

Transparency and openness: Requirements and expectations

By Helene N. Andreassen, UiT The Arctic University of Norway ORCID:
0000-0001-9450-803X

Date: October 2020



Manuscript:

One of the most trend-setting journals in linguistics is called *Language*, published in the US since 1925 by the Linguistic Society of America. In 1994, the editor of *Language* at the time, Sarah G. Thomason, wrote an editorial note where she expressed her concerns regarding the quality of the data that authors referred to in the manuscripts submitted to the journal. I want you to pause the video here, and spend a couple of minutes reading her thoughts.

In Sarah Thomason's view, the reviewers of *Language* – and herself – detected too many errors in the data, too many cases where the data seemed not to be fully attested. Stressing the crucial importance of data accuracy in empirical science, she urged authors to provide sufficient details about their data sources, as well as the methodology they had used when collecting their data. She also encouraged readers and reusers of the data to be critical in their evaluation of the data, and to consult the primary sources when available.

A lot of things have happened since 1994, but there is still much room for improvement. And once again, I am using the field of linguistics as an example. In 2017, two surveys were published. First, a survey of 100 monographs from the period 2003 to 2012 showed that the large majority of authors did not describe the location of their data, nor did they offer any data citation next to the language examples in the text. The second survey, which focused on papers published in 9 scientific journals in the same period, revealed a similarly miserable situation.

The failure to cite data properly in publications is worrying, because it reduces the transparency of the research, and as a consequence, also its reproducibility and replicability. And I want to stop here for a moment and explain these three terms, because you need to understand these in order to understand the overarching goal of good research data management:

First: Transparency in research means that we are explicit about the evidence that supports our scientific claims. This might include our methodology, our data collection, our data analysis, and our interpretation of the results.

And transparency is a prerequisite for the two other concepts, reproducibility and replicability.

Reproducibility of research means that a study contains enough information to allow a repetition of it, using the same original set of research data. Replicability of research, on the other hand, means that a study contains enough information to allow a repetition of it, but now on a different set of research data.

Replicability is particularly relevant for quantitative empirical science, such as the medical sciences and the life sciences. But there are more and more people now who discuss how we can replicate qualitative research, that we typically find in the humanities and in the social sciences. To give one example, people are currently discussing the strength of so-called "conceptual replication", where a study is replicated using a different method.

In recent years, we have heard a lot about the “replicability crisis” or the “reproducibility crisis” that a lot of published research cannot be replicated or reproduced. Non-replicability or nonreproducibility of a study might lead other people to question its scientific credibility. And if such cases are too many, it might have a negative effect on the level of trust that our community has or should have in science.

There are many reasons for the current replicability crisis, and together they disclose the gap that exists between the image that we have of science, where researchers are thought to self-correct through replication, and real life in academia, where the publication race and funding applications more and more dominate everyday work life.

One reason for the current replicability crisis is questionable research practices. These are practices that don't fall under the category of fraud, but which are damaging enough, and unfortunately quite widespread. A few examples of such practices are failing to report all methodological choices, excluding data points, and deciding to collect more data after determining the significance of the existing results.

In order to promote a more healthy research and publication system, scientific journals, funders, institutions and national governments now work together and demand more transparency in published research. And they are not alone. Because in recent years we have also witnessed the increasing activity and importance of different non-profit organisations that are working on the international level. One example is the All European Academies, who have issued the European Code of Conduct for Research Integrity. Another example is the Research Data Alliance, who work on building social and technical infrastructure to support open sharing and reuse of data.

As members of the scholarly community, one of our duties is to publish research. Let us therefore have a closer look at the scientific journals. They are interested in publishing high-quality, transparent research, and they increasingly require their authors to provide open access to their research data and code when they submit a paper manuscript. To illustrate this, let us go back to *Language*, the prestigious linguistics journal mentioned in the beginning of this video. As you can see, the journal now offers a separate, very detailed section on how authors should inform about their data.

You will find similar guidelines in many other scientific journals, and this is why I encourage you to pause the video now and spend a couple of minutes reading through the example guidelines, to see what YOU may expect from journals that you submit your manuscripts to.

In other journals, even stricter measures are taken, and the verb “should” in the guidelines is replaced by the verb “must”. This is a quite important step because it underlines that sharing of data is default and non-negotiable when publishing.

Let us now turn to the funders, who are interested in optimising the impact of their funded research on future innovation and research. The funders increasingly require the research projects to work in line with the so-called FAIR principles, and also to share all their research outputs with open access. To illustrate this, let us have a look at Horizon Europe, the next EU funding programme for innovation and research. First, they believe in the FAIRness of data – and we will return to what this actually means in a separate session. They also believe that we need to focus on openness in science in order to speed up progress and deal with challenges in our society, and they intend, and I quote, to “fully embrace and support Open Science as the new research *modus operandi*”.

Let us now turn to the funders, who are interested in optimising the impact of their funded research on future innovation and research. The funders increasingly require the research projects to work in line with the so-called FAIR principles, and also to share all their research outputs with open access. To illustrate this, let us have a look at Horizon Europe, the next EU funding programme for innovation and research. First, they believe in the FAIRness of data – and we will return to what this actually means in a separate session. They also believe that we need to focus on openness in science in order to speed up progress and deal with challenges in our society, and they intend, and I quote, to “fully embrace and support Open Science as the new research *modus operandi*”.

Projects that receive funding from Horizon Europe will need to document how they intend to collect and use data in a so-called data management plan. They will also need to publish their research with open access, and they will need to share their data unless there are good reasons for not doing so. The European Commission will also make the connection clear between openness and ethics by incorporating aspects of research integrity in all guidance documents, and also by encouraging research projects to adapt the European Code of Conduct for Research Integrity, issued by the All European Academies.

So, what about our institutions and their role in all this? In many countries, the individual institutions have responsibility for implementing national strategies on research integrity and openness. As a result, they dedicate more and more resources to training and guidance in good research practices. Many institutions also look to the international communities to capture current challenges and solutions.

We therefore witness institutions building competence on research data management and open access in general, we see open science policies being written, we see data protection officers being employed, and we observe a gradual increase in doctoral education programmes that include discussions about the purpose, content and implications of transparency and open science.

This being said, not all institutions have come very far in developing the support services needed to help the researchers. And especially research data management is quite new to many. If your institution is one of these, don't panic. If you want to learn how to work in line with good practices, there are many free resources online that can help you improve your knowledge and skills. This course is one such example.

Before I end this video, I want to quote the EU maxim “as open as possible, as closed as necessary”. As indicated in the author guidelines of many journals and in funder guidance documents, not all research data can be shared and are not expected to be shared. There might be ethical, legal, commercial and security reasons for not sharing, and no one expects researchers to breach the contract of trust signed with informants or other partners. This brings me to an important distinction to make, the one between transparency and openness.

All research projects can work in line with the transparency principle, to have all information available needed to replicate the research. In most cases, this transparent research can also be fully open, and places us here. Some of us need to go here. This is when information in our data needs special protection, when the data must be archived with some level of access control. Keep in mind that this is not wrong in any way, to be up here to the left.

What we should avoid is going down here, to the non-transparent part of the figure. We should be sceptical about sharing such research, for instance data that are not described or organised in any understandable manner, and where data points have been taken out and not informed about. But it

would be even worse if our data were neither transparent, nor shared. This would make our research invisible as well as non-replicable, which is of no good either to us as researchers, or to our community in general.

Summary:

In this video, we have looked into the expectations and requirements that are out there when it comes to managing and sharing our research data in line with good practices. I have also talked about why these expectations and requirements exist in the first place, that they are a response to some problematic and questionable research and publication practices that unfortunately still occur. If you come from an institution that doesn't have any formal requirements on transparency and sharing, if you are not committed to any funder policy, or if you are publishing in a journal with no section on data in the author guidelines, this doesn't mean that you shouldn't read up on and work in line with good practices. It is in fact also your responsibility and in your interest as a member of the scholarly community to help improve the health of our research and publication system, and to help maintain the contract of trust between academia and society in general.