Quality, reliability, and data (II)

In my previous editorial, I addressed the topic of the (preservation of-) data in the context of research. Usually I do not get many reactions on an editorial but this time it was quite different. The reactions fell into two categories:

- Is this problem new? Because in the past ...
- Did you know that since [recent] we have the availability of... / have to use... / are required to use ...?

In principle, the problem is not new. In the past, like at this moment, giving others the opportunity to re-do a certain experiment has been one of the very foundations of academic research. The problem I wanted to point out is that due to the impact of Moore’s Law, not only the availability of new technology exponentially increases but also the time-window during which the technology can and will be used in practice (and therefore also can be re-used) becomes correspondingly shorter. One could wonder if there is room for a new discipline which could be called “paleo-technology” where people do research on the preservation of economically outdated technologies for reasons, amongst others, such as mentioned earlier.

Regarding recent developments in tools/guidelines/regulations, I was, on one side, amazed on how much is becoming available, and on the other hand, how little of it is used. Therefore, I would like to use this editorial to share a few of these recent developments so that future authors can make use of it.

First of all, there are national research organizations that currently not only appreciate but insist on an open and transparent data policy. Examples of this are the Dutch organization for Scientific Research NWO.1

Next to this, there are regulations being implemented that deal with the preservation and preservation terms of data. The Netherlands Code of Conduct for Academic Practice (VSNU), for example, states that raw data must be kept available for a minimum of ten years.2 Fortunately, there are also organizations that give very practical guidelines on how to do this. An example of this is given in http://www.researchdata.nl/en/services/data-management/selecting-research-data/3 where the following points of attention are stated:

- Technical: which data formats, software (standard or tailor-made: research-specific tools), hardware?
- Metadata: available and sufficient? Technical information, codebooks, information on data structure, contextual information, information on intellectual property rights, links with publications or related data (in a collaboratory, e.g.).
- Data: which data from which point of the digital life cycle: raw data, intermediate data, published data?
- Clarity on intellectual property rights, for example, copyright, patent and/or database rights, privacy protection?
- Infrastructure available for preserving the data? Either a data archive or an institutional or thematic repository.
- Costs: how are the costs to be covered for selecting, converting, preserving, and making the data available?

It would be interesting to see future papers explicitly addressing some of the above points. Finally, our own journal does have possibilities to upload supporting material with inclusion of the above guidelines.4

I hesitated to devote, again, an editorial to this topic. What convinced me to do it was, on one hand, the impact that current and future requirements on research data management will have on publishing research results, the fact that pragmatic guidelines on “how to do this” become available and that, at this moment, only very few authors seem to use them. Reactions welcome!

Aarnout Brombacher

Eindhoven University of Technology, Eindhoven, The Netherlands

Chief-editor QREI

REFERENCES
2. Netherlands Code of Conduct for Academic Practice (VSNU)

DOI: 10.1002/qre.2286