

LCRDM

The National Coordination Point Research Data Management (LCRDM) is a national network of experts on research data management (RDM) in the Netherlands. The LCRDM connects policy and daily practice. Within the LCRDM experts work together to put RDM topics on the agenda that ask for mutual national cooperation.

The complexity of collaboration

Yan Wang,
Iza Witkowska

In our first blog post on *Collaboration in research* we argued that in order to share data within research collaboration 'Trust is not enough'. In this second post in the series we will elaborate on the increasing complexity of collaborations in research, since this is the main reason why trust is no longer enough. We will address the issue of complexity from three perspectives: the size of the project, the increased alienation of the partners and the different levels of maturity in data sharing practices

The size of projects

Funding agencies encourage researchers to establish collaborations with a broader scope of partners. This trend leads to large project consortia. To illustrate the complexity of larger projects we describe the [EPOS](#) and the [H2020 Integrity](#) use-cases.

[EPOS](#) concerns a pan-European effort to establish a large-scale infrastructure for the provision of open access to a variety of aspects of European Geo-science. Utrecht University coordinates one of the so-called EPOS Thematic Core Services (TCS), i.e. the multi-scale laboratories TCS. This TCS currently comprises over 90 experimental labs from all over Europe who need to share their collected data via an established infrastructure. For this purpose, the data sharing policies of the EPOS partners need to be harmonized throughout the whole infrastructure and labs needed to commit themselves to the open policies in use at the EPOS-level.



The [H2020 Integrity](#) project is developing teaching programs and tools aiming to qualify students in responsible research. The project, again coordinated by Utrecht University, has 11 partners from 9 EU countries, including 2 commercial partners. Data sharing in this project is done on the work package level. Some partners collect and share data personal data. Agreements on data sharing are made as soon as data collection will start within a work package. The data sharing issues are complex because of the different confidentiality levels of the data collected and the different levels of interest in the data (commercial vs. open science), even when the ERC [FAIR data](#) principles apply.

Alienation

Larger projects with partners from different geographic regions, industries or scientific domains also leads to alienation of the partners. Public and private collaboration has become a common practice in research projects. The participation of private organizations comes in various forms, such as providing industrial data as research input, industrial pilot of research results, prototype development and so on.

The [Dutch Monitor on Food Waste](#) project has 35 external stakeholders in the supermarkets and food producers sectors in the Netherlands. The main project partner is [Wageningen Food and Biobased Research](#), the external stakeholders have a data supply role. This project is an example of bringing together different data flows: from public open data to confidential commercial data. The project partners from the commercial side are only willing to share their data when the data infrastructure is totally secure and results are published either on an aggregated level, or stay completely confidential. At the same time public open data, possibly with a [CC-BY-SA license](#) is used, which prescribes that data resulting should be published with the same license. In addition, there are no agreed data standards among the partners, therefore the collected data may appear in different formats. Subsequent data curation and comparison efforts demand intensive work from researchers.

Also the earlier mentioned [H2020 Integrity](#) project has two commercial partners. In the previous paragraph we argued that this leads to complexity even when FAIR data principles are leading. FAIR allows different shades of Open, even closed. The example shows alienation where the non-commercial university partners in a work package are led by the Open Science principles of their institute and commercial partners are led by commercial interest.

Considering data sharing in public/private collaboration, the [TU Delft](#) inclines to follow the regulations of the private partners on the handling sensitive data. When researchers submit publication drafts with results generated from industrial data and commercially sensitive data is sensed with a possible disclosure, the publication process is paused until a resolution is made.



Different levels of maturity

European Research Council ([ERC](#)) promotes collaborative research and specifies various related requirements in its funding scheme. For instance, in the Horizon 2020 program's [guidance on gender equality](#), it indicates that geographical diversity is one of the considerations when reviewing the project proposal. All of these imply that data sharing may take place across a broader scope of partners.

Projects at the [Faculty of Social Sciences](#), Radboud University often collaborate with UMC's, brain research institutes (e.g. Donders Institute), public health organizations or commercial partners, and these research projects always involve personal data. These projects experience different maturity levels of data sharing. UMC's are usually well organized in both infrastructure and guidelines in sharing data and so are external organizations who are professional in data sharing, anonymization, and responsibility towards patients. Smaller projects partners, or Universities are less mature in safe data sharing.

The same issues arise with the data emerging in the context of transnational access programs, where personal data comes into the scope. This is where suitable policies are currently elaborated on within this part of the program.

Conclusion

The complexity of project is likely to impose a barrier for data sharing. First of all the size of the project has an impact since different data sharing policies, regulations and habits need to be harmonized. Also these differences between partners can be found to be large and even conflicting regulations may occur. Different partners have different maturity levels in data sharing practices. The latter might be an advantage since the less mature partners will be able to learn from the more developed partner, but it also creates a risk for Open Science, since it is likely that the more developed partner has implemented more restricted guidelines. These might be hard for the other partners to follow.

In the next blog post in this series on *Collaboration in research* we will expose how the research projects handled data sharing in their research workflows.

