The research data management life cycle

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

Date: October 2020

Manuscript:

Each research project contains different types of activities, and we can divide the process into three broad phases: The planning phase, the active phase, and the publication phase. In each phase, there is work to be done on your research data, and we typically call this the research data management lifec ycle.

In the planning phase, which is here, you develop your research question and you plan the research process. In this phase, you also search for relevant literature. For instance, you might want to find out what has already been done on the topic, or you want to read up on your selected method, or you want to consult research on a related topic that you want to use as a starting point.

At this point in the research process, it is also wise to search for existing research data that are relevant for your project. Perhaps you will find data so similar to those that you wanted to collect that you don't need to collect data yourself. Or perhaps you will find data that help you further narrow down, or slightly shift the focus, of your planned data collection. Or perhaps you will find data collected with the same method that you had planned to use, that confirm the success of using this method. Or slightly worse, perhaps you will find data that indicate problematic issues with using this method for answering your research question.

It is not very common yet in all disciplines to share data in public repositories, so it might happen that you don't find very much that is relevant for you. But it is still worth it to spend some time thinking about potential reuse of existing data, and search for them.

When you know which data you need in your project, this being already existing data or data that you need to collect yourself, you should write a data management plan. A data management plan covers all phases of the research process and functions as a dynamic roadmap for your research data management.

This is what we call the active phase of the research process, and this is when you carry out the data collection, when you analyse the data and when you incorporate the results and data visualisations into your research paper manuscript.

In order to control your data files, you should organise them properly, with understandable naming and an intuitive folder structure. You should also write up some documentation that describes your data, which helps you remember details about them when you write your paper, or when you share your data with others.

You should also make sure that you store the data in secure, quality-assured systems, and that you have good backup routines. You don't want to be one of those who lost their laptop with the only copy of their research on it. There are a few of those out there, unfortunately. Decisions about storing are particularly important if your data contain sensitive information. In order to reduce the risk of sensitive information getting out, these data need to be stored with special protection.

It might be, when going through your data, or after analysis, that you discover that you need additional data. This makes the active phase a potentially iterative phase.

Finally, we have the publication phase, when you have finished your research paper manuscript and you submit it to a journal. If you write a doctoral monograph or a book, this is when you submit the entire manuscript for evaluation.

At this point, you should make the research data you have used in the paper available to those who are going to review your work. The easiest way to do this is to submit your data to a data repository and then share the link to the dataset with the reviewers. In many cases, you may ask the data repository for a private link, which means that you can wait until *after review* before you actually publish the data and make them available to everyone.

Here, I need to mention that not all data can be shared with others. There might be ethical, legal, commercial or security issues that prevent you from doing so. Whether you can archive your data files with open access or whether you need to have some kind of access control, depends on the degree of sensitivity of your data. It is your responsibility as researcher and author of the data to find out what you can and cannot do with regard to sharing. Still, you should always have as a goal to archive your data in a repository. This way, people can find information about your data, and for your own sake, this way you know that the data are safely taken care of in the years to come.

If you have used other people's published data, you simply refer to these in your manuscript. If you have integrated other people's data into a larger dataset, you need to check what they allow when it comes to republication of their original data. Most often, this is indicated through a reuse license.

So, now you know what the life cycle of research data management looks like. As you can see, there is work to be done on your research data throughout the research process, but if you plan well and establish good routines early, you will most likely work more efficiently and effectively on and with your data, which in turn might contribute to the quality of your published research.

It might be that you need to adjust aspects of your project during the process, and here your knowledge about the planned life cycle of your data will most likely help you.

In this course, we will go through the entire life cycle. We will focus on the most important aspects in each phase, explain terminology and propose some good practices that have been established by the international community. We hope this will help you see how you can proceed when working with research data in your own projects.