Teaching research data management skills using resources and scenarios based on real data

Veerle Van den Eynden, Jared Lyle, Alexandra Stam
UK Data Service, ICPSR, FORS

IASSIST 2016
31 May 2016
http://tinyurl.com/gvf86nw
Our experiences

- Jared Lyle, ICPSR
- Veerle Van den Eynden, UK Data Service
- Alexandra Stam, FORS (anon)
Overall goals of good data practices

• Complete and self-explanatory data collections for that users can understand and use
• Do no harm
Challenges

• Domain-specific for social sciences: surveys, interviews – may not apply to other sciences on the surface, but principles do apply, e.g. labelling, documenting, enhancing, disclosure review

• Level of experience participants have with data and software – make exercises software agnostic (e.g. not R, Stata), else time goes on learning the software

• Sophistication of user – do they know what data files are, how data are structured, what labelling is?

• Balance messy vs simplistic data exercises
What works

• Single data collection that can be used for teaching variety of topics
• Materials contributed in advance by participants: DMP, consent forms, ethical challenges/dilemmas, research summary … small items that can be shared with participants, or worked into scenarios
• Guided step-by-step exercises as warm-up exercises
• Mixed audiences, senior researchers, challenges that cannot be resolved
• Take existing data collection and mess it up
Know your audience

- Data professionals – training in data curation, review, …
- Researchers – training in good data practices
- Junior vs senior, e.g. postgrads
Topics to cover

• Data review: what is missing? Out of range codes, what doesn’t make sense,
  • Quality of data
  • Documentation, labelling, metadata
• Data > people, so ethical / legal implications:
  • Disclosure review
  • Anonymisation
  • Consent in qualitative research (UK)
• How much time does your audience have
• Lots of other RDM aspects can be relevant too
Training methods

- Hands-on: exercises, scenario discussions, demos (e.g. encryption, checksums, backups)
- Real data / real research
Developing training exercises - UKDA

- **Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003**
  - interview recordings, transcripts and diaries
  - Safeguarded, embargoed and permission access
- **Conflicts and Violence in Prison, 1998-2000**
  - interviews, also coded responses
  - permission access
- **Polish and Lithuanian Workers: Opportunities and Challenges for Trade Unions, 2004-2006**
  - Survey, incl. freetext responses
  - safeguarded
- **Life on antiretroviral therapy: People's adaptive coping and adjustment to living with HIV as a chronic condition in Wakiso District, Uganda**
  - Interviews, ethnographic observations, questionnaire survey
  - safeguarded
Why suitable for RDM exercises

• Research topics that many researchers would understand
• Methodologies that most researchers understand
• Mixed methods: survey and interview
• Sensitive topics, so legal and ethical aspects are important aspects to showcase
• Manageable size data collections
• Well documented
• Some messiness in the data
How do I develop these into exercises

- Leave real messiness in, e.g. tracked changes in documents, poor labelling, inconsistent formatting, disclosive information in freetext variables, poor file names,…
- Insert new messiness e.g. fake names, disclosive variables, disclosive info in file properties, …. based on what we typically see in data deposits
- Reduce size of dataset, e.g. up to 20 variables, up to 10 responses for survey file; extracts of interview transcript (few pages); amended codebook
- Ensure data can be used openly in workshop, so scramble data around or apply pseudonyms
- Get permission of depositor
- Write instructions, research summary, dataset overview
- Write discussions / responses, marke-up versions (e.g. anonymisation)
Example: context


- The research in Entebbe, Uganda, analyses the experiences of people living with HIV following access to life-saving antiretroviral therapy (ART). The study aims to understand how people have responded to a new chance at life, what factors enable people to adjust to living with HIV as a chronic condition, and what support measures affect this adjustment.

- Adjustment to a new life on ART poses medical, social and economic challenges, especially in settings of poverty. People must take treatment for the rest of their lives, and they are usually recovering after a period of serious illness and disruption to their social and economic lives.

- Three ART delivery sites with different modes of delivery and support were compared:
  - HIV clinic in Entebbe hospital
  - 3 government health centres with referral links to Entebbe hospital
  - The Aids Support Organisation (TASO), Entebbe branch

- A quasi-experimental research design was used to compare people affected and unaffected by HIV and ART.
Example: data collection

- Life-history and illness narratives from 38 participants, recruited from the three study sites: notes taken during interview and narrative written out afterwards
- Semi-structured interview from same 38 participants examining key issues of HIV self-management: interview taped, transcribed verbatim and translated to English
- Questionnaire with 263 people living with HIV and 160 community control group participants: questionnaire completed by trained enumerators
- The survey questionnaire measured Illness Perceptions (IPQ brief standard), Quality of Life (WHOQOL-BREF standard), mood (Hopkins symptom checklist), mental adjustment to HIV (MAHIVS scale) and socio-economic status. In addition, participants provided information of their age, gender, household size, years in education, religion and marital status.
- Ethical approval for the study was obtained from the Uganda Virus Research Institute and the University of East Anglia, UK. Overall approval was granted by the Uganda National Council for Science and Technology.
Example: data and documentation files

<table>
<thead>
<tr>
<th>File Name</th>
<th>Date</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>InterviewExtracts</td>
<td>08/03/2016</td>
<td>File folder</td>
<td></td>
</tr>
<tr>
<td>2012_01_18__Interview_3__Final_Copy__CASES__gone_to_printer</td>
<td>15/07/2015</td>
<td>Adobe Acrobat D...</td>
<td>985 KB</td>
</tr>
<tr>
<td>Background_description_for_the_qualitative_data</td>
<td>15/07/2015</td>
<td>Microsoft Word D...</td>
<td>21 KB</td>
</tr>
<tr>
<td>Background_description_for_the_quantitative_data</td>
<td>15/07/2015</td>
<td>Microsoft Word D...</td>
<td>35 KB</td>
</tr>
<tr>
<td>FINAL_Info_Sheet_26_consent_forms_Life_on_ART (1)</td>
<td>15/07/2015</td>
<td>Microsoft Word 9...</td>
<td>61 KB</td>
</tr>
<tr>
<td>HIVUganda_codebook_partial</td>
<td>28/07/2015</td>
<td>Adobe Acrobat D...</td>
<td>398 KB</td>
</tr>
<tr>
<td>HIVUganda_infosheet</td>
<td>14/07/2015</td>
<td>Microsoft Word 9...</td>
<td>61 KB</td>
</tr>
<tr>
<td>HIVUganda_surveydata_extract</td>
<td>15/07/2015</td>
<td>Microsoft Excel W...</td>
<td>15 KB</td>
</tr>
<tr>
<td>journal.pone.0105154</td>
<td>15/07/2015</td>
<td>Adobe Acrobat D...</td>
<td>202 KB</td>
</tr>
<tr>
<td>ResearchSummary_HIVUganda</td>
<td>15/07/2015</td>
<td>Microsoft Word D...</td>
<td>20 KB</td>
</tr>
<tr>
<td>Extract_interviewnarrative</td>
<td>28/07/2015</td>
<td>Microsoft Word 9...</td>
<td>50 KB</td>
</tr>
<tr>
<td>Extract_interviewobservations</td>
<td>28/07/2015</td>
<td>Microsoft Word 9...</td>
<td>40 KB</td>
</tr>
<tr>
<td>Extract_interviewtranscript</td>
<td>28/07/2015</td>
<td>Microsoft Word 9...</td>
<td>48 KB</td>
</tr>
</tbody>
</table>
At the time AK was born, his father was an old man and even though he loved his son much he blamed him for not having educated her. “pataa potele shukuru lako kwa polole shukuru kwa polole,” meaning that he loved his son but the bad thing he did was not educating her.

I then asked why her father did not educate her and she told me, “Tutumia wajumoke inavyoancee kama hali yake hali yuaweza kusimamia.” meaning that the father blamed her for all the dead children and used to call her Kasilata so he abandoned her. She went on and said, “Kasusisikie inavyoancee kama hali yake hali yuaweza kusimamia,” meaning that she was born amidst a lot of sorrow.

At that time AK was born, his father was an old man and even though he loved his son much he blamed him for not having educated her. “pataa potele shukuru lako kwa polole shukuru kwa polole,” meaning that he loved his son but the bad thing he did was not educating her.

Narrative
AK said that she was twenty-one years old but did not know her date of birth. However, he was sure that in 1998, the year in which she gave birth to her first child, she was fifteen years old. She is a Musoke by tribe and was born in Nambo which, at the time, was part of Masaka district although her ancestral homeland is in Katunge in Kamango District in Western Uganda.

She was born in a family of twelve children but nine of them died before the age of two and only three grew into adults. AK does not know why her siblings died as babies. Of the three that made it to adulthood, two children including AK were girls and there was one boy. Today, AK is the only living child of her parents as her two siblings have in fact passed on. The boy died in 1994 at the age of twenty-seven and AK said that he died after working and having dreams for three days. Then her sister died in 2010 from illness.

I then asked AK how she felt about the deaths of all her siblings and she told me, “Tunaitwa Katunde yake yangwana ambao anaweza kupa mmoja wa wageni mmoja wa wageni meaning what can I do? God has kept me alive and I eventually die from the diseases that gave me, I will be up to him.”

AK’s father was a farmer; he owned cattle and grew cotton as a large scale and this was his source of income. Her mother was a house wife but she died during child delivery when AK was still a very young girl. AK told me that her parents managed to take very good care of their family despite their little sources of income, “Tukusikia shukuru lako kwa polole kwa polole,” meaning what was hard about taking care of people those days? You would slaughter a goat and cultivate your food to eat.

After AK left school, she went back to Katunde in Kamango District and towards the end of 1997 met a man...
Example: transcript extract

Interview 2

Jk: Derrick

DATE: 02/11/2011
DURATION: 12:50HRS-13:50HRS
LOCATION: BEHIND THE HOUSE, UNDER A MANGO TREE
WEATHER: IT WAS A VERY HOT AFTERNOON

Int: Do you accept to be recorded?
Res: Yes I do.

Q1. Int: Thank you so much. How much time has passed since your diagnosis? Res: It has been one year and two months.

Q2. Int: Which factors (things, events, people, circumstances) have changed the most since then? Res: Since I got to know my status, I started accessing treatment and so I have seen some improvement. I no longer get sick, I had become weak but I regained my strength and I can do my work.

Int: You have said you no longer get sick. What were you suffering from? Res: It was fever that used to disturb me.

Q28. Int: Have you had your relationship with people in your life changed since your diagnosis? Res: I don't have any problems with people but any way people talk and those who know about it know the status.

Int: What do they say? Res: That so and so is infected, things of the sort.

Int: Are there any changes in your daily life? Let me say of having fun? Res: I married a wife that wasn't the case then. That is what had changed.

Q3. Int: Which things have most encouraged/promoted you your adjustment to living with HIV? Res: I would say taking drugs and minding about taking drugs.

Int: Nothing else? Res: Maybe even the eating habits.

Int: How? What kind of food do you eat? Res: This village food of ours like cassava, sweet potatoes, greens that if you get them.

Q4. Int: What is the most important thing/factor/circumstance that has helped you cope? Res: I would think that they are those things I have mentioned. Nothing much may be to concentrate on what I do like someone can have a job and then he denies his thoughts to it.