DataFirst
University of Cape Town

Research Access to African Government Data

The Impact of Open Source Tools

Lynn Woolfrey

eResearch Africa Conference, Southern Sun Newlands, Cape Town, 8 October 2013

Research Access to Official Data

Referring to raw data from government surveys [microdata]

Allows independent evaluation by researchers to

-confirm

-correct

-extend

government data and the official policies based on the data

Improves researchers’ quantitative skills

Provides real life data for teaching purposes

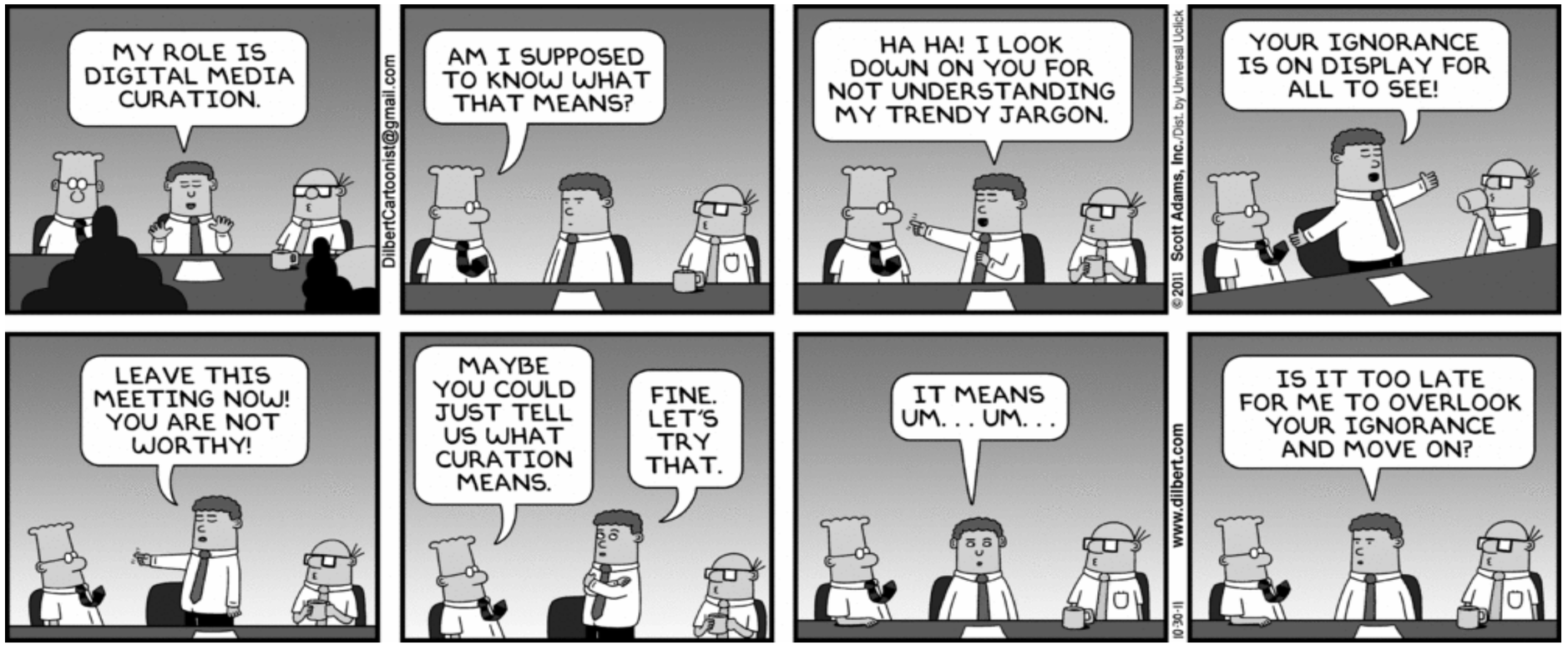
Official Data Curation

A large amount of data is collected in African countries

e.g. The WB microdata portal lists 322 African studies (http://microdata.worldbank.org/index.php/home)

Microdata is collected mainly by National Statistics Offices

Currently this is poorly curated and seldom shared with researchers



”Managing data to ensure they are fit for contemporary use and available for discovery and reuse.”
-UC San Diego
http://rci.ucsd.edu/data-curation/definition.html

Projects to Assist governments with microdata curation

A number of donor projects have been established to assist governments to curate their data

The aim of these projects is to improve all dimensions of Data Quality:

- Accessibility
- Accuracy
- Relevance
- Timeliness
- Comparability
- Interpretability

Accelerated Data Program

The ADP is one such project, established in 2006 and funded by the OECD
The Program assists National Statistics Offices (NSOs) by:

- Installing Open Source data curation tools –
  National Data Archive (NADA 3.0) data portal software
  NESSTAR Publisher metadata creation software
  http://www.surveynetwork.org/home/software
- Training data managers in the tools

DataFirst
University of Cape Town

2009 Survey

Survey on access to African government microdata

Has it helped?

In 2009 10 African Statistics Offices were participating in the Accelerated Data Program (ADP)

Has the project helped researchers to access this type of data?

So….

A survey was undertaken in 2009 to investigate

- Access to African government microdata (raw data, not the aggregated statistics routinely made available)
- The impact of the ADP on microdata access

Survey method

Units of analysis: NSOs in 53 African countries

Primary Investigator: The presenter, an African researcher at an African university (UCT)

Method: A formal request for data was sent to an official email contact from the NSO website.

A recent (but not current) household survey dataset was requested (in the official language of the country)

Online data request forms were used if these existed

2009 Survey

Data - 3 ADP: Ghana, Nigeria, Senegal;
3 non-ADP: Botswana, Kenya, SA
Receipt but no data
1 ADP: Uganda; 6 Non-ADP - Egypt, Malawi, Mauritius, Mozambique, Namibia, Sao Tome
No response - The rest (including 6 ADP)

Successful data sharing needs more than technology and should be underpinned by data curation policies and skills development

Survey limitations

Also, to be fair ….

The 2009 survey did not measure ADP impact on ALL dimensions of data quality

Access is only one aspect of data quality and is tied to other data quality improvements

So…

A follow up survey was undertaken in 2012

DataFirst
University of Cape Town

2012 Survey

Follow-up survey of ADP's impact on all aspects of data quality

Follow up survey 2012

By 2012 23 NSOs were participating in the ADP and had online data portals.

The survey requested data from these 23 ADP NSOs to measure ADP impact on access

AND other dimensions of data quality

Units of analysis: The survey only reviewed 2009 and 2012 ADP NSOs, not non-participants

Method: Data request sent via online forms on NADA data portals

2012 Survey

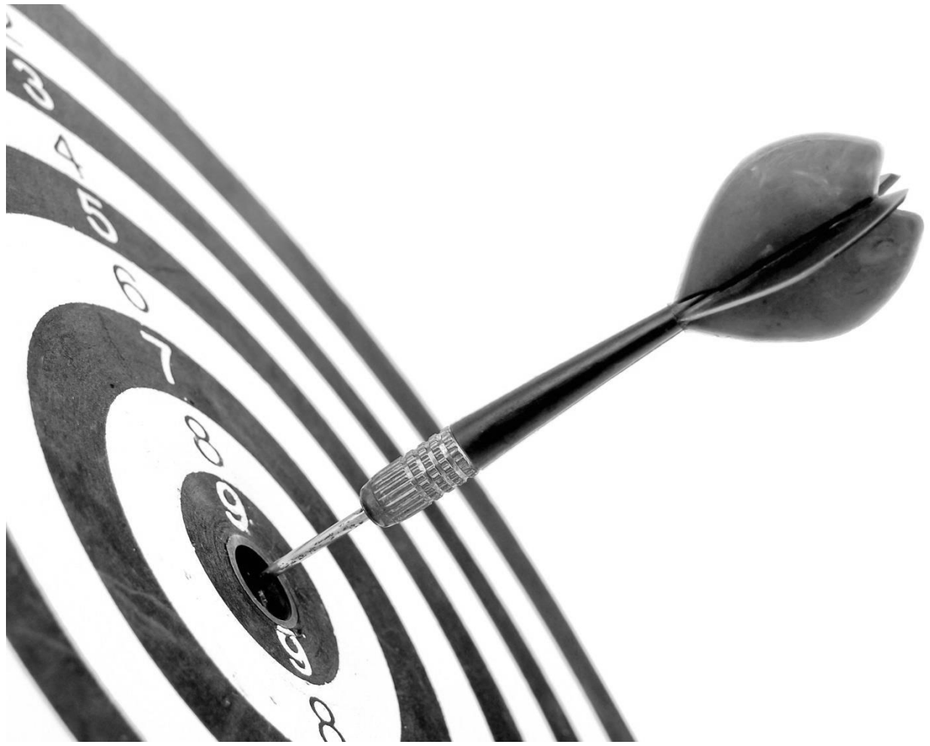
Accessibility 1:
Data from 7 ADP NSOs
2 of these (Gambia, Uganda) are 2009 ADP NSOs which did not provide data in 2009
2 (Tanzania, Tunisia) are new to ADP

Accessibility 2

ADP Data Discovery tools have enabled NSOs to publicise their data – so that we know what data is out there

Accessibility 3

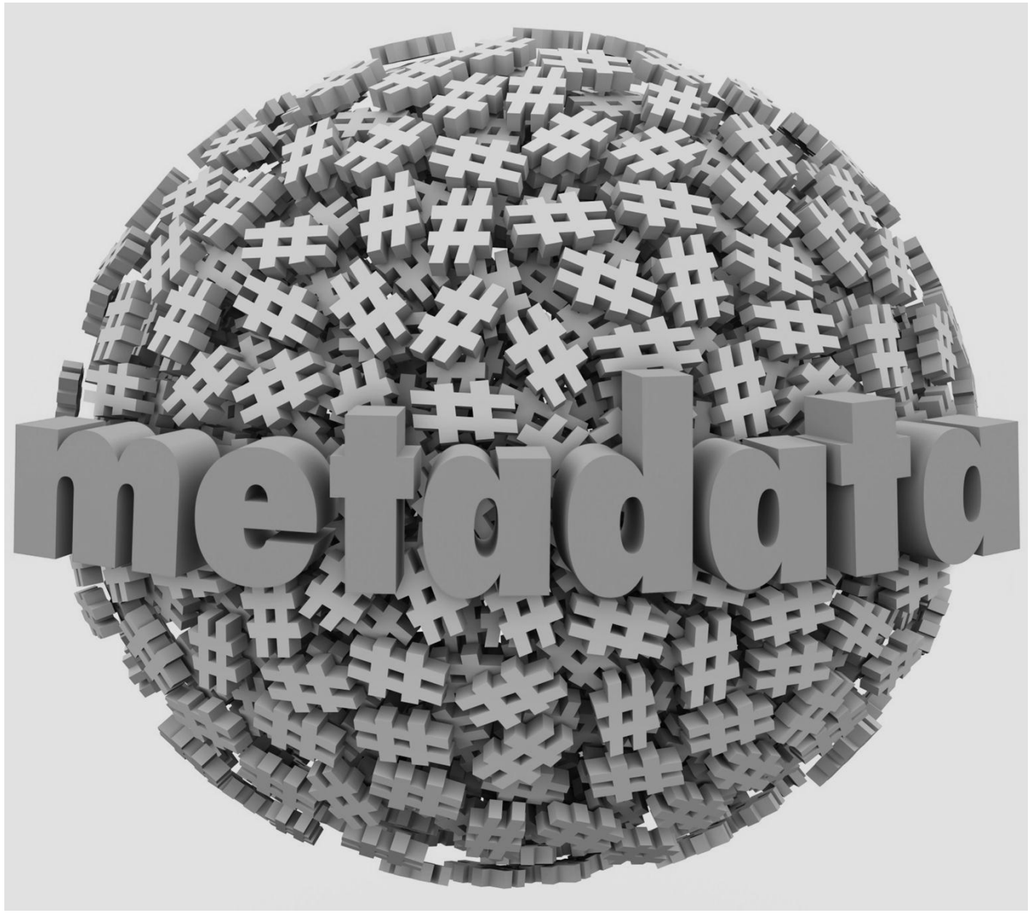
The ADP tools accommodate public, licensed and enclave data access to ensure data security



Accuracy, **Relevance** ,**Timeliness** will improve with data usage (through feedback from data users) But usage depends on Access and suggests the need for advocacy around Open Data

Comparability

ADP tools introduce standard data descriptions and terminology which assist cross-national data comparisons



Interpretability

This needs good documentation and all ADP NSOs have created metadata (usage and access information)

Conclusions

GOOD NEWS:

ADP infrastructure and training seems to have improved

- Data Discovery
- Data Comparability
- Data Interpretability

CHALLENGES:

Access continues to be hampered by lack of data curation policies at NSOs.

This reflects meagre commitment to Open Data on the part of political decision-makers

Recommendations

- Advocacy to African policy-makers around the role Open Data plays in a strong research-policy interface
- Training – more and longer-term, rather than the short workshops currently being held by the ADP
- On-going technical support for NSOs using the tools

For full results see DataFirst Working Paper 22 http://www.datafirst.uct.ac.za/home/index.php?/Papers/

Questions?