdata access in South Africa can provide some insight here: Interaction with researchers over 10 years of microdata provision through the Data Service at the University of Cape Town shows that greater access has encouraged user feedback and led to official data producers being more aware of their accountability for data turn-around times. They have also tailored their products to be more relevant to policy researchers: One example is the provision of more detailed geography with South African census microdata.

Accuracy

Data accuracy refers to how well the data represents reality. Disseminating government data at an appropriately disaggregated level can lead to better statistics, as aggregation can obscure poor data quality (Palangkaraya, 2012: 4). Providing researchers with detailed data exposes government statistics to more scrutiny. But it also allows government data producers to crowd-source among academics for data quality input. This can improve their collection and collation methods and initiate a virtuous cycle of microdata use and feedback to producers, leading to more accurate official data.

Interpretability

Open data, is, of course, accessible, but to be truly "Open" the data needs also to be easy to decipher. This is a product of good data documentation (US Census Bureau, 2013:ii). Inaccurate or inappropriate usage of microdata can also be mitigated by the provision of usage information. The tools provided with ADP training allow metadata and key documents to be shared with the data. The 2012 Survey revealed that all participating NSAs with online portals were sharing information on their data. Often these records had not previously been available online. Users of South African microdata available online have provided input which has improved the documentation produced by the NSA. The implication here is that good and readily available data documentation can sometimes follow data sharing, rather than precede it.

OVERCOMING BARRIERS TO OPEN AFRICAN MICRODATA

Policy Support

African Planning Ministries and Statistics Directors can work with donor organisations to develop strategies for better curation of microdata at NSAs and other government institutions. This will allow them to take advantage of the virtuous cycle of reuse. The 2009 and 2012 African Microdata Access Surveys showed that unless there was prior Ministerial-level commitment to a more open data approach, training and tools were not effective in enabling broad access (Woolfrey 2013). The absence of policies for sharing can amount to implicit policies not to share. Funding will then not made available for long-term preservation and dissemination of national microdata (Kiregyera, 2005: 70-72; Woolfrey, 2013). Ministerial permissions to share could release funding and ensure any data that is shared is done securely through official channels using appropriate platforms.

Donor organisations can take this into account when funding training initiatives at NSAs. At NSA's not yet subscribing to Open Data, it may be more effective to work towards functional systems and staff

competencies for in-house curation of micro-datasets. Systematic, quality-centred microdata handling is vital for producing good aggregated data products, whether or not the microdata is shared. Advocacy for a more open agenda can form part of any donor support, using Open Data success stories from the region. NSAs which are already committed to this approach can be supported with tools and training to actualise their Open Data agendas.

Dedicated Curation Units

Many of the software tools for curation are Open Source or free. Donor agencies provide free training as a component of statistical capacity building programmes. However, these have not been effective in establishing long-term microdata curation best practice at African NSAs (Woolfrey, 2013). There may be a need for a more coordinated approach to training data managers. This is terms of donor coordination and a more consistent approach by NSAs as to who receives this training. There is little coordination among donor bodies in promoting their respective tools. These include a range of platforms to disseminate aggregated data products. The result is a proliferation of aggregated data tools at NSAs. The quality of the underlying data fed into these systems may be downplayed and technical support for these systems may end with donor project completion dates. This adds to the general confusion at these institutions concerning curating data for reuse.

NSAs also do not maximise the benefits of donor funded training: A common approach is to have each department represented at training workshops. Staff already have allotted tasks and are too overworked to implement the new knowledge going forward. Alternatively microdata management training is given mainly to IT departments, as the task becomes more technology-driven. As a result, statistical knowledge does not always inform data management decisions. A solution may be for African NSAs to establish dedicated microdata curation units within their User Services or Dissemination Departments. These units could be staffed by a tech-savvy junior statisticians or information professionals. They would be responsible for liaison with IT managers for systems support, and with senior statisticians on statistical issues such as disclosure control. Curation departments could be the focus of microdata management training instead of diffusing this training across departments to staff who may not ultimately be involved in curation efforts.

CONCLUSION

A more Open Data approach from official data producers in African countries could ensure government departments get the microdata they need for decision-making. This stance could also ensure microdata are available to policy analysts in academia to inform national planning. Opening up African microdata may lead to other data quality improvements. Open Microdata needs to be well-managed and securely shared. Statistical Directors can leverage donor funded data support projects to enable safe sharing of official microdata. This has the potential to improve the evidence to policy chain in African countries. More customised technology transfer, working with dedicated curation staff, may see better uptake of the resources provided. A move towards more Open Data in the region can be encouraged by Open African Data success stories. Finally, experience amassed through monitoring curation support projects could drive more informed donor funded Open Data initiatives in the future.

REFERENCES

Cooper, A. 2005. A proposed methodology and infrastructure for standards development and implementation within a national statistical agency. The Fourteenth Conference of Commonwealth Statisticians, Cape Town, 5-9 September, 2005. [Online]. Accessed 20150127 from http://www.statssa.gov.za/commonwealth/presentations/A_Cooper.pdf

Digital Curation Centre [website]. What is digital curation? [online]. Accessed 20150127 from http://www.dcc.ac.uk/digital-curation/what-digital-curation

Hoffman, J. 1995. Linking economic research and policymaking: the role of the AERC. Nairobi: African Economic Research Consortium [Online]. Accessed 20150127 from http://dspace.africaportal.org/jspui/bitstream/123456789/32227/1/SP20.pdf?1

Jacobs, J.A. Born-digital U.S. 2014. Federal Government information: Preservation and access. Report for the Center for Research Libraries Global Resources Collections Forum: Leviathan: Libraries and government information in the age of big data, March 17 [online]. Accessed 20150127 from http://www.crl.edu/sites/default/files/attachments/pages/Leviathan%20Jacobs%20Report%20CRL%20%C6%92%20%283%29.pdf

Kiregyera, B. 2005. A case and some actions for improving statistical advocacy in poor developing countries. African Statistical Journal 1: 70-84 [online]. Accessed 20150130 from http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/African.Statistical.Journal_Vol1_3.Articles_5.Case.SomeActions.pdf

Lane, J. 2003. Uses of microdata: keynote speech. In Statistical confidentiality and access to microdata: proceedings of the Seminar Session of the 2003 Conference of European Statisticians, Geneva, 10-12 June 2003, 11-20. New York: United Nations Economic Commission for Europe. [Online]. Accessed 20150114 from http://www.unece.org/stats/publications/statistical.confidentiality.pdf

Palangkaraya, Alfons. 2012. Evidence-based policy: Data needed for robust evaluation of industry policies. Melbourne Institute of Applied Economic and Social Research, University of Melbourne [online]. Accessed 20150128 from https://www.ipria.org/publications/reports/Evidence%20based%20policy%20-%20Data%20needed%20for%20robust%20evaluation%20of%20industry%20policies.pdf

Porter, R.W. and Hicks, I. 2007. Knowledge utilization and the process of policy formation: towards a framework for Africa. Washington: Academy for Educational Development. [Online]. Available: http://sara.aed.org/publications/cross_cutting/knowledge_utilization/html/utilization.htm

Statistics Canada Quality Guidelines. Fifth Edition. 2009. Ottawa: Statistics Canada (Catalogue no. 12-539-X) [online]. Accessed 20150127 from http://unstats.un.org/unsd/dnss/docs-nqaf/Canada-12-539-x2009001-eng.pdf

United Kingdom Cabinet Office. 2013. G8 Open Data Charter and Technical Annex [Cabinet Office Policy Paper) [online]. Accessed 20150130 from

https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

United Nations High-Level Panel of Eminent Persons on the Post-2015 Development Agenda. 2013. A new global partnership: Eradicate poverty and transform economies through sustainable development [Online]. Accessed 20150114 from: http://www.un.org/sg/management/pdf/HLP_P2015_Report.pdf

United States Census Bureau. 2013. U.S. Census Bureau Statistical Quality Standards [online]. Accessed 20150128 from: https://www.census.gov/quality/standards/Quality_Standards.pdf

United States Office of Management and Budget. 2006. Standards and guidelines for statistical surveys [online]. Accessed 20150127 from: http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Woolfrey. H.L. 2010. Archiving social survey data in Africa: An overview of African microdata curation and the role of survey data archives in data management in Africa, Unpublished MPhil Thesis, University of Cape Town.

Woolfrey, H.L. 2013. Leveraging data in African countries: Curating government microdata for research. Cape Town: DataFirst (Technical Paper 22) [online]. Accessed 20150127 from: http://www.datafirst.uct.ac.za/images/docs/DataFirst-TP13_22.pdf