Research Data Management: a Researcher’s Point of View

Lyne Da Sylva, professeure agrégée
École de bibliothéconomie et des sciences de l’information
Université de Montréal

Portage & Research Data Management In Canada
RDA 10th Plenary Collocated Event
18-09-2017
Outline

- Informal survey
- Some thoughts on directions for research
AN INFORMAL (AND NON-SCIENTIFIC) SURVEY
Methodology

- E-mail communication with a handful of researchers from various backgrounds and institutions
  - 15 researchers
  - 10 disciplines
  - 3 “sectors”:
    - natural sciences (n=5)
    - humanities (n=5)
    - social sciences (and business) (n=5)
  - General question: “what do you think of the upcoming Tri-Agency policy for publicly-funded researchers”? 

© Da Sylva, 2017
I am making a small informal survey among a few researchers and I was wondering if I could ask you for your opinion.

It has to do with the upcoming policy by Canadian funding agencies (NSERC, SSHRC, CIHR) concerning the management of research data.

Simply put, it will involve the requirement, for funded researchers, of proposing a "data management plan" - a description of the data they will collect and/or produce, and the methods/procedures/tools they will use to ensure data conservation and sharing.

This stems from the preoccupation of allowing the whole research community to benefit from data acquired and produced with public funding.

How do you see this, from your point of view (as a researcher)? In a few words (or a few sentences if you have more to say).

Thank you so much in advance,
Answers

- 13 e-mails: rather short
  - 49 to 341 words (mean of 171)
  - 308 to 2062 characters (mean of 1028)
- 2 face-to-face or telephone conversations
  - a few minutes each
  - I took some notes
Prior Knowledge of Issue and/or Policy

Apparent Awareness

- Yes: 2
- Probably Yes: 1
- Unclear: 6
- Probably No: 4
- No: 2
Feeling of Concern

Apparent Concern

- Yes: 8
- Probably Yes: 0
- Unclear: 5
- Probably No: 1
- No: 0

© Da Sylva, 2017
Overall attitude

- Positive
- Positive but...
- Undecided
- Worried
- Negative

© Da Sylva, 2017
Concerns

- Context-based
- Data-based
- Process-based
- Researcher-based

© Da Sylva, 2017
Concerns (cont’d)

Context-based

Immaturity of the ecosystem
- limits of technological environment
- necessity for training and support for researchers

Bureaucratic obstacles
- volume of required documentation
- requirements on formats

© Da Sylva, 2017
Concerns (cont’d)

Data-based

Variability
- variability among disciplines (types of data, size of datasets, modes of data collection)
- variability among practices (variable interpretation of acceptable publication delays)

Intelligibility of the published data
- for other researchers (for reuse)
- granularity of useful data
Concerns (cont’d)

Process-based

Ethical issues
- confidentiality of data
- ethical use of data

Costs
- time
- processing necessary to make the data intelligible
Concerns (cont’d)

Researcher-based

Reluctance to share data
  • desire to retain exclusivity on the data
SOME SUGGESTIONS
awareness-raising

provide
- tools
- money (for the infamous grad student...)
- better infrastructure (integrating the entire data lifecycle)

provide a « data protection period » (5 years)

quickly give access to data
- « the best data collectors are not necessarily the best analysers »

shift the sharing obligation
- from grant-application requirement to « dissemination »
NOTABLE QUOTES
THE GOOD...
I think this is a good idea – up to a point.
The idea of making data collected through publicly available funds accessible to all is a laudable idea and I fully agree with the principle. It is the application of the principle that may be a little more difficult ...
In any case, if we manipulate data, we should always describe it and justify what we are going to do with it and show how the community can benefit, if possible.
Any good researcher should be able to explain what his or her data is, and how the collection and preservation of this data was done.
THE BAD...
This is an extremely important issue. However, even with all the best intentions in the world for access to knowledge, dissemination and sharing for the primary stakeholders (i.e. future researchers), bureaucratic obstacles abound.
(Political Science Researcher)

- In the humanities in general, but in digital humanities in particular, these problems are obviously much more complex and sensitive than in the depersonalized social sciences, let alone the natural sciences ...
THE UGLY!
Right now, we’re living in the « Wild West » when it comes to data management.
It won’t be easy: you should see the state of the directories on my computer – it looks a lot like my desk!
For my part, it would involve documenting the following:

- the methodology for analyzing source data
- the data processing chain
- the structure and description of the databases
- the metadata of the sources used
- the software used
- the formats (both for processing and for archiving)
- the distribution methods
- the repository provided for archiving

I see this would be a pretty large document to include in a grant application, but maybe I'm overcomplicating things.
SOME THOUGHTS
Only 3 out of 15 researchers were apparently aware of the issue

Attitude not sector-specific
Willingness to pursue discussion

- 7 out of 15 prompted me to get in touch with them to further the discussion, or even asked me to keep them posted about related events
  → a sign of concern
POSSIBLE RESEARCH DIRECTIONS
Some prompted by responses to survey

Study shifts in the conduct of research “before and after”

- practices of data citation
- impact on researcher productivity or evolution of scientific discovery
- shifting patterns in grant applications
- impact on evaluation and tenure

Propose “improved infrastructure”

- e.g. integrating the entire data lifecycle

Explore parameters of data sharing

- « data protection period »
- open standards and links with Open Access, Linked Open Data, etc.
Conclusion

- Apparently necessary
  - Awareness-raising
  - Training and support
- Research prospects: plentiful
MERCI!

Lyne Da Sylva
lyne.da.sylva@umontreal.ca
http://dasylva.ebsi.umontreal.ca