

Open/Embargoed/Restricted

Transcript of video

By Henning Hansen, UiT The Arctic University of Norway

ORCID: 0000-0002-9186-1441

Date: December 2022



In this session we will talk a little about how different kinds of data can be shared, and how the level of sensitivity of the data sets the limit for the level of openness. As a starting point, most research data can be shared openly, without breaching any rules or regulations. With that said, there are several restrictions that need to be taken into account when it comes to sharing data. Legal, ethical or commercial reasons, as well as security, may constitute aggravating circumstances, which in some cases may prevent the sharing of data altogether.

As a first step, before any data can be shared, the ownership of the data needs to be determined, and regulations and requirements from the employer or university or any external funders or publishers need to be reviewed. If you are working within a collaborative project, make sure you include the sharing of data in your signed agreement. When the sharing of data has been given the green light, it is important to determine the level of sensitivity and confidentiality for your data. Are there any commercial reasons why it would be unwise to share the data? Does the research data contain sensitive information, for example personal data? Can, and should the data be anonymised? Are there certain parts of the data that should not be shared openly? What are the potential risks involved in sharing the data publicly? These are examples of types of questions that need to be raised before the data is shared.

Research institutions have different ways of classifying data according to sensitivity. Some operate with colour coding for the classification of research data, for example green for non-sensitive data and red for sensitive data, or more refined versions such as the example from Carnegie Mellon University. Others, for example North Carolina State University, use a simpler but still distinct system. The basic distinction between data that can be shared openly on one side, and data with limited or no access on the other side, permeates most systems. The classification of research data as sensitive or non-sensitive data can be applied to entire datasets or to individual files or parts within a specific dataset. Thereby, it is sometimes possible to share certain parts of the data openly. However, keep in mind that not all files must necessarily be deposited in the same repository. By equipping them with the necessary metadata, it is still possible to connect them to one another.

Even in cases when a dataset contains information that cannot be shared openly, there are often ways of sharing at least some information regarding the data. Depending on the nature of the data and the level of sensitivity, it may for example be possible to share the processed data as well as the affiliated metadata openly, while access to the raw data remains restricted or closed. If that is not possible, another option is to share access to the metadata only. In that way, the dataset can be indexed, and the existence of the dataset may thereby become apparent to the research community and other relevant or interested parties. In some cases, it is not possible to share any of the information mentioned above.

To sum up, it is important to keep in mind that sharing data is far from a black and white scenario. Rather, it is more of a continuous scale from green to red. Just because a dataset contains information that should not be made publicly accessible, that does not necessarily mean that nothing about the dataset can be shared. It is always advisable to strive after sharing as much information as possible openly, and thereby follow the recommendation from the European Commission (and others) of sharing research data as open as possible and as closed as necessary.