

Depositing a dataset into a repository

Transcript of video

By Helene N. Andreassen, UiT The Arctic University of Norway

ORCID: 0000-0001-9450-803X

Date: December 2022



In this video, I am going to show you how to deposit a dataset in a repository. I am using my own institutional archive which is on the Dataverse platform. I click the 'new dataset' button and the metadata template opens with some fields already pre-filled.

What I first need to do is to add the title of the dataset. This dataset is replication data for a paper that has been written, so I put the prefix 'Replication data for:' and then the title of the research paper.

I then add the name of the author. Since this is an institutional archive, the affiliation has already been pre-filled.

I then add the name of the contact person which in this case is the same as the author. And then the email address.

Another field that is obligatory is the description field, but I am going to do that later. But I add today's date.

I am then going to add the subject. This is a dataset in linguistics, so I am going to click the Arts and Humanities category.

Then keywords are required, and you need to add at least one, but what I am going to do is add as many as relevant. In this case, I am adding information about the languages that have been studied. And I am adding information, well keywords, that are related to the research question.

The producer is important, because this is the organisation that has the administrative and economic responsibility for the dataset. The distributor is the owner of the repository.

I am saving the dataset, and now I have already some metadata in place. But I want to add more. I am clicking the 'add and edit' metadata-button. And the metadata template opens once more, but now there are a few more fields available.

What I am first going to do is to add the ORCID of the author. This is a unique ID for this person, which is easy to create on the ORCID website.

I am also going to enter the description, which I have now finished. I am adding the information, well, the description about the dataset. What I also want to do is add the abstract for the paper that is

related to this dataset. I am taking that one too, and putting it in a separate field. And just to show what is what, I am adding some information in the beginning of the description, showing what is the dataset abstract and what is the publication abstract.

This has already been done. I am going further down.

And I am going to the 'related publication' field. In this case, the research paper has been submitted, it has been accepted, and it's about to be published. I already have all the information I need about the

bibliographic reference. I can also say that if the research paper hasn't been published yet, or hasn't been accepted by the journal, that it's only submitted, in that case you cannot write all the information here. You have to wait, until it's authorised to inform about where the paper has been published. When your paper has been approved, you can ask whether it's okay that you put information about the forthcoming publication in this field.

OK, so going further down. Another field that I want to draw your attention to is the contributor field. In this case the dataset is a doctoral project. I am entering information about the supervisor, who has been contributing advice on the data collection. I also add information about the grant agency, which has funded this research.

In this case the researcher has collected, or has taken data, from two different time periods within the data material. He has taken data from 1958 to 1970 and also from 2008 to 2020. And this is important for people who want to look at the data, to know from which period they come. But the collection of the data, the extraction of the data from the primary data, happened in 2020.

We also want to inform about the kind of data, so these are corpus data.

And we also want to inform about the software that has been used to analyse the data. He has been using Praat, which is a phonetic and analytical tool, and then RStudio. And I also inform about the version that has been used.

As you can see there are quite a few fields, and not every field is relevant for every dataset. You just need to get acquainted with the metadata template, and decide what is irrelevant.

What might be relevant is the geographic coverage of your data, in the case you have data that are taken from a specific geographic location. That is not the case for this dataset. You also have 'Social Science and Humanities' metadata, where you can inform about the data collection methods, and also information about the management of the data.

I am going to save my changes, and what I am now going to do is add files. I am in the files-tab.

And I click the button 'upload files.' I click 'select files' to add, and I go into my computer and find the folder where I have prepared my data files. I open them, and as you can see in this repository you could add a small description for each file. But I decide not to do so in this case.

Now I have the four files in my dataset. I go over my metadata to see that everything is in order.

But I am not done yet, because as you can see, in the versions-tab here, my dataset is only a draft version. So now I need to proceed to get my dataset published. And this differs from one repository to another, how it happens. In this repository, the researcher is not allowed to publish the dataset himself. He has to send it to curators in the repository, who go through the dataset and then, if everything is in order, publish the dataset.