“$100 Is Not Much To You”: Unseen Barriers in Data Dissemination

LOUISE BEZUIDENHOUT

UNIVERSITY OF THE WITWATERSRAND
UNIVERSITY OF EXETER

Data Citation and Open Data

- Data citation, data sharing and the Open Data movement
- Citations have value through notions of credit and attribution
- Open Data driven forward by the credit of data citation
- Publication data and citation data as the end-products of a long chain of scholarly activities

Open Data and the "Digital Divide"

"Without open data policies, it is not possible for developing countries to close the digital divide" (CODATA-PASTD 2015).

Open Data also commits scientists in developing countries to share their data

The "Nitty Gritties" of Open Data

- Open Data – and hence data citations – depend on
  - Data to share
  - Deciding *what* data to share, *when* and *where*
  - Seeing value in sharing
  - Having support for sharing activities
  - Having the ability to share

- Very important considerations in LMICs with
  - Low resources for research
  - Relatively isolated researchers
  - Slow pace of research and daily challenges

Looking Closely at Data Practices

- Fieldwork – ethnographies of 4 laboratories in sub-Saharan Africa
- 56 semi-structured interviews with postgraduate and research scientists, laboratory observations, policy review and governance interviews
- How do research environments alter pre-sharing activities and thus engagement with Open Data?

Data Generation
- Working in the lab
- Finding the data to initiate research

Data Processing
- Storage
- Curation
- Accessibility for re-use

Data Sharing
- Publication
- Contribution to databases etc
- Altmetrics
- Profiling data online

Data Re-Use
- Finding data online
- Access
- Downloading data
- Processing re-used data

Key Challenges

- **Physical**
  - ICT provision/infrastructures and personal internet provision
  - Power outages, service delivery and border controls
  - Research funding, agency to spend
  - Equipment availability, technical support, outsourcing analyses
- **Personal**:
  - The ability to use available resources relative to other commitments (e.g. teaching loads)
  - Location of work (home or office)
  - Skills necessary to make use of data
- **Social**:
  - Institutional policies such as data sharing guidelines
  - Formal and informal intellectual property regimes
  - Procedures for procuring and reimbursing resources (e.g. paying for software)
  - Support for data sharing, education
  - High student turnover

Micro-Challenges of Research

- Not all inhibiting factors “big” issues
- Micro-economic factors - under $100
- Ubiquitous in daily research, but difficult to address:
  - Not addressed by funding schemes
  - Core funding not available
  - So common they are not talked about
- What impact does this have for scientists’ ability to engage with Open Data activities?

Getting Data to Share

- ***KY1/13:*** *in Kenya [research] is difficult and they get frustrated. If they do any research it is just a little for promotion. It is tough with no research culture and funding.*

- ***SA2/1:*** *so again, you know they call us to meetings and they say we have funding for this and that. And I think "great stuff", but I wish they would ask me what the real issues are. I'll probably tell you 100 other things outside of the money [permitted to be spent on the grant]*

Accumulating Data to Share

- ***SA2/7:*** *But also with the students you know we have a very high turnover of students and then those students leave and they haven't stored their data properly or they leave with their hard drive.*

- ***KY2/13:*** *it is difficult to build up a body of data when the research is all short-term and ends with the student.*

- ***KY1/9:*** *students publish information in theses and publications and that's where they end up - on a shelf.*

What Data to Share?

- ***SA2/7:*** *probably people who are established who will do [publish in OA journals] for there because now myself if I'm here I must get the chemicals, I must get funding.*

- ***KY2/1:*** *if we want to publish we must pay out of our own pocket. It is important to publish, but it is also expensive.*

- ***KY2/13:*** *How much do we do to develop our own data? What processes do we need to convince people that the data are good?*

The Act of Sharing

- **KY1/9:** *you know, these fees for joining, they add up quickly and you must choose [what to join]*

- **KY1/3:** *here I'm using wifi, so the moment you step out of the college you're shut off and again in the estates [less-formal residential areas] where we stay as of now the internet is a bit expensive. It's not affordable. So I do as much as I can here so that when I go back home I'm going to rest*

Exhausting Enthusiasm Before You Start?

- ***SA1/3:*** *I think it leads to better science*
- ***KY1/1:*** *I won't release data unless I first of all publish*
- Implications for data citations, Open Data and Open Science discussions

Top Down AND Bottom Up

- To build robust systems in Open Data that facilitate data citation we need to have corresponding scrutiny of “bottom up” processes
  - Having data to share
  - Making sharing possible
  - Patterns of sharing (linked to pace of research?)
- Processes of data generation, storage, curation and dissemination need further scrutiny in low-resourced settings

Carrying the Discussion Forward

- Facilitate discussion and openness about daily research practices – it's ok not to be perfect
  - Workshops under discussion in Kenya involving university management, government and scientists
- Accumulate "good examples" of scientists with robust data management systems
  - Discussions with research networks, international networking organizations and academies of science
- Micro-credit system to assist in building robust data generation, curation and dissemination practices
  - Hope to secure funding for a pilot project

Thank you

- Collaborators at Exeter: Prof Brian Rappert, Dr Ann Kelly and Prof Sabina Leonelli

www.devworlddatasharing.weebly.com

www.datastudies.eu