

# DataFirst Technical Papers



Leveraging data in African countries: Curating government  
microdata for research

*by*  
*Lynn Woolfrey*

---

Technical Paper Series  
Number 22 Version 2

## About the Author(s) and Acknowledgments

Lynn Woolfrey - Manager DataFirst, University of Cape Town. [lynn.woolfrey@uct.ac.za](mailto:lynn.woolfrey@uct.ac.za)

## Recommended citation

Woolfrey, L., (2013). Leveraging data in African countries: Curating government microdata for research. A DataFirst Technical Paper Number 22, Version 2. Cape Town: DataFirst, University of Cape Town

---

© DataFirst, UCT, 2013

DataFirst, University of Cape Town, Private Bag, Rondebosch, 7701, Tel: (021) 650 5708,  
Email: [info@data1st.org](mailto:info@data1st.org)/[support@data1st.org](mailto:support@data1st.org)

# Data Quality Project Working Paper

## Leveraging data in African countries: Curating government microdata for research

Lynn Woolfrey,  
DataFirst, University of Cape Town

### Abstract

Governments mandate their National Statistics Offices to collect empirical data through censuses and surveys to determine appropriate national policies. Further investigation of this data by academics can help evaluate the effectiveness of government action. In many countries governments have created policies and institutions to share official data with researchers to allow independent evaluation of both the data and the policies informed by the data. In 2009 a survey was undertaken to assess ease of access to government survey microdata in African countries. This investigation revealed that the sharing of African government microdata is constrained by several obstacles. African National Statistics Offices, the main African data producers, have limited resources to curate microdata and ensure its long-term availability. Consequently many African data producers do not follow international best practice with regard to survey data management or make the microdata from official surveys available for research purposes. A further obstacle in Africa is inadequate producer-user communication channels. Concerns around the confidentiality of respondent information also present a barrier to data usage for research, as does the bureaucratic nature of government institutions involved in data production. A follow-up survey undertaken in 2012 examined improvements in data access and other dimensions of data quality in the intervening years, focusing on Statistics Offices participating in a donor project to advance data curation in resource poor countries. The survey showed that the provision of appropriate data curation tools can improve data sharing for policy feedback. However, this is more effective in countries with sound data usage policies driven by African decision-makers who appreciate the role of data utilisation in national development.

## 1. Reasons to repurpose Government Microdata

The collection of national statistics via government censuses and surveys to inform government planning can be seen as a vital component of governance. The provision of this national survey microdata to researchers can be beneficial for governments. Re-use of this data provides independent feedback on the effectiveness of government action. Re-analysis of quantitative data can prevent inappropriate policy decisions based on inaccurate research, and serve to improve the survey research methods of official data producers.

The public's trust in official data will be increased if they perceive academic research based on government statistics as an extension of official information. Providing access to government data can also be viewed as essential to promote more people-centred government, as further research on official data serves to contribute to the findings of government statisticians concerning social policies to improve lives.

The open exchange of data within a research-policy interface has come to be viewed by policymakers as a pre-requisite for a successful national research and innovation system. That is, government expenditure on survey research and the preservation of survey data for reuse is thought to build national research capacity and provide the material for high-quality research output. This research in turn is expected to lead to the technological and social innovation necessary for economic growth

### **2. Curating African Government Microdata for Re-use**

In some countries official information policies have been formulated to encourage the efficient dissemination of official data by government agencies, not only in aggregated form, but at a unit-record level required for in-depth analysis (microdata). In these countries National Statistics Offices (NSOs) and research organisations provide the institutional capacity for data curation. For example, in Europe a network of national Survey Data Archives (SDAs) form part of the European research support infrastructure (CESSDA ESFRI Project, 2008). SDAs provide the advocacy, institutional links and skilled staff to facilitate the use of government microdata by researchers in Europe. In Canada the national statistics office has established research data centres at universities, which provide controlled environments for the reuse of official data (About the CRDCN, 2011).

In most African countries, however, the sharing of microdata from government censuses and sample surveys is constrained by several obstacles. There has been a substantial amount of African microdata collected over the years, predominantly by African National Statistics Offices. However, these organisations have limited resources to curate the microdata to ensure its long-term availability for research. Consequently many African data producers do not follow international best practice with regard to data curation, or share the microdata from the surveys they conduct (Dupriez, 2008).

However, with growing emphasis on the importance of statistical data as a national resource for scientific investigation to support innovation and sound national decision-making, some African leaders have begun to support the preservation of microdata and its reuse by academics. Evidence-based policymaking is seen as an imperative to support political legitimacy. A vibrant research policy interface, however, is increasingly seen to depend on access to original microdata files for researchers, as opposed to the aggregated data generally made available to the public.

The needs of international development organisations have begun to change data management practices in African countries. These agencies require country-level data to monitor their development projects in the region. Development agencies, in partnership with African governments and regional organisations, have, over the years provided funding to African National Statistics Offices for the production and dissemination of national statistics (Woolfrey, 2010). Recently, these donor agencies have begun to provide funding support for microdata repurposing for further research. While arising from the data needs of development organisations, this support can assist African governments to improve their data resources for better national planning.

---

### 3. African National Statistics Offices as Data Curating Institutions

In African countries NSOs are the key data curation agencies within National Statistics Systems and are responsible for coordination of data collection and data management within these systems. They form part of the country's civil service, either as autonomous government departments, or as departments under a Ministry such as the Ministry of Planning. These institutions are mandated to collect economic, demographic and social statistics to support government planning. However, traditionally there has been no culture of data archiving or data sharing at government institutions in Africa.

Despite government lip-service to the value of evidence-based policymaking, and official claims regarding commitment to harnessing empirical data for economic growth, in many African countries scant government funding is allocated to NSOs. NSOs in most countries of the region are chronically underfunded and suffer from shortages of basic equipment such as computers and vehicles. Government and donor expenditure is allocated mainly for data collection, and very little funding is provided for the long-term preservation and sharing of national data. Skills shortages and high staff turnover due to low salaries in the public sector also result in a paucity of data curation expertise in these institutions. African NSS are decentralised, with data being produced by a number of government agencies. No systematic inventory of data available across institutions in NSS is undertaken, which hampers data discovery by researchers. The data co-ordination role mandated for NSOs in African countries is seldom accomplished, as these agencies have barely enough human and financial resources for their primary task of producing national statistics for government ministries (Lufumpa & Mouyelo-Katoula, 2005:31-32; Woolfrey, 2010).

### 4. Support for the Curation of African Microdata

Data curation systems are needed at African NSOs to ensure efficient data preservation and dissemination and the creation of data documentation to support the usage of official data. International donor organisations have recently begun to focus on the provision of software and guides to improve the management of national data in African countries. One such project has been undertaken by the International Household Survey Network (IHSN). The IHSN was established in 2004 to improve data quality and data usage in developing countries. Its membership is comprised of organisations that provide funding and technical support for household survey programmes, and includes, among others, PARIS21, UNSTATS and the World Bank (Members and partners, 2011). The IHSN is based at the World Bank. The network has developed Open Source data curation tools including the National Data Archive (NADA) software, which is a web application that allows the creation of searchable online data portals for microdata preservation and dissemination. In 2006 the IHSN's Accelerated Data Program (ADP) <http://adp.ihsn.org/> began distributing the microdata portal software to African NSOs and providing data curation training to NSO staff. The portal software was bundled with metadata creation software provided by

NESSTAR <http://www.nesstar.com/> and both tools are available free on the IHSN site <http://www.ihsn.org/home/software/ddi-metadata-editor>. Since 2008 DataFirst, a data service at the University of Cape Town, South Africa, has assisted the IHSN to develop their data curation software further and undertake software installations and training at African NSOs.<sup>1</sup>

The focus of the IHSN work is well-curated data to assist governments with fact-based planning, and for the World Bank's project initiation and monitoring purposes. DataFirst's interest in this is as a conduit to better access to African government data for researchers. The aforementioned benefits will only accrue to African countries if governments share their data with the researcher and donor communities. A survey was therefore undertaken to investigate whether government data producers in these countries are prepared to make the unit record data from national surveys available to researchers.

## 5. Survey on Access to African Government Microdata<sup>2</sup>

### 5.1 Survey Purpose

The survey aimed to investigate how easy it is for researchers to find and obtain government microdata in African countries. The focus was on access from NSOs, as the main data producers in African National Statistical Systems. As previously discussed, the benefits of corrections and extensions to government survey research findings depend on researchers being allowed access to the original microdata. This allows them to obtain a more accurate economic view of their country than that derived from research using aggregated data (Lane, 2003: 12). Access to unit record data, rather than the aggregated data provided in reports or online interactive information portals was thus the focus of this survey.

The process of applying for data also tested the effectiveness of data user-producer communication channels at these NSOs. Historically, little interaction has taken place between official data producers in Africa and those interested in using their data. This has impacted on the *discoverability* of African data, which is an important measure of data quality, along with *accessibility*. Regular communication between data producers and user groups ensures data users know what data exists and what constitutes appropriate usage of this data. The nature of data user-producer communication also has implications for the *relevance* of African data, which is another core component of data quality. Without feedback from the data user community, it is difficult for data producers to provide statistics that will fulfil existing data needs.

---

<sup>1</sup> In 2008 the author participated in data portal installations and training at the Lesotho Bureau of Statistics and the Mozambican Institute of National Statistics, and has undertaken training and policy development work in the field of data curation at African NSOs, under an OECD contract.

<sup>2</sup> The author presented the results of this survey as a paper entitled "African microdata access survey 2009" at the 57<sup>th</sup> Session of the International Statistical Institute, held in Durban, 16 – 22 August, 2009.

---

## 5.2 Survey Method

The 2009 microdata access survey used contact points available on the websites of the NSOs in the study. As the main public channel of communication for national data producers worldwide, websites were the obvious point of contact for engagement on data.

During March and April of 2009, a request for survey data was sent to email addresses obtained from the websites of NSOs in all fifty-three African countries, in the official language of the country to avoid a bias in the responses. A specific household survey was identified, where this detail was available from NSO websites. Otherwise a query regarding the availability of microdata from any household survey was submitted. No recent datasets were requested, as it was assumed these would be more difficult to obtain. At the time of the survey NSOs in ten African countries were using the IHSN's NADA software to curate their data. These were NSOs in Ethiopia, Gambia, Ghana, Lesotho, Liberia, Niger, Nigeria, Senegal, Sierra Leone and Uganda. Online microdata request forms provided on their online data portals were completed and submitted for these NSOs.

## 5.3 Survey Limitations

Internet access is a vital component of data sharing, and the availability of websites and functioning email contact points at NSOs indicates a certain minimum technological level maintained by these institutions. Website and email information and email responses are essential components of the survey in order to assess the accessibility and data discovery/user communication dimensions of quality management at African NSOs with regard to data sharing. For this reason, the original plan was to use only email contact points or web-based forms for this survey. However, problems with outdated website and email details became an obstacle to communication. This resulted in a decision to re-submit the data requests by fax if no response was received to the original data request submitted by email or via online data request forms. This was to take into account poor bandwidth and other infrastructure problems of some African countries. A further justification for employing other communication channels was that the survey was also an attempt to measure the *willingness* of African data producers to liaise with potential data users and to share their data.

Bureaucratic resistance to placing data in the public domain can be an obstacle to effective data exchange on the continent, and this element needed to be explored by circumventing email and internet problems wherever possible. This approach was hampered to some extent by the fact that many of the fax numbers provided were out of date or did not function. In the end contact details from the United Nations Economic Commission for Africa (UNECA) website were used in a further attempt to contact the NSOs. The strategy did not assist the survey because the email and fax details from this source did not elicit many more responses and it would seem that the information on this website was outdated (the NSO list was subsequently removed from the UNECA site).

A further limitation in the study was the decision not to use telephone contact points to request data. A small number of African NSOs did not have

email or fax contact details on their websites, only telephone numbers. These however, were not used as primary contact points in the study. It is probable that a better response rate would have resulted from the inclusion of telephone inquiries to NSOs, not necessarily in terms of data acquisition, but with regard to obtaining information on data access policies. However, it was decided to use telephone contacts only to confirm fax and email details, and not to request data. This decision was partly based on time constraints, but also on the belief that email and fax communication should suffice as contact points for data user-producer communication. Given that many researchers interested in African survey data would be based at foreign institutions, expecting the potential data user to include costly and time-consuming international telephone call to NSOs would make this method of data acquisition cumbersome. However, telephone points were used to verify fax details, in an attempt to limit technological constraints as a cause of non-response. This was not entirely successful as telephone lines sometimes did not work, which meant not all email and fax details could be confirmed in this manner.

Although the author works for a data service in Africa, and has met senior staff at several African NSOs, all attempts were made to avoid using personal contacts, so as not to bias the study. The aim of the research was to ascertain the extent of availability of African government microdata for researchers generally, and using personal contacts may have led to the circumvention of restrictions put in place for the anonymous researcher.<sup>3</sup>

### ***5.4 Survey Results***

The survey assessed microdata access from NSOs in 53 African countries. The results are summarized below and included in Figure 1.

#### *Category 1: No website access (8):*

NSOs in 8 African countries did not have websites or these could not be accessed (Angola, Comoros, Democratic Republic of Congo, Eritrea, Libya, Mali, Somalia and Zimbabwe). Fax and email details for all NSOs except the Ministry of National Planning in Somalia were found on the UN Economic Commission for Africa's website, but these yielded no response, possibly because they were out of date (the details for heads of NSOs were recorded as from 2003).

#### *Category 2: No contact (13)*

Contact problems with emails bouncing and/or fax numbers not working meant that 13 African NSOs were uncontactable (Benin, Burkina Faso, Cape Verde, Central African Republic, Congo, Djibouti, Guinea, Guinea-Bissau, Mauritania, Rwanda, Seychelles, Swaziland and Sudan)

---

<sup>3</sup> As it turned out, the researcher inadvertently contacted someone she knew when emailing the Kenya National Bureau of Statistics (KNBS). Although the request was subsequently dealt with by other staff at the KNBS, this may have played a role in the eventual acquisition of microdata from the Kenyan NSO.



---

*Category 3: No reply (18)*

In 19 cases NSOs were contacted by email or by email and fax but these requests elicited no response. These included NSOs in Algeria, Burundi, Cameroon, Chad, Côte d'Ivoire, Equatorial Guinea, Ethiopia, Gabon, Gambia, Lesotho, Liberia, Madagascar, Morocco, Niger, Sierra Leone, Tanzania, Togo, Tunisia and Zambia).

*Category 4: No policies (4)*

In 4 cases contact yielded information that the NSOs had no policies in place to share microdata for research. The Egyptian NSO only makes aggregated data available. The Mozambican Instituto Nacional de Estatística (INE) had no microdata access policy, although they have systems in place for data sharing (the NADA software was installed at the INE in 2009). The Instituto Nacional de Estatística in São Tomé e Príncipe also did not have a data sharing policy (São Tomé). The author met the President of the Mozambican INE at a conference in August 2009 who explained that the INE have a policy to make their datasets available to researchers through the Mozambican Department of Science and Technology (Loureiro, 2009). However, this information was not on INE's website or mentioned by staff at the INE during the survey. Links on the Department of Science and Technology's website where researchers can register (presumably for data access) were broken (Register researchers, 2011). A meeting with the Director of the Statistics Department of the Moroccan High Commissioner's Office for Planning (HCP) at the same conference elicited further information on data sharing at the Moroccan NSO. Mr Taamouti explained (and later confirmed by email) that the NSO did not have a microdata dissemination policy and data requests are treated on a case-by-case basis (Taamouti, 2009).

*Category 5: No follow-up (3)*

Staff of 3 African NSOs initially responded to data requests but lack of follow-up procedures stymied data access (Malawi, Namibia and Uganda). Staff at the National Statistical Office of Malawi (SOM) agreed in principle to supply the data, but then ran into complications regarding a means of sending the data files, which were fairly large. Staff at the Namibia Central Bureau of Statistics initially responded with requests for further details, but then did not follow up. The Uganda Bureau of Statistics staff responded in a timely manner but provided the information that the selected dataset would need to be obtained from the Office of the Presidency. Requests for a different dataset were not followed up. Thus while data discovery tools are in place at this NSO (they have a data portal) and the staff are willing to assist, they do not have established and automatic procedures for microdata provision.

*Category 6: Onerous access requirements (1)*

The Mauritian Central Statistics Office required payment of 4000 Rupees (\$130) to obtain the requested data. The NSO also required and a visit to the

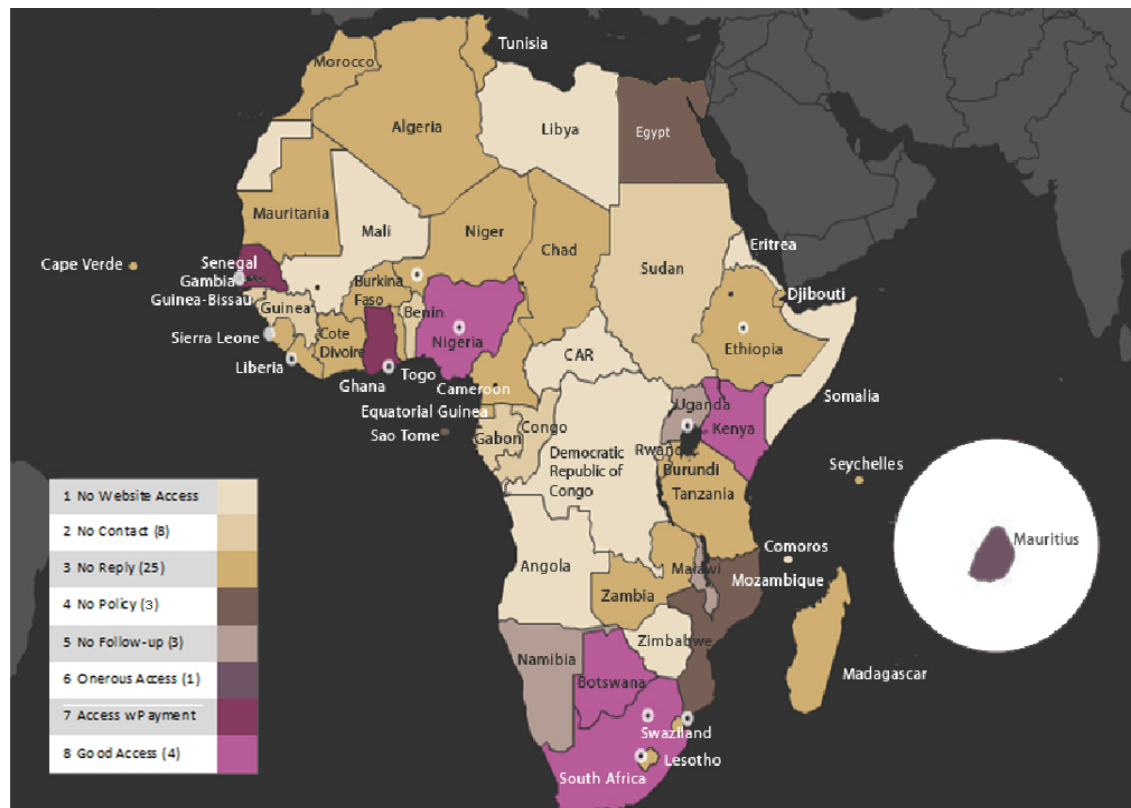
country to swear an oath of confidentiality before the Director of Statistics, or the appointment of an in-country representative to undertake this on the researcher's behalf. While the data request was handled in an efficient manner these onerous conditions hamper access to Mauritian data for researchers.

### *Category 7: Data provision with payment (2)*

NSOs of a further 2 countries charged for data access (Ghana and Senegal). The Ghana Statistical Service required payment to obtain microdata files. The amount was \$320 for the microdata from three surveys. This is not costly for the amount of data provided, but would be an obstacle to access for researchers from smaller, less well-funded research institutions on the continent. During July the IHSN's data curation software was installed at the NSO in Senegal, L'Agence Nationale de la Statistique et la Démographie (ANSD). A data request form provided was completed online and also faxed to the ANSD (29 July 2009) and staff responded on 31 July 2009 with an invoice for 411 331 FCFA (\$902) for the data from their 2005 poverty survey. This amount would be a definite obstacle to access for most researchers in the region.

### *Category 8: Good data provision (4)*

Finally, 4 NSOs provided relatively obstacle free access to the microdata from their national surveys (Botswana, Kenya, Nigeria and South Africa). The Botswana Central Statistics Office (CSO) responded with a formal letter from their Director providing permission to obtain the data, which was subsequently sent via email. The Kenya National Bureau of Statistics (KNBS) emailed the data files and supporting documentation. Microdata from the Nigeria Bureau of Statistics (NBS) could be downloaded from their NADA portal. Statistics South Africa data was downloaded from proprietary (NESSTAR) online platform. Metadata to support the usage of the data was also available from the Nigerian and South African data portals.



Survey of Microdata Access in African Countries. 2009. © 2010 Lynn Woolfrey

## 5.5 Comment on Survey Results

### 5.5.1 Technology and human resource constraints

In cases where NSO websites were not accessible or links were broken and activation emails not sent, outdated technology and skills shortages would seem to be the main obstacles to data sharing. Problems with technological infrastructures and lack of technical support are further evidenced by the fact that staff in fourteen of the fifty-three NSOs targeted by the study use public email accounts (Gmail, Hotmail and Yahoo accounts) rather than institutional emails. This would seem to indicate problems with the reliability of mail-servers at these NSOs. There is evidence that IT staff in some cases are barely maintaining websites. While outdated technology plays a role here, irregular website maintenance appears to be due to a shortage of staff trained in ICTs, particularly those with web maintenance skills.

This explanation was supported by staff at the Directorate of Demographic and Social Statistics of the INE in São Tomé and Príncipe. They provided the information that the INE's website was outdated because the staff member responsible for maintaining the institute's website had left and had not been replaced. Constraints related to technology and staff resources meant that even where NSOs were willing to provide their data for specific research projects, the means eluded them. Some NSOs supplied data as compressed files via email but in some cases files were not in a format suitable for sharing. For example, the Malawi National Statistical Office were willing to send the requested data to the researcher but the data files were too large and numerous to

be sent via email and NSO staff could not suggest a secure alternative means of transmitting the data files (this NSO did not have a NADA data portal at the time).

Interviews conducted in 2009 with staff of NSOs in Lesotho, Mozambique, Namibia, Swaziland and Zimbabwe revealed that these organisations have a chronic shortage of skilled IT personnel, and many trained IT staff are lured away from the government sector to better paid private sector employment (Woolfrey, 2010). Websites are the principle conduits via which data producers are able to interact with data users and the public. Ideally these should be utilised as efficient data discovery and data dissemination channels. However the skills to establish, maintain and update websites are in short supply. IT infrastructures in these NSOs are also not up to the task.

Funding shortages are possibly at the root of technological and skills obstacles to effective data sharing in these institutions. This explanation was confirmed to some extent by staff of the Zimbabwe Central Statistical Office who explained that their website was not accessible because they had not paid their ISP bill. The chronic state of under-funding evident in these African NSOs reflects the low priority some African governments have allocated national statistical production because they do not see the value of empirical data for government planning, and do not routinely make use of this resource for policymaking. Poor financial support for data curation also reflects a lack of understanding on the part of policymakers in some countries of the role a data rich research-policy interface can play in good governance

### *5.5.2 Confidentiality issues and bureaucratic structures*

Confidentiality concerns were cited by NSO directors as a reason for not placing national data in the public domain. To some extent this does restrict data availability, as human and technological resources are required to ensure data is anonymised before being provided to researchers. However, The Mauritian example in this study is an indication that, despite sound technological infrastructure, access to data for legitimate research can be hampered by overly-bureaucratic approaches to data sharing on the part of government functionaries.<sup>4</sup>

The bureaucratic culture prevailing in government structures can be seen to hamper data sharing by African NSOs. Government functionaries working with national statistics are accustomed to providing their data to a select group of policy-makers and donor bodies and are unfamiliar with a service orientation that includes the wider research community in their client base. This leads to a dearth of institutional resources at African NSOs for data sharing, including those required to create anonymised versions of the data files.

Inflexible hierarchies mean that staff at NSOs can do little without support for data sharing from their directors, who themselves will need authority invested in them by government ministers to make their data widely available. Unless policy decisions at ministerial level support the placing of national data in the public domain, NSO functionaries will continue to resist making decisions in this regard, for fear of overstepping their responsibilities.

NSOs in Nigeria and South Africa provide anonymised data via their

<sup>4</sup> At the time of the study Mauritius was ranked 51 on the Global Economic Forum's *Networked Readiness Index 2008-9*, ahead of both South Africa (52) and Egypt (76). Only Tunisia had a higher ranking in Africa on this index (38).

websites. These two countries have more financial resources available to them than most countries in the region. However, it is contended that support for knowledge utilisation at ministerial level in both countries plays an important part in this open approach to data sharing. Government support for the sharing of research results is evidenced by South Africa being the only African signatory to the 2004 OECD Declaration on Access to Research Data from Public Funding. Government commitment is vital for data sharing. Tools provided by organisations such as the IHSN mean data sharing is technically feasible, but this needs the necessary government support to become a reality.

### ***5.6 Further Research on ADP Countries***

How did the ADP partner NSOs fare in the survey? On the face of it, no better overall than the other NSOs. Ideally, the tools and training provided by the IHSN/ADP should streamline microdata storage and supply in African countries. The 2009 study showed, though that data access is not widespread among those NSOs which now have these data curation tools (Microdata Access Survey, 2009). Only one of the NSOs that provided free, immediate access – the Nigerian Bureau of Statistics – had a NADA data portal. NSOs in Ghana and Senegal have ADP data publishing tools, but delay access to await payment. While making use of the internet, the other NSOs that gave free, fast access - Botswana, Kenya, and South Africa, do not employ the IHSN/ADP tools. For other ADP participants, obstacles seem to be largely around optimal exploitation of the technology. For example some NSOs do not use the software's online data request form (Niger, Uganda) or the links to data request forms are not active (Liberia) Some participating NSOs are still using the generic request forms provided with the software, rather than customising them (Ghana, Lesotho, Mozambique). Often potential data users are required to print out and fax or mail data request forms to the relevant NSO, even though the form can be submitted online.

The 2009 survey showed that easy access to national data requires more than an initial investment in technological resources. It would seem that appropriate technology is not able to support these goals without government buy-in and the concomitant allocation of data curation resources on a long-term basis. Training and skills exchange initiatives also seem fruitless unless these are provided for an extended period of time. To further examine the impact of the project to confirm this finding a second survey was carried out in 2012

## **6. Follow-up Study 2012**

The follow-up study assessed the data curation practices of African NSOs participating in the ADP and specifically those who have bought into the project enough to have implemented the full suite of data curation tools. The aim of the 2012 study was twofold: Firstly the study sought to confirm previous findings that, despite available technology, African NSOs are constrained from sharing their data by other obstacles. The 2009 survey showed that even without proper data sharing tools, NSOs could provide data for research. Are NSOs that now

possess suitable data curation platforms still not sharing microdata resources with the African research community? By 2012 most of the NSOs in the study had adopted in-house or proprietary platforms to make aggregated national data available online. However, for reasons stated earlier, the 2012 investigation was only concerned with investigating ease of access to survey microdata, and not than other types of data.

The second aim of the 2012 research was to assess the impact of ADP work by examining the project's impact on data quality improvements in general. Data accessibility is only one component of data quality, albeit a key one. An assessment of the IHSN/ADP needs to be in the context of the project's support for data discovery as a vital element in data access, and the projects' role in promoting other aspects of data quality – relevance, accuracy, comparability and interpretability. It is contended that the data advocacy work of the project and the training it provides in metadata creation could be vital first steps to open access to quality data products in the region. Has the project improved other dimensions of data quality? If this is the case, then widespread uptake of the IHSN tools could be beneficial for the systematic preservation and efficient dissemination of internationally comparable African government data. The 2012 research focused on ADP work in boosting data quality for new members to achieve this goal. The research also attempted to find out whether long-term project participation in the project has led to overall data quality improvements at NSOs.

### 6.2 2012 Survey Countries

The second 2012 research was a survey of data portals of ADP NSOs in African countries, to determine research access to microdata from these NSOs. The survey was conducted from July to November 2012 and at that stage thirty-three African NSOs were ADP participants. These included seven of the eight “successful” NSOs – those in categories 7 and 8 in 2009. According to the information on the ADP website at this time Twenty-three of the thirty-three ADP NSOs had online data portals at the time.<sup>5</sup> The research was aimed at assessing the data curation value of the data portals and the training provided for using them. NSOs in Botswana and Kenya are working with the ADP but currently do not have data portals. The 2009 study showed that these NSOs have already established a favourable climate for microdata sharing. It would be interesting to re-examine data dissemination by these NSOs if they eventually adopt the IHSN technology. Surely the introduction of facilitating technologies can only improve their data service? However the 2012 survey only examined data curation at NSOs with existing data portals.

The South African NSO, Statistics South Africa (SSA) was also excluded from the follow-up survey as it is not project participant. SSA uses proprietary software, NESSTAR, for data preservation and dissemination. SSA data and other Southern African data is also disseminated by DataFirst at the University of Cape Town. DataFirst's online data portal makes use of the IHSN software and is the African test site for the platform <http://www.datafirst.uct.ac.za/catalogue3/index.php/catalog>. The author manages

---

<sup>5</sup> Burkina Faso, Cameroon, Cote d'Ivoire, Egypt, Ethiopia, Gambia, Ghana, Guinea, Guinea-Bissau, Lesotho, Liberia, Malawi, Mali, Mozambique, Niger, Nigeria, Senegal, Sierra Leone, Southern Sudan, Tanzania, Uganda, Tunisia and Zambia

this site and is therefore in a good position to evaluate whether the technology is being used to optimise data curation to advance data quality at participating NSOs and what barriers to effective data curation at these NSOs remain.

### 6.1 Dimensions of Data Quality

Quality data can be defined as data that is optimally usable (US Census Bureau, Methodology and Standards Council, 2006a:1). As the original focus of the research, *data accessibility* is an important component of data quality. This refers to how easy the data is to find and obtain (US Census Bureau, Methodology and Standards Council, 2006a:1-2). Better technology was shown in the initial survey to be a less important enabler than assumed. What then do the “successful” NSOs – those who provided easy access to their data - have in common if it is not enabling technologies? Initial findings suggest that commitment by NSO directors is a key enabler for sound data curation leading to easy access. Has the project made a difference here, perhaps through data advocacy?

Another key aspect of data in this context is its *relevance*, that is, its ability to meet users’ data needs. Data users may be government agencies, academics, or private commercial clients. Better data access should make a difference here as users of the data can provide feedback on its usefulness for social research and national planning.

*Accuracy* is another key dimension of data quality, which can be defined as “the difference between an estimate of a parameter and its true value” (US Census Bureau, Methodology and Standards Council, 2006a:1). As with relevance, the accuracy of the data can benefit from independent assessment by data users. The data service at DataFirst has allowed corrections to South African public access data over the years which would not have been possible without “crowd-sourcing” error detection through placing the data in the public domain.

The fourth quality dimension which will be dealt with here is data *timeliness*, as indicated by the amount of time between the reference period of the data and its availability to the public. How does the project assist faster access to released data and shorter waiting times for official releases.

The fifth aspect of data quality pertinent here is *comparability*, judged by whether data from the same statistical project can be compared at different times, between different geographical locations (international comparability) and between different “domains”, for example between industries, or household types (Rosen & Elvers, 1997:627; Statistics Canada, 2003:90). Judging data comparability is only possible with standard frameworks (Rosen & Elvers, 1997:626; Statistics Canada, 2003:15;). The use of a standard curation platform across NSOs in African countries should support data comparisons. Use of this platform requires the adoption of international data and metadata standards by project NSOs, which also improves data comparability.

The final dimension of data quality is its *interpretability*, which is dependent on the availability of useful documentation to support sound data analysis (US Census Bureau, Methodology and Standards Council, 2006a:2-4; Statistics South Africa, 2008:16; Statistics Canada, 2003:6-7). Has the work of the ADP allowed more accurate interpretation of the data held by participating African NSOs?

---

## 6.4 Survey Method

IHSN data portals are designed to allow online data discovery and access microdata and supporting documentation. The researcher used the data portals of ADP NSOs in Africa to discover what data they hold, and to apply for a one dataset from each NSO. Steps involved included:

(1) Finding data portals from NSO homepage links and (2) Registering on the portal as a data user. Registering on the portal involves completing an online registration form which generates an email with a link that the researcher must select to authenticate and activate their account; (3) Submitting a data request. Requesting data requires the completion of an online data access form for the dataset of interest. Requests for public access data initiate an automated email with a link to the data which can be downloaded immediately. Licensed data requests are viewed by the portal administrator and approval generates an automatic email to the user with a link to download the data; (4) Downloading the data and (5) downloading survey documentation, such as questionnaires and codebooks.

### 6.4.1 Accessibility

Technological limitations in under-resourced countries and the lack of standards for data exchange are impediments to the free flow of data. (Committee on Issues in the Transborder Flow of Scientific Data and National Research Council, 1997:29). The ADP focuses on provision of enabling technologies and the promotion of data standards to overcome barriers to data accessibility. There are three components to this data quality dimension. One is data preservation, that is, the long term archiving of microdata in usable formats, including data migration to ensure long-term availability. The second is data discovery, that is, finding out what data is available. The third component is data access, the process involved in getting the data in hand.

#### 6.4.1.1 Data Preservation

Statistical advancement programmes of international donor organisations aim to preserve national data of supported countries to enable reanalysis of this data for donor projects, but also for more extensive use by government functionaries in these countries. Incidents of data loss due to lack of formal data curation policies and practices emphasise the value of these projects for under-resourced nations. The tools and training of the ADP have served to highlight the need for data curation in African National Statistical Systems, and provided the means for this. The portal software allows for multiple data and document formats, ensuring the migratability of stored elements.

#### 6.4.1.2 Data Discovery

The project seems to have made a positive impact on data discovery, although the situation is still not ideal. Research undertaken in 2006 showed that public information available on NSO sites did not even include a list of surveys conducted by each African NSO, let alone metadata on these surveys (Woolfrey, 2010). The advent of data portals at NSOs and the training of local teams to



create metadata for the surveys has provided public information on African datasets, if not the data itself, at least making this data easily discoverable.

NSO websites would likely be the first port of call for researchers wishing to access African government microdata. Data portals that can be easily accessed from homepages are thus vital. These are set up by ADP staff working with ICT departments during training at the NSOs. However the survey revealed that 4 ADP project NSOs listed on the ADP site as having data portals did not provide links their portals from their homepages (Cote d'Ivoire, Guinea-Bissau, Malawi and Mali). For the purpose of this research attempts were made to solve this problem by using the data portal links listed on the ADP's website. However, with the exception of Guinea-Bissau, the ADP links were broken. Even if portal links are available on the ADP site, the absence of homepage links renders these portals effectively invisible to researchers.

The 2012 study first examined data curation at the 11 NSOs which had portals in 2009. Had participation in the ADP in the intervening years improved data discovery at these institutions? Research in 2006 yielded little information on what surveys had been conducted by African NSOs. The focus of the 2009 study was on data access, and did not include an investigation of the role of the data portals in data discovery. In 2012 these portals were analysed in this regard. Are these being used to inform researchers of what data exists? Portals for two of these NSOs – Lesotho and Liberia - could not be assessed as their homepage links were broken. The Lesotho site had been hacked.

The new study showed that long-standing ADP NSOs had survey datasets listed on their portals, with searchable metadata. However, some have a long way to go in this regard: Only 2 datasets are listed on the Sierra Leone site. However for others the portals provided information on a fair number of datasets, for example the Ethiopian NSO included metadata for 104 studies on their portal. Survey numbers for other portals were: Gambia (10), Ghana (24), Mozambique (11), Niger (59), Nigeria (38), Senegal (30), and Uganda (16).

The 2012 study also investigated data portals created after the 2009 study. A list of these was obtained from the ADP website. Two of the 12 sites could not be accessed: The Malawi NSO's link from the ADP site was broken, as were homepage links to portals of NSOs in Mali and Southern Sudan. The Mali portal, however, could be accessed from the ADP site. The 10 accessible portals listed their surveys, with metadata. These included: Burkina Faso (10 studies), Cameroon (27), Cote d'Ivoire (27), Egypt (30), Guinea (17), Guinea-Bissau (11), Mali (16), Tanzania (13), Tunisia (4 - but all public access datasets) and Zambia (25). For these the 2012 survey showed a definite improvement in data discovery possibilities from the pre-ADP days. In the absence of other sources the portals play a key role in informing researchers of surveys that have been conducted by African governments. The work of creating standardised metadata for all surveys held by African NSOs is vital for data sourcing and data comparisons. The data descriptions are harvested for other useful discovery tools such as the World Bank's Central Microdata Catalogue <http://microdata.worldbank.org/index.php/catalog/central>.

#### 6.4.1.3 Data Access

Both surveys showed that it takes more than making appropriate technology available to initiate microdata sharing by government data producers.

Six African NSOs gave relatively easy access to their data in 2009. Botswana, Kenya, Nigeria, and South Africa provided free, online access. NSOs in Ghana and Senegal shared their data. But this was not free, and access was delayed to process payment. Of the 10 ADP NSO with data portals, then, 3 published their data. This showed that, while appropriate technology expedites data access it seems it does not lead to data sharing unless this practice is officially endorsed by NSO Directors acting on open data policies already in place in the institution and endorsed by national policies.

In 2012 23 ADP NSOs had data publishing portals. Three of these - Ghana, Nigeria and Senegal - had portals in 2009, and are still providing data. Two NSOs which had ADP portals in 2009 but didn't publish data now provide data as well as metadata (Gambia and Uganda). Of the 23 ADP participant NSOs with online data portals by 2012, two new members - Tanzania and Tunisia - are now publishing their data.

Improving data discovery mechanisms should promote data access as awareness of the existence of data leads to greater demand. This has been the case in South Africa, where the use of microdata in research has increased considerably with the advent of online access to government microdata from Statistics SA, The South African Data Archive and DataFirst (Woolfrey, 2010). However, data access needs to be both easy and free to affect this. The 2012 study showed that despite available technology and training 4 NSOs have not moved from using the ADP tools for data discovery to maximising their utility by disseminating data (Ethiopia, Mozambique, Niger and Sierra Leone).

#### **6.4.2 Relevance**

Factors relating to relevance include the data's potential to address current issues, for example to provide empirical resources to improve public policy. Measuring the relevance of official data to the user community involves identifying the users of the data and communicating with them to ensure data producers understand their needs, and are attempting to meet these needs. This interaction is also necessary to ensure that users can understand how the data should be applied (Lynn, 2004:576-577). The data portal software produces reports at user and study level, so that detailed information on types of users and data requests is available. Ideally these could be used by official data producers to refine data production to better match national data needs. The survey was unable to test the contribution ready access to user statistics has made to the production of more pertinent data in the targeted African NSOs. However, DataFirst's data service in South Africa has shown that greater access has encouraged user feedback and led to the government statistics office establishing stakeholder consultation workshops for input regarding the type of data collected by official surveys.

#### **6.4.3 Accuracy**

Data accuracy refers to how well the data represents reality. Errors in data collection and data preparation will have an impact on the accuracy of the final data product. Sharing data reinforces transparent survey research methods, which can be improved upon through being under scrutiny, allowing for more accurate

final data products.

The ADP tools allow NSO staff to create standardised data descriptions (metadata) which assist appropriate and informed data usage. DataFirst has been creating metadata for South African surveys for four years using these tools. Their experience has shown that the process of metadata creation interrogates the data and can highlight data anomalies. Data curators are able to introduce metadata elements that can assist them to provide more accurate data. One example is the use of versioning. Corrections to data can be documented to inform data users and version control can ensure researchers that they are using the latest and most accurate dataset for their analysis. Versioning avoids the situation where researchers produce incompatible findings due to inadvertently working with different iterations of the same dataset.

The survey of the data portals which could be accessed showed that most of the NSOs include versioning in their metadata. The Ivoirian NSO provided this for only 2 of its 27 studies, and the Tunisian NSO did not have their four studies versioned, but the other NSOs had versioned the majority of their studies. This will be an important data quality element if these NSOs move on to provide online access to their datasets. Any data quality additions should result in new version numbers being assigned, to keep researchers abreast of changes to datasets.

#### **6.4.4 Timeliness**

Survey data needs to be as current as possible to be useful, and for this reason government data producers strive to meet reasonable release deadlines. NSOs may delay the release of datasets to undertake extra data cleaning activities. However, a compromise must always be reached between the timeliness of data releases and data quality, as reducing data collection and data cleaning periods can adversely affect the accuracy of the survey data (Rosen & Elvers, 1997: 626).

Presumably placing data or even metadata in public domain will allow data users to better monitor turn-around times for surveys, and increase pressure on NSO to make this data available more quickly. The survey revealed that most datasets listed on NSO portals were fairly current, from 2004 onwards, 2 NSOs had surveys listed for 2011 while four had studies listed for 2012. It was impossible to determine from the portals what recent surveys were missing. However if researchers are able to receive launch information for surveys, it would be possible for them to use the NADA data portals to track the time from the completion date of a survey to the time of data availability, exposing unacceptably long delays in data release.

#### **6.4.5 Comparability**

This quality concept relates to sets of data, and refers to how easily survey datasets can be compared with other surveys, over space (different geographical locations and domains) and over time (creating a statistical time-series) (Rosen & Elvers, 1997:626). Comparable data should ensure changes indicated are actual and not related to deviations in survey methodology or unreliable measurements. The creation of comparable data is only possible if data providers adhere to international standards with regard to definitions and concepts (Rosen & Elvers,

1997:627).

As a result of the ADP work, for the first time the data output of African NSOs is being curated using metadata and storage standards, and this opens up possibilities for data comparisons and other interoperability. Metadata is created for data portals according to international data documentation standards such as the Data Documentation Initiative (DDI) and Dublin Core. These standardised descriptions allow a greater degree of comparison among local datasets. More importantly, between country data comparisons can now be undertaken. Creating metadata may also encourage more stringent adherence to standard data collection and data preparation methods, which will further benefit attempts at data comparisons.

#### 6.4.6 Interpretability

Data interpretability is a function of good supporting documentation. Even if initial investigators do share their data, secondary analysts need to be able to understand it enough to reuse it. Therefore poorly documented research has limited value as a research tool (Clubb et al., 1985:55-56). Inaccurate or inappropriate usage of data can be mitigated by the provision of detailed information with each dataset produced.

Ideally, supporting documentation should include both explanatory and contextual information on the survey dataset described (UK Data Archive, 2002:2). The documentation provided with datasets should also include information on the strengths and limitations of the data (Statistics Canada, 2000:3).

The tools provided to ADP project NSOs include software for the creation of standardised metadata for their surveys. All NSO portals examined in 2012 had some form of metadata for their studies, which would assist preliminary data discovery. However, key documents such as questionnaires, codebooks and final reports should also be made available to ensure correct interpretation of the data.

The 2012 study assessed availability of documentation on the NSO sites. At the time of the assessment, November 2012, 4 more portals were inaccessible due to broken links on NSO homepages (Egypt, Mozambique, Niger and Sierra Leone). The original 7 portals still had broken links (Lesotho, Liberian, Malawi, Mali, North and South Sudan and Zambia). This meant only 13 portals were part of this assessment. The majority of these did make key documents available. There were exceptions though: Only 2 of the 10 Burkina Faso datasets had linked documentation. The Ivoirian and Guinean data portals only provided basic metadata online, and no downloadable documentation.

The implication here is that good and readily available data documentation may need to follow data sharing, rather than precede it. Facilitating data sharing encourages fuller documentation, as researchers working on official statistical data are often able to supply feedback to national data producers on weaknesses in data documentation, leading to improvements in these records. This could not be measured during the survey but has been the case in South Africa, where government data is available online. User feedback over the years has led to more comprehensively and accurately documented national data (Woolfrey, 2010).

## 7. Interpretation of results

In terms of data access, the 2012 follow-up survey proved as disappointing as the 2009 study, or more so given the time that has been available to affect data sharing. There have been success stories, and some NSOs are now utilising the IHSN resources to provide data for further research.

Portal functionality was assessed more realistically in the 2012, in the context of overall improvements in data quality brought about at NSOs by the ADP. Rather than viewing the success of the Project in terms of rapid improvements in data access, the 2012 assessment concentrated on the project's role in advancing data curation methods to assist long-term preservation and discoverability and other improvements in the usefulness of the data, such as data that can be easily interpreted.

The 2012 investigation revealed that data discovery has improved at project NSOs, and the standardised metadata created by these NSOs has ensured the cross-country comparability of the data. It has also become easier to interpret data files with the assistance of user documentation now freely available from some sites.

Overall, though, official data curation is still hampered by the same obstacles identified in the 2009 study: poor technological infrastructures and a paucity of skills in both IT and data curation. A continued barrier to access, even where the technological and human resources exist, is lack of support for Open Data from African political decision-makers.

### 7.1 Minimising Technological and skills Constraints

A combination of Open Source technology, World Bank financial backing and ongoing training “road shows” combine to give the ADP the potential to make a real difference to data usage in African countries. Donor organisations support the work of the ADP because they recognise that technology is a key enabler for Open Data. The tools created by the IHSN are free and relatively easy to install and use. Data curation training with the tools is provided by the ADP to NSO staff, to ensure their effective usage. Why then, has the uptake not been as hoped? One reason may be that the project has not paid enough attention to the infrastructure fundamentals of each country in their data development work. A “one-size-fits all” approach to technology transfer may not be appropriate in this context. ADP countries are at different stages with regard to the development of their technological infrastructures. This is evident from the countries' rankings on the World Wide Web Foundation's Web Index. This is a rating of countries according to their use of the internet for the economic and social advantage of their citizens.<sup>6</sup>

---

<sup>6</sup> The ranking is made up of scores for (1) Web impact – the impact the internet has in a country, e.g. the extent of business internet use, (2) Web readiness – the technological and institutional frameworks to affect web usage, e.g. affordability of internet access, and (3) Web content and web use, e.g. internet penetration

DATA ACCESS AND 2011 WEB INDEX RANKING FOR AFRICAN COUNTRIES							
RANK	COUNTRY	WEB INDEX	IMPACT	READINESS	THE WEB	ACCESS 2009	ACCESS 2012
30	Tunisia	50.68	54.06	45.5	41.61		✓
36	South Africa	44.49	46.86	49.61	32.1	✓	✓
39	Egypt	41.05	49.67	22.29	30.16		
41	Mauritius	36.67	34.44	56.95	29.08		
42	Kenya	32.84	37.35	29.32	25.25	✓	
45	Ghana	27.68	27.35	34.96	28.5	✓	✓
46	Senegal	25.38	31.09	32.14	7.69	✓	✓
48	Nigeria	23.57	28.86	18.64	18.11	✓	✓
49	Uganda	20.25	21.75	22.54	22.14		✓
50	Morocco	19.39	21.05	23.25	19.75		
51	Tanzania	18.64	23.47	18.62	12.24		✓
53	Cameroon	15.1	19.81	18.11	7.76		
54	Mali	13.67	20.82	11.88	3.03		
56	Namibia	13.57	13.24	33.55	8.6		
57	Ethiopia	10.89	19.69	4.37	0		
58	Benin	10.43	12.26	22.25	6.61		
59	Burkina Faso	8.51	15.49	0.44	5.83		
60	Zimbabwe	1.94	6.66	0	3.78		

Source: World Wide Web Foundation, 2012

Africa's top performers on this scale are Tunisia (30), South Africa (36), Egypt (39), Mauritius and Kenya (42). These are shown to be countries where the internet plays a role in the economy and is set to be a development resource. The other seven African countries in the rankings are in the bottom ten countries who are unable to take advantage of the internet to grow their economies. All but three of the African countries ranked in the top ten were able to provide data in at least one of the surveys. The three exceptions amongst the good performers are Egypt, Mauritius and Morocco. The Mauritian NSO was willing to share data and this could have been obtained, with some effort. Egypt and Morocco currently have not put policies in place to routinely share data. It seems that while conservative data policies are responsible for the lack of data from NSOs in the Web Index top 10 countries, the poor response from the rest is a product of infrastructural and skills constraints which limit the benefit they can derive from technological support provided by projects like the ADP.

Technology transfer is a key enabler of better data usage, but more, and more tailored interventions are needed. At NSOs where IT skills are scarce, this may involve providing ongoing technology support so that initial technological advances can be sustainable. Provision of medium-term telephonic support such

as that provided by commercial software providers could assist where high staff turnover leads to skills depletion at NSOs. This may be a costly solution but less costly than abandoning the work of the ADP, and leaving a microdata curation vacuum at these NSOs.

## 7.2 Overcoming Political and bureaucratic barriers

Staff at four of the NSOs contacted in 2009 expressly stated that their institutions did not have data sharing policies. Two of these NSOs have been project participants for a number of years: Egypt since 2009 and Mozambique since 2007. This indicates that in fact these institutions have a policy not to share microdata.

Broadening of the assessment of this data improvement project to include examining the overall data quality improvements at NSO provides a more optimistic picture than focusing on data access alone. However, the findings in this study with regard to data discovery and data interpretability were not all good news. It is contended that the situation will only improve once government data becomes available, along with the metadata. Data sharing policies show the good faith of governments to make their findings available for re-examination, and a willingness to improve their data products. Widespread reuse should enable national data producers to “crowd-source” for data quality improvements and encourage more rigour in data collection and data processing.

Optimal data access should be easy and free, or at least not prohibitive. Data access policies which place unreasonable demands on researchers – such as those of the Mauritian NSO - need to be adjusted to make data sharing viable. NSOs that charge for data – such as those in Ghana and Senegal – cater only for affluent institutions, often in developing countries. This is not beneficial to the growth of empirical research in Africa. African countries need a critical mass of researchers utilising national quantitative data to provide the lobbying for data quality improvements, including easy access and better documentation to assist analysis.

The ADP has undertaken data advocacy work to overcome bureaucratic resistance to data curation improvements for better data access. It may be that more prominence needs to be given to this aspect of their work, aimed at senior African government decision-makers. The continent’s data sharing success stories can be provided as examples and to assuage fears around disclosure. An example of a cross-over issue is donor support for better bandwidth to NSOs, which is technology related but can benefit from policy changes driven by donor lobbying and financial backing.<sup>7</sup>

## 8. Conclusion

The survey revealed that, with few exceptions, data preparation and provision for research purposes is not part of NSO agendas. Communication

---

<sup>7</sup> The author participated in a UN Economic Commission for Africa Expert Group Meeting in Addis Ababa in June 2012 where delegates raised issues of inadequate bandwidth allocation to NSOs hampering data dissemination. The group recommended that UNECA request member governments to lobby for better telecommunications services, including improved bandwidth allocations to NSOs.

around microdata access is poor. Data sharing policies and procedures have not been established, and thus data requests from academics are either denied outright, or dealt with on a case-by-case basis. The former situation wastes national resources by preventing reuse of data by researchers for policy analysis which could aid better government planning, and provide innovative input for economic growth. The latter can lead to onerous requirements for data access, again restricting usage and limiting the benefits to African countries of data reuse. Regional and international support and local enthusiasm for data sharing in Africa is hampered by the paucity of technological and human resources in African NSOs, which are the main data producers on the continent. This is partly the result of limited understanding policymakers have of the value of a data-rich research-policy interface for sound national planning. An appropriately trained workforce supported by enabling technologies is a necessary condition for the effective utilisation of national data resources, but these need to be underpinned by sound data usage policies driven by government decision-makers who appreciate the role of data in the development of modern societies, and are committed to using their data resources for national growth.

The ADP and other donor projects aimed at better data curation have the potential to make a difference through bolstering technological and institutional infrastructures to support the evidence to policy chain in African countries. More customised technology transfer, and data advocacy centred on African data success stories may see better uptake of the resources provided. Finally, experience amassed by the ADP and similar projects through monitoring and evaluation could drive more informed donor funded data initiatives in the future.

## References

[website] *About the CRDCN*. (2011). Retrieved December 5, 2011, from <http://www.rdc-cdr.ca/about-crdcn>

[proceedings] The CESSDA ESFRI Project: setting up a one-stop shop for European data. (2008). *IASSIST 2008 Conference: technology of Data: collection, communication, access and preservation*. Stanford University, Palo Alto, California, 27-30 May. Retrieved September 5 2010, from

<https://www.stanford.edu/group/ADS/cgi-bin/drupal/session/the-cessda-esfri-project-setting-up-a-onestop-shop-european-data>

[personal communication] Discussions with national statistics office staff. (personal communication, 2009).

[proceedings] Dupriez, O. (2008). Establishing data archives in developing countries: some initial steps. *IASSIST 2008 conference: technology of data: collection, communication, access and preservation*, Stanford University, Palo Alto, 26-30 May. Retrieved August 31 2009, from <https://www.stanford.edu/group/ADS/cgi-bin/drupal/establishing-data-archives-developing-countries-some-initial-steps>



---

[journal article]Lufumpa, C.L. and Mouyelo-Katoula, M. (2005). Strengthening statistical capacity in Africa under the framework of the International Comparisons Program for Africa (ICP-Africa). *African Statistical Journal* 1: 30-47

[website] Members and partners. (2011). Retrieved December 5, 2011, from <http://www.surveynetwork.org/home/index.php?q=about/membership>

[report]Organisation for Economic Cooperation and Development [OECD].(2004). *Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communique: Annex 1: Declaration On Access To Research Data From Public Funding*. Retrieved September 30, 2009, from [http://www.oecd.org/document/15/0,3343,en\\_2649\\_201185\\_25998799\\_1\\_1\\_1,00.html](http://www.oecd.org/document/15/0,3343,en_2649_201185_25998799_1_1_1,00.html)

[website]Register researchers link on Mozambique Science and Technology portal. Retrieved December 5, 2011, from [http://www.mct.gov.mz/portal/page?\\_pageid=615,1225737&\\_dad=portal&\\_schema=PORTAL](http://www.mct.gov.mz/portal/page?_pageid=615,1225737&_dad=portal&_schema=PORTAL)

[personal communication]Taamouti, M. (Personal communication, 2009).

[Thesis] Woolfrey. 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. Cape Town: University of Cape Town.

[website] World Wide Web Foundation. (2012). About the web index. Retrieved November 30, 2012, from <http://thewebindex.org/about/the-web-index/>

[website] World Wide Web Foundation. (2012). The web index. Retrieved December 1, 2012, from <http://thewebindex.org/data/index/>

# About DataFirst

---

DataFirst is a research unit at the University of Cape Town engaged in promoting the long term preservation and reuse of data from African Socioeconomic surveys. This includes:

- the development and use of appropriate software for data curation to support the use of data for purposes beyond those of initial survey projects
- liaison with data producers - governments and research institutions - for the provision of data for reanalysis
  - research to improve the quality of African survey data
  - training of African data managers for better data curation on the continent
  - training of data users to advance quantitative skills in the region.

The above strategies support a well-resourced research-policy interface in South Africa, where data reuse by policy analysts in academia serves to refine inputs to government planning.

---



[www.datafirst.uct.ac.za](http://www.datafirst.uct.ac.za)

Level 3, School of Economics Building, Middle Campus, University of Cape Town  
Private Bag, Rondebosch 7701, Cape Town, South Africa

Tel: +27 (0)21 650 5708

