National Coordination Point Research Data Management

Collaboration in research
Trust is not enough

A blogseries on issues and solutions in data sharing in collaboration

DOCUMENT VERSION:
MAY 2020

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.

more information: www.lcrdm.nl

Contents

5 Trust alone is not enough
5 Peter Hinrich, Jacquelijn Ringersma
7 The complexity of collaboration
7 Yan Wang, Iza Witkowska
10 How to share the data?
10 Jacquelijn Ringersma, Peter Hinrich
12 A recipe for agreement
12 Monic Hodes, Yan Wang
14 Recommendations for data sharing and support
14 Jacquelijn Ringersma, Peter Hinrich, Margriet Miedema
17 Colophon

Colophon

Collaboration in Research | **Trust is not enough**
A blogseries on issues and solutions in data sharing in collaboration

PUBLICATION DATE | May 2020

DOI | 10.5281/zenodo.3819736

AUTHORS | Jacquelijn Ringersma (WUR), Peter Hinrich (SURF), Yan Wang (TUD), Monic Hodes (UU) Margriet Miedema (LCRDM)

EDITORS | Margriet Miedema (LCRDM), Jacquelijn Ringersma (WUR), Peter Hinrich (SURF)

OTHER CONTRIBUTORS
LCRDM task group 'Policy and guidelines for collaboration on research data'
Wouter Kool (LEI), Judith Pijnacker (SURF), Jolien Scholten (VU), Iza Witkowska (UU; TU/e)

Peter Hinrich (SURF), **Monic Hodes** (UU), **Patrick Hoetink** (UT), **Margriet Miedema** (LCRDM), **Wouter Kool** (LEI), **Judith Pijnacker** (SURF), **Jan Rijnders** (RU), **Jacquelijn Ringersma** (WUR), **Jolien Scholten** (VU), **Yan Wang** (TUD), **Iza Witkowska** (UU; TU/e)

WITH CONTRIBUTIONS OF:
Barteld Braaksma (CBS), **Toine Kuiper** (TUE), **Francisco Romero** (UMCG), **Henk van der Zijden** (Hanze Hogeschool)

DESIGN | Nina Noordzij, Collage, Grou

TRANSLATION | UvA Talen

COPYRIGHT | CC-BY 4.0

LCRDM supported by

Trust alone is not enough

Peter Hinrich,

Jacquelijn Ringersma

The problem

Scientific research is increasingly being conducted in collaborative partnerships that transcend national and institutional boundaries. This is based on the notion that greater diversity amongst project partners will enhance the quality of the output. As a result of this development, collaborative partnerships are getting increasingly large and complex. Project partners need access to each other's research data and resources. This is not always simple in practice, however, due to incompatible data sharing requirements and guidelines from the various institutions, participating private partners and/or funding organisations. Legal and technical barriers may emerge, particularly with regard to confidential data or intellectual property.

Cross-institutional research is becoming increasingly important. This is the case for both researchers and institutions, as such initiatives can enhance opportunities for indirect and contract research funding. Public-private collaboration is another important factor in the advancement of the Dutch knowledge economy and the progress of scientific research. In 2017, Dutch companies spent more than 3 billion euros on research in collaborative projects with universities [ref]. According to the EU monitor, in 2015 the Netherlands received approximately 962 million euros in EU research funding [ref] (not only for academic research).

The increasing requirements that are being imposed on research and research data can result in barriers in the absence of proper coordination of rules and guidelines concerning data collaboration.

The task group

From September 2019 to March 2020, the National Coordination Point for Research Data Management (LCRDM)'s task group for 'Policy and guidelines for collaboration on research data' conducted a small study of the issues and barriers that researchers encounter with regard to data collaboration in projects where the partners are from multiple institutions. Examples include national and international inter-university projects, as well as projects involving partners from the private sector. Based on several use cases, the task group conducted an inventory of the various levels of complexity amongst collaborative partnerships, as well as of the most urgent issues when collaborating on research data. We have used the results of this inventory to make several recommendations that could support researchers with regard to data sharing dilemmas in complex research projects.

Result: A blog series

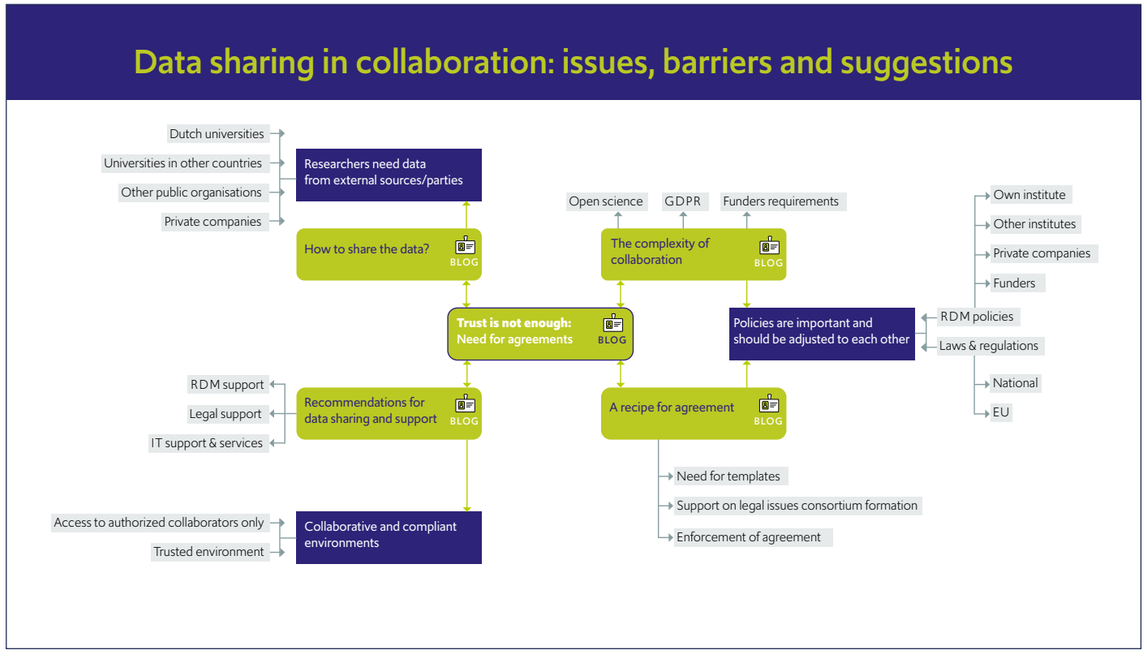
The outcome of the task group's efforts is a series of five blog posts. In it, we will present the results of the study at several different levels, in addition to noting how solutions to the issues could be pursued.

In this first post, we will outline the problem that has been identified and the approach taken by the task group. In subsequent posts, we will address the problem from a variety of perspectives: complexity of collaboration, data sharing in collaboration and institutional policies relating to data sharing. In the final blog post, we will present a number of general recommendations. They could help supporting RDM entities with regard to advising researchers at their institutions about working with data in complex collaborative projects.

The most important conclusion that can be derived from the work of the task group has already been presented in the heading of this blog post: mutual trust amongst partners alone is not sufficient for collaboration in research projects and data sharing. The conditions under which the partners will share data and other resources must be agreed in advance. Trust alone is not enough.

Approch and working method

The task group was composed of experts from various backgrounds (data stewardship, policy, project management, privacy) and institutions. We conducted a number of interviews with representatives of cross-institutional research projects. In these interviews, we asked about the guidelines encountered in their projects, whether these guidelines had resulted in barriers and, if so, how these barriers had been addressed. The experiences of the task group members were obviously used as well.



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.

How to share the data

Jacquelijn Ringersma,

Peter Hinrich

Collaboration in research is an increasingly important condition for research funding. In the first blog post [www.lcrdm.nl/trust-is-not-enough] of this series on *Collaboration in research*, we demonstrate that, in this type of collaboration, data sharing based on trust is no longer sufficient. In the second post [www.lcrdm.nl/complexity-of-collaboration], we describe how the complexity of the collaborative partnerships can influence data sharing. In this third blog post, we will present two use cases to illustrate how data were shared in the projects described, as well as the choices involved.

Proper agreements and secure infrastructure in public-private partnerships

Wageningen Food & Biobased Research (WFBR) is a partner in the National Food-waste Monitor of the Dutch Ministry of Economic Affairs. The Food-waste Monitor presents the scope of food residues in the Netherlands, based on public data about waste processing, feedstuff production, consumer waste, primary productions and renewable energy. [ref]. In addition, a supplementary initiative on 'Fighting food was-

te together' collects data from supermarkets, as well as from other companies in the food-supply chain [ref]. Companies provide their data, and WFBR uses them to quantify waste volumes and issue recommendations for reducing food waste.

This project is an example of research that would not be possible without bringing together a variety of data streams, ranging from public data sources to corporate data. For companies, the act of providing access to their sensitive data constitutes a commercial risk, and they are not willing to do so except under certain conditions. For this reason, agreements are made with the data suppliers, which agreements are documented in a three-year 'Data Transfer Agreement' (DTA). In addition, a secured collaborative environment has been created for companies, in which each participating company provides data in a protected area, with access restricted to project researchers. The DTA includes a specification of the data to be delivered, the frequency of delivery, as well as who will be allowed to use the data, during which period, and how the data may be published. In view of competitive interests, the project restricts publication to aggregated data only [ref].

Consortium agreements of the reseach funder

Wageningen Plant Research (WPR) coordinated the Horizon 2020 project Eden-ISS (Safe food in space). The objective of the project was to investigate how fresh food can be grown for and during space missions, as well as how participants in space missions deal with fresh food. An experimental station was constructed in Antarctica, where conditions are comparable to those encountered on space missions. In addition to funding from the European Research Council (ERC), the project has received funds directly from WPR, as well as from a number of participating commercial companies. In the project, the ERC prescribed the manner in which data sharing is to be handled. This was documented in the consortium agreement. In practice, this meant that parts of the data were shared among certain partners only, while other data could be shared by other members of the consortium. The data could only partly be published upon completion of the project (photographs taken from the experimental station in Antarctica are available here). Although the EU has guidelines specifying how data is to be processed and provides a template for doing so, it is difficult for the researchers themselves to enter into a contract. For this reason, the various parties were assisted by legal specialists in drawing up the consortium agreement. They also monitored whether the consortium agreement complied with any data policy that the participating institution might have.

Conclusion

Collaborative agreements between research partners are essential to safeguarding the societal objectives while respecting the desire to protect private parties. In this regard, it is important to consider the interests of the private partner (usually resulting in protection of the data), as well as the interests of the public partner (usually aimed at open science). It is essential to be able to provide research-funding entities from the public sector (e.g. NWO, ZonMw, ERC) with a template for a consortium agreement, having data sharing as one area of focus.

As demonstrated by the example of the WFBR, a secure collaborative environment is an important condition for public-private collaboration involving confidential data. Research institutions can share their knowledge and experiences in this area as well. If the requirements overlap to a large extent, demands for this type of environment can be bundled and realised jointly (e.g. within the context of SURF).
In the next blog post of the *Collaboration in Research* series, we will present an example of a code of conduct for scientific practice that specifically mentions data sharing. In doing so, we will demonstrate how such codes and guidelines can enhance clarity for researchers with regard to data sharing in large, complex collaborative projects.

A recipe for agreement

Monic Hodes,

Yan Wang

Nowadays, researchers are increasingly collaborating with researchers from other disciplines, organizations and other countries. Cross-institutional cooperation in the field of research data and resource sharing is a great adventure: different rules, customs and methods, who and what are leading? In the first three blog post in this series on *Collaboration in research* we showed that in collaboration "Trust is not enough" [www.lcrdm.nl/trust-is-not-enough], how the complexity of collaboration impacts data sharing [www.lcrdm.nl/complexity-of-collaboration] and we showed how researchers deal with data sharing in collaborations [www.lcrdm.nl/how-to-share-data]. One of the conditions for successful data sharing is a proper consortium agreement. In this fourth blog post we will further elaborate on this. Is an agreement a recipe for successful collaboration, and is there a recipe for agreement?

A good example: The Tasmanian responsible conduct of reseach policy

The purpose of the Tasmanian code of conduct is to guide institutions and researchers in responsible research practices. The policy describes the principles and practices for responsible conduct of research, for institutions and researchers. In addition, the policy also describes what agreements must be made if you work with multiple parties. When the university is involved in a collaborative project with one or more external parties or collaborating researchers, several issues must be discussed and recorded at the start of the research. You can think about conflicts of interest, project risks, ethics approval and safety clearance, intellectual property and copyright; authorship and publication, etc.

ERC consortium agreement

The European Research Council (ERC) presents larger research collaborations with a framework for a consortium agreement. A typical Horizon 2020 agreement addresses various aspects, which can include:

- Intellectual Property Rights (IPR)
- Future use and commercialisation of the project's results.
- Knowledge transfer processes between project partners

We specifically mention the elements above since these will influence data sharing practices in collaborations. Since 2017, the Open Research Data (ORD) pilot has been applied to all thematic areas of the Horizon 2020 program. However, in general the project consortia tends to follow the minimal requirements set by the EC and in large projects, partners rely on their national policy to comply with the EC requirement. The H2020 guide on open access offers three opt-out (partially or entirely) possibilities even the project is opt in the ORD pilot, to accommodate different data sharing situations. However the data sharing agreement is a bit challenging to make among partners due to the fact that there is not a unified standard across the whole EC. In some cases, project consortia are suggested to opt out the open access pilot to have more flexibility of data deposit.

Institutional data sharing guidelines

Most research institutions at the Netherlands have an RDM policy in place (see here for an overview of RDM policies in Universities and University colleges). The part that indicates the topics for making the right agreements when collaborating with multiple parties in a research project is often missing. In Wageningen University & Research data sharing guidelines are being developed. These guidelines help the researchers to

understand why data can or cannot be shared depending on the type or classification of the data and e.g. who owns the data or who funded the research. In addition the data sharing guidelines include a framework for a data sharing, which researchers can follow when they make a collaboration agreement with external parties. Although the guidelines are not yet final, and not yet included in the institutional policy they provide a handy "shopping list" which help a lot of researchers in collaborative projects.

Conclusion

Research collaboration could benefit from agreements on data sharing practices at the beginning of a project. Although the EU offers a framework for such an agreement, much is left to the integrity of the researchers and the participating institutes. An national code of conduct which specifically mentions data (and other output) sharing would be beneficial. The Wageningen University & Research concept guidelines might be evenly beneficial for researchers from other institutes.

Recommendations for data sharing and support

Jacquelijn Ringersma,

Peter Hinrich,

Margriet Miedema

Research partnerships are increasingly cross-boundary and cross-institutional, and this development is making them larger and more complex. In many cases, various institutions, private project partners and participating countries have incompatible data sharing guidelines. Barriers to data sharing often emerge with regard to legal and technical aspects. From September 2019 to March 2020, the National Coordination Point for Research Data Management (LCRDM)'s task group for 'Policy and guidelines for collaboration on research data' conducted a small study of the issues and barriers that researchers encounter with regard to data collaboration in projects where the partners are from multiple institutions. The tasks group's results are presented in the Collaboration in research blog. In the first blog post [www.lcrdm.nl/trust-is-not-enough], we outline the problem and the approach taken by the task group. The second [www.lcrdm.nl/complexity-of-collaboration], third [www.lcrdm.nl/how-to-share-data] and fourth [www.lcrdm.nl/recipe-for-agreement] blog posts address the complexity of collaborative partnerships, use cases involving data sharing, and examples of institutional or other guidelines or frameworks for data sharing agreements, respectively.

In this final post in the series, we state our conclusions and recommendations for policy makers and parties providing support in the field of Research Data Management (RDM). This will improve support for researchers, enabling them to share data in research projects without encountering barriers.

First, our conclusions

To date, the process of setting up a research collaboration has often been based on researchers' personal relationships, and too few agreements have been made with regard to the data to be shared. Our main conclusion is that simple trust between partners is no longer sufficient due to the scope of the projects and the diversity of the project partners. Obvious though it may seem, it is important for researchers to make good agreements and to document them in contracts.

Research projects are increasingly being confronted with regulations and requirements by research funders, local or other governments, and institutions. These regulations include the GDPR, copyright and database law, as well as arrangements relating to data security and the data policies of specific institutions. In many cases, requirements and regulations are not coordinated with each other, which increases the complexity of contract creation. Ensuring that the agreements made between parties are consistent with these requirements sometimes requires specialised knowledge.

In many disciplines, public-private partnerships are of major importance to both institutions and researchers. Their status, funding and societal relevance are largely dependent on these partnerships. Given the commercial interests of private parties, they often prefer restrictive data sharing. By contrast, researchers working in Dutch institutions are often inclined to share their data under FAIR or Open conditions. This can sometimes create tensions.

Large, complex collaborative partnerships have a major need for standardised storage and data-exchange infrastructure. In current practice, institutions and partners use their own infrastructures, and it is not always a given that external parties can use these facilities just as easily. Issues of data security and the access rights management for data exchange can generate complex infrastructure requirements.

Recommendations

Joint policy by research institutions

In collaborative partnerships with private partners, researchers do not always have a strong negotiating position. As a consequence of Open Science policy, university researchers are increasingly making their data accessible according to FAIR and Open Science principles. At the same time, they are facing pressure to collaborate with the business community, for purposes of funding and because of the importance of valorisation. Companies are not bound to Open Science, and they call for data protection. This places researchers between a rock and a hard place. For this reason, research institutions must collaborate in order to coordinate policies and guidelines and strengthen the position of researchers. Although we are aware that this is a long-term solution, we are requesting the Steering Group for Education, Research and Valorisation (SOOV) of the Association of Universities in the Netherlands (VSNU) and the various Dutch institutions to include this item in the strategic agenda.

Legal support

As illustrated by the use cases in the Collaboration in research blog, researchers stand to benefit from sound legal support in the initial phase of a project. To whom will the data belong, or who will be in charge of the data? How can a consortium agreement be made? Research institutions should combine their knowledge of partnership agreements, in addition to developing and sharing templates for this purpose. The National Coordination Point Research Data Management (LCRDM) could contribute to the sharing of such knowledge through the LCRDM platform.

Secure collaborative infrastructure

The various use cases demonstrate the importance of technical solutions that enable collaboration across institutional boundaries. Although technical possibilities do exist to this end, they require targeted support from the very beginning of the project. In some use cases, a collaborative infrastructure had been implemented, while project partners in other use cases shared data at the level of work packages (rather than at project level) in order to avoid the issue of larger infrastructure. Data sharing at the end of the project is becoming increasingly common (e.g. in Zenodo).

Many institutions of higher education and research offer solid support for IT infrastructure. The recent initiative of the Dutch Research Council (NWO) to establish a Digital Command Control (DCC, a central organisational unit for all data-related questions) indicates that the demand for such support is being taken seriously by both institutions and funders.

Better support collaboration from the very beginning of the project

Support for data sharing is needed from the very beginning of the consortium-formation process. Although no data curation occurs during this phase, it is necessary to consider infrastructure, agreements, authorisation for confidential data, data standards, data sharing protocols, data ethics and other matters from the outset.

At most Dutch institutions, support for Data Management is not yet so holistic. Researchers must approach different departments for different issues, and they tend to get lost in the maze of well-meaning support facilities. Legal or contract-management departments should be a structural component of the support provided for Data Management, which currently still tends to consist largely of support for data storage (IT), data curation and data preservation (Libraries). Such structural collaboration could also enable these legal departments to increase their specialisation in the data-related aspects of the agreements. Larger, more complex collaboration in research calls for a new form of Data Management support.

1) Data ownership is a complicated issue which we did not dive into for the purpose of this blog series. With 'owns' we mean the institution that has the final say over whether or not sharing the data.