

New perspectives on data management and open research data in the Nordic countries

On October 3rd, a workshop on the topic of good data management in the Nordic countries took place in Stockholm. The workshop was organised by the Swedish Research Council in collaboration with NordForsk and NeIC (Nordic e-Infrastructure Collaboration). The focus of the workshop was strategic work on data management issues, enhancing FAIRness and open access to research data, and exploring potential for Nordic collaboration.

The aim of the event was to bring together Nordic stakeholders who work strategically with data management issues and to enhance FAIRness of research data. The participants came from a range of organisations in the Nordic countries, representing, among others, researchers, universities and research funders.

Sometimes talking about open data and digitisation sounds like something that will happen in the future. However, already today, we live in a world of digital information everywhere around us, facing us with new challenges and expectations on not only keeping up relevance and good quality of research data, but also enabling increased online access to data. In order to enhance the quality of research data, there is today a strong need for further evolving traditional methods of good data management, creating new ways of enhancing quality, persistence, relevance, openness and reusability.

At the beginning of the workshop, Sofie Björling, director of the Department of research infrastructure at the Swedish Research Council, welcomed all participants and highlighted the wide range of topics on the agenda. This report follows the structure of the workshop, by summarising the presentations given by speakers, as well as pointing out a selection of comments from the plenum.

The current state of Open Access and Open Data in Nordic countries

The workshop started with a survey of the current state of Open Access and Open Data within the Nordic countries, based on a review of current practices of Nordic data repositories, presented by **Andreas Jaunsen** of NeIC. The review will soon be published on [NeIC's website](#).

One of the main challenges appeared to be access to data. One of the conclusions of the survey was that even if almost all repositories included in the study provided unrestricted access to their metadata, a majority (70%) of the repositories do not provide unrestricted access to all of their data, but mostly to certain parts of it. Concerning findability and persistence, approx. 60% of the repositories do not issue PIDs, and a majority (56%) of the repositories do not employ any metadata standard.

The survey showed that while metadata is typically openly accessible, the data itself may be open or restricted. Moreover, the survey of these repositories concluded that a large majority

of the repositories in the survey sample were not certified and do not follow standards like, for example, Core Trust Seal. The scope for Nordic collaboration was highlighted, as only few of the repositories surveyed are results of collaborations between the Nordic countries.

The FAIR principles as guiding principles in open access to research data

The FAIR principles are a set of guiding principles to make data Findable, Accessible, Interoperable and Reusable, describing how to manage scientific data in order to facilitate discovery, use and reuse.

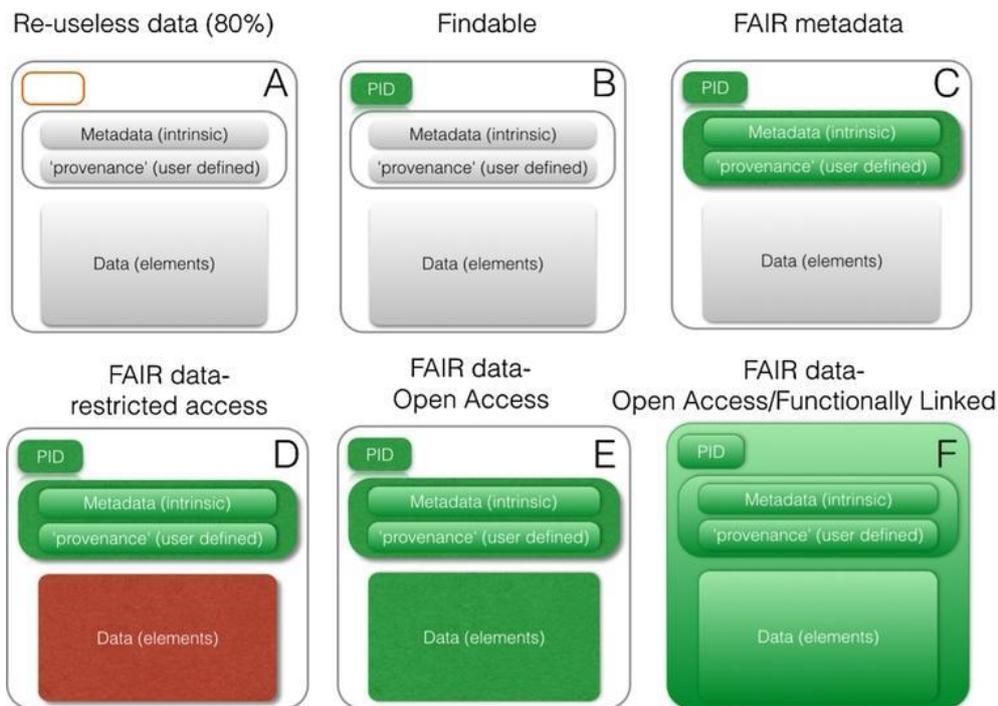
Sarah Jones, Associate Director at DCC (Digital Curation Centre, UK), and also involved in work of the European Commission Fair Data Expert Group, summarised activities on a European and international level, in particular regarding the European initiatives on turning FAIR data into reality, such as the FAIR Data Action Plan, and connections between FAIR and Open Data.

Jones pointed out that FAIR data is not equal to open data: data can be FAIR, but not open or open but not FAIR, which implies that there is a measure for both openness and FAIRness. FAIR concerns managing and sharing data in a way that optimises its structure, interoperability and usability. Furthermore, FAIR enhances the sharing of data in a way that is useful (as opposed to just uploading data on the internet). Somewhat relatedly, Open Data is about accessibility and reusability of data.

Data is supposed to be useful and accessible to as many as possible, and here interoperability is one of the key challenges. A further challenge is presented by the way researchers understand FAIR. Jones highlighted that researchers often find the discourse surrounding FAIR too impenetrable due to the use of jargon. The use of the FAIR framework to structure the DMP is further made difficult by the fact that some of the concepts related to FAIR overlap.

- Some avenues for further efforts were mentioned, such as: Support disciplines to develop their interoperability frameworks for FAIR sharing
- formal learning to promote data management skills, for example in the form of MA programmes for professionals (data scientists and data stewards)
- informal learning: summer school, shadowing, on-the-job training
- data as a criterion for promotion of researchers at universities
- the need for metrics to be defined by the research community itself because FAIR implementation will most likely differ between disciplines
- the need for funders to coordinate their investments, while thinking about how their actions fit in with the EOSC.

Data as increasingly FAIR Digital Objects



Source: <https://content.iospress.com/articles/information-services-and-use/isu824>

After those presentations, group discussions took place, focussing on identifying the main priorities for making open access to research data a reality.

Some of the priorities mentioned by the participants are:

- **enhanced knowledge of data management/FAIR/open research data at all levels**
- **coordination of work in order to avoid double constellations**
- **adoption of DMPs (data management plans) to define and make responsibilities visible**
- **the role of the funders in pointing out that DMPs are required**
- **investments in infrastructure, technical solutions and support**
- **incentives and recognition for researchers to share their data, which leads to**
- **culture change mechanisms for promoting data stewardship**

Good data management as a pre-requisite to increasing quality in research

The focus of the next session of the workshop was good data management as a pre-requisite to increasing quality in research. The first of the two presentations during this session was an example of the work with a policy for open research data.

Elin Stangeland, adviser at the University of Oslo Library, presented the University of Oslo's policy and guidelines for research data management. In particular, the presentation demonstrated what actions have been taken to follow up work with research data within the

institution, in order to ensure implementation with the purpose of increasing quality and reusability in research at the University of Oslo.

In 2018 the Norwegian Research Council introduced a requirement for DMPs to be approved within the institution and to subsequently be openly available online. UiO are working locally to facilitate these requirements, including reviewing the need for a DMP template. She raised the question of whether a national and international collaboration in establishing templates for DMPs would be a good approach.

Echoing training in data management skills as one of the priorities identified by workshop participants, the University of Oslo is also putting efforts in this area. Stangeland mentioned an ongoing project on skills development for researchers and research support staff, which is now ready to pilot courses in December 2018 and January 2019. For researcher oriented courses DMPs will be linked closely to the researcher's own project management (the so-called "what's in it for me" perspective). In the longer term online courses will be considered. Stangeland also raised some challenges, for example relating to how sharing data openly, or contributing to development of infrastructure, does not have a lot of impact on career progression at the University of Oslo as this is still very traditional, oriented towards publications and citations.

Some further challenges mentioned were how to track the impact of the universities' initiatives concerning research data, as well as how to verify quality, reusability and reproducibility. How much infrastructure should be in place before the policy is formalised at the university level was a further difficulty mentioned.

The second talk, given by **Johan Åhlfeldt**, research engineer and system developer at the Centre for Digital Humanities, University of Gothenburg, focussed on outcomes from FAIR data management in humanities research. Åhlfeldt presented the Pelagios Commons project as an example of digital outcomes from research that could not have been possible without open access to data. The Pelagios Commons collaboration involves 30 projects worldwide, which publish their data online. The data, which has a strong cultural heritage component, is described in a way that makes machine-readable, machine-driven retrieval of data possible. There is also an interface for visualisation in a way that is readable for humans.

In humanities research, digital collaboration has been adopted as a work process, already today combining different kinds of digital sources, cultural heritage data (e.g., manuscripts, texts, artefacts), open governmental data (e.g., administrative data, geographic data), research data made available by others, and the researcher's own data collections.

Some of the challenges raised by Åhlfeldt concerned importance of rethinking new ways of dissemination of research information, long-term sustainability and availability, ontologies, management of digital objects and pushing the limits by good examples. On the question of what he sees as the main priority, Åhlfeldt mentioned persistent identifiers, as well as being able to interlink datasets through use of common authority files (e.g., persons) and a common infrastructures that can provide identifiers to online machine-readable data.

Towards FAIRness in research data in the Nordic countries

The talk given by **Per Öster**, director of Research Infrastructures & Policy at CSC - IT Centre for Science, focussed on connections between data management and FAIRness, support, needs and roles in the data management cycle.

Öster highlighted that the CSC provides support in all phases of the research process. In each stage, there is an interest in sharing of data. The issue of financial responsibility was again mentioned. For example, the Terms of Use of certain free commercial services stipulate that access to the service can be terminated at any time, without cause and without notice. The fact that these services are popular among researchers highlights that while there is a strong demand for research data sharing, the research system should be careful to adopt commercial services. A further challenge pertaining to data management is that published data is very specific to a particular project, which can make it difficult to reuse. Therefore, it is important to specify under what circumstances data can be (re)used, as well as where any data has been deleted.

How could Data Management Plans be used to promote FAIRness in research data?

The first presentation on this topic was given by **Maggie Hellström** from Lund University, ICOS Carbon Portal ENVRIplus and domain specialist for SND. The talk focussed on promoting FAIRness with some reflections from an Earth Science perspective. Hellström pointed out that it is important to, first, look at each of the components of FAIR (i.e., F, A, I and R) and to identify where the focus should be for oneself as a data provider in the research data lifecycle and second, engage users to get feedback on whether one was successful in actually making data FAIR.

The perspective of researchers was again highlighted. Hellström stated that researchers still have a number of questions regarding FAIR data management. Some examples mentioned were:

- whether all data, including raw data, should be shared
- financial responsibilities
- responsibilities for doing the work involved in making data FAIR
- whether FAIR applies to data generated from “older” research projects
- whether there are any sanctions for not complying with FAIR

In sum, it seems that many researchers are still hesitant regarding FAIR because they think it involves substantial extra work. However, Hellström pointed out that the researchers might not realise that they have already done some of this work as part of their research practice. **Adil Hasan** (UNINETT/Sigma2 in Norway), whose talk was entitled “*easyDMP*: making data management simpler,” presented an example of a DMP tool. The *easyDMP* is a common tool that breaks down the DMP recommendations into manageable parts for each stage of the data management process. The tool also includes an administration interface which facilitates the sharing of DMPs within an institution or a data centre. Hasan highlighted the challenge of integrating the DMP within the whole research process. DMPs should not be created at the start of a project, only to be ignored at a later stage. Hasan stressed that researchers in Norway actively access data even after the data has been archived. This shows that there is a need and a demand for DMPs to become integrated within the whole research cycle.

The research funder's perspective on DMPs was presented by **Jyrki Hakapää** of the Academy of Finland, who stressed the Academy's view that high-quality research should go hand in hand with responsible research. The Academy recommends certain storage services, which can be discipline-specific such as FSD (a national social science archive, linked to CESSDA) and FIN-CLARIN (language data service, linked to CLARIN-ERIC), or aimed at all research disciplines such as services provided by CSC. The researcher is also free to find their own, reliable, discipline-specific archive services.

Researchers applying for a grant with the Academy of Finland are required to submit a DMP. Reviewers of grant applications receive guiding questions such as "What is the intended level of open access to research results?" and "Does the DMP responsibly support the reuse of research data?" According to Hakapää, evaluating DMPs has proven to be easier in some disciplines such as medicine rather than others, such as the humanities.

The DMP tool *DMPTuuli* is recommended by the Academy but applicants are not obliged to use it and can create their own DMP by following the recommendations on the Academy of Finland's website. It was mentioned that the basic form of the DMPs in *DMPTuuli* is the same for all disciplines, but what is required for a particular research project is discipline-specific. Completing and evaluating a DMP is naturally more complex when a project spans across disciplines, which is why it is crucial for the disciplines make themselves mutually understandable.

On the question of what weight a DMP has in the review process, Hakapää answered that the DMP is not the most important part of an application. However, theoretically it is possible for an application to be rejected due to a poorly executed DMP. Currently, the Academy of Finland does not follow up DMPs during the research project. Once the project is finished, the researchers are asked to explain how the data has been collected, stored and how it can be opened if possible. The Academy does not have sanctions for not opening data, and following up the DMPs is regarded as more a question for Higher Education Institutions. CSC are helpful in this regard as a service to support the ideas that have been presented in the DMP.

Norwegian national policies and principles on Open Research Data were presented by **Ulrike Jaekel** at the Research Council of Norway. Jaekel summarised the current situation concerning the enormous growth of data over the past years, as well as the lack of secure and standardised management of data and metadata descriptions. Jaekel described the principles of the Norwegian national strategy on access to and sharing of research data and the Research Council's Policy on Open Access to Research Data. DMPs are required by the Research Council for all funded projects from 2018 and should be compliant with the FAIR principles.¹ Projects are requested to submit a first version of the DMPs available to the RCN once the project has been notified for funding and before a contract is signed with the Research Council. If the project manager decides that the project does not need a data management

¹ https://www.forskningradet.no/en/Article/What_is_a_data_management_plan/1254032567145

plan, an explanation for this must be provided to the RCN when submitting the revised application. An updated version of the DMP needs to be submitted with the final report. Moreover, DMPs should be public and made openly accessible, if possible. The researchers have freedom in choosing which DMP tool they want to use and the content of the DMPs are not evaluated by the Research Council. The R&D-performing institutions or companies are responsible for assessing the content of the DMPs and ensure that these are in accordance with the institution's own guidelines and policies.

Panel discussion and conclusions

After the presentations, the participants in the workshop discussed a range of topics, such as the approval of DMPs, roles and responsibilities at different stages, for example concerning data storage specified in the DMP, but also long-term preservation after the project has finished.

The final session was a panel discussion on lessons learned and potential for Nordic cooperation. The participants concluded that there is good Nordic coordination within specific domains. Moreover, since it is common to share data within research infrastructures as well as within certain disciplines, interested stakeholders should look to these infrastructures and disciplines for learning opportunities. The panel highlighted the need for incentives for researchers, as well as good examples and success stories, which can be used in training courses on research data management.

A further important point concerned the responsibility for the data of the institution after the research project is completed. In other words, should the institution report back to the funder? The participants stated that, at least in Norway and Sweden, the responsibility is within the HEI where the data was created, which demonstrates the importance of a research data office at a university level.

Storage of data was highlighted as an important issue in several presentations during the day, in particular regarding roles and (financial) responsibilities. It was stressed that the researchers need to be made aware of what options they have, as well as what kind of support is available at different levels in life-cycle of research data.

Regarding the whole research cycle, the participants discussed the benefits that DMPs provide with respect to planning of necessary resources, efforts and services connected to data management needs. Some examples of information that can be gleaned from a DMP include the type of data created/collected, the need for computational capacity and storage. DMPs can also be beneficial for defining roles and responsibilities for data management during different stages of the data life cycle. Responsibilities for long-term data management and for supporting infrastructure and service providers need to be clarified.

During the workshop, the question emerged regarding responsibilities for storage costs, which is even connected to the long-term management of data sets in a way that allows for long-term conservation and availability (storage including authentication and conservation of the

context, as well as archiving) over time. DMPs can thus benefit several actors who are part of the life cycle of data and a common template (and tools) can clarify roles and help highlight the needs of data management.

In sum, Open Science is still insufficiently mature: for example, some funding agencies require and evaluate DMPs, while others do not. A natural next step would be to examine whether DMPs are implemented and define responsibilities for implementation. A final step would involve looking at whether the benefits of Open Science are actually realised. The panel agreed that it would be good to have a Nordic knowledge network as a starting point. The arrangement of the workshop was appreciated, and several of the participants pointed out that the topics discussed are very important, which means that there could be a need for continued Nordic events about these topics in the future.

Text: Sanja Halling and Svetoslava Baumann