h e g
Haute école de gestion
Genève

ORD capability of scientific communities
Mandate 2 commissioned by Swissuniversities

Carried out by the HEG, March - October 2021
Report V2 – 2021.10.22

Authors:

Gaia Bongi
Marielle Guirlet
Elise Pelletier
Grégoire Urvoy
René Schneider

Table of Contents

Table of Figures

List of tables

Abbreviations and acronyms

Acronym	Designation
DaSCH	Data and Service Center for the Humanities
DH+	Digital Humanities
DLCM	Data Life Cycle Management
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR	Findable, Accessible, Interoparable, Reusable
FHGR	Fachhochschule Graubünden
GDR	Gestion des données de la recherche
HEG	Haute école de gestion
OA	Open Access
ORD	Open research data
OS	Open Science
PID	Persistent Identifier
RDM	Research Data Management
SIB	Swiss Institute of Bioinformatics
ZHdK	Zürcher Hochschule der Künste

Summary

This report presents the results of a study commissioned by swissuniversities to the Geneva School of Business. Focusing on research communities, it characterizes the practices and needs of researchers in Open Research Data (ORD) and contributes to orienting the ORD support strategy. It highlights the varied definitions of "research community" formulated by researchers, as well as their lack of knowledge about their membership in a data community. The report also demonstrates that there is no clear model for research communities and that these are informally organized. It also relates the impacts of being a member of a research community on ORD practices and highlights the following findings:

- concerns, barriers and motivations for ORD are not significantly different if the researcher belongs or not to a community
- standards are more widely used by scientists who are members of a community
- the level of skill development depends in part on the support by a community, institution or infrastructure

In view of these last two results, the use of standards and the level of skill development seem to be good candidates for assessing the level of maturity of a community.

The report also presents possible levers for the creation and development of a research community, namely

- the engagement of key players
- the simultaneous adoption of bottom-up and top-down approaches
- awareness, education and training

Finally, it makes the following recommendations:

- define the concept of "research community”
- accompany the construction and consolidation of research communities
- provide the appropriate level of support for the definition and adoption of standards

The level of priority for each recommendation is summarized in the form of a decision matrix, based on the severity of the recommendation (its level of importance), the need of the target group and the effort required for its implementation.

Introduction and background

Within the framework of the National Strategy for Open Research Data initiated by its Open Science Delegation, swissuniversities commissioned the HEG and the FHGR to conduct two complementary studies: (1) to characterize the open research data (ORD) landscape in Switzerland (Erstellung einer Übersicht über die ORD Landschaft in der Schweiz); (2) to characterize the practices and needs of ORD researchers with a focus on research communities (ORD-Fähigkeit von wissenschaftlichen Gemeinschaften). This second study was carried out by HEG between March and September 2021. This document is the final report of this study.

The results should help to guide the strategy for supporting Open Research Data (ORD). By characterizing the current practices and needs of researchers, and by distinguishing these by community and possibly their granularity by community, they will help to optimize the resources devoted to Open Research Data, as well as to consolidate the strategy and the resulting action plan.

The data collected for this work was done in two phases. First, semi-structured interviews with representatives of data sharing infrastructures and of research communities were conducted. Then, Swiss academic researchers were consulted by means of an online questionnaire.

The rest of this document presents the research methodology adopted, followed by the literature review on the themes and key concepts of this work. The definition of these concepts is given in Appendix 1. It then details the results obtained from the analysis of the collected data and presents their synthesis as well as recommendations.

Methodology

1) Objectives

The main objectives of this work are:

1. Characterize the level of maturity of the research community with respect to the ORD
2. Identify the needs of the communities in terms of support for the improvement of their practices and their ORD culture
3. Characterize the granularity of these needs by community
4. Specify the ORD actors to be involved, by category of need and possibly by community
5. Make recommendations to support communities in their ORD practices

These objectives aim to answer the following research questions:

1. How can we evaluate the level of maturity in ORD of academic research in Switzerland? On the basis of which criteria? What is this level of maturity? What differentiation exists between communities? What do these results show about the culture of data sharing in the different communities? What do they reveal about the extent of the needs to reach a full level of ORD practice for the different communities?
2. What are the criteria for assessing specific needs for community advancement in ORD? What are these needs?
3. Are the needs all the same within the communities or are there differences at the sub-community level? To what sub-level (for each community)?
4. Who are the stakeholders that need to be involved? Is there a need to differentiate communities with respect to these stakeholders? How can they contribute, at the institutional level, to increasing and improving research data sharing?
5. Which kind of support to the communities will be the most efficient for improving their ORD practices? Which other recommendations about this support can be made ?

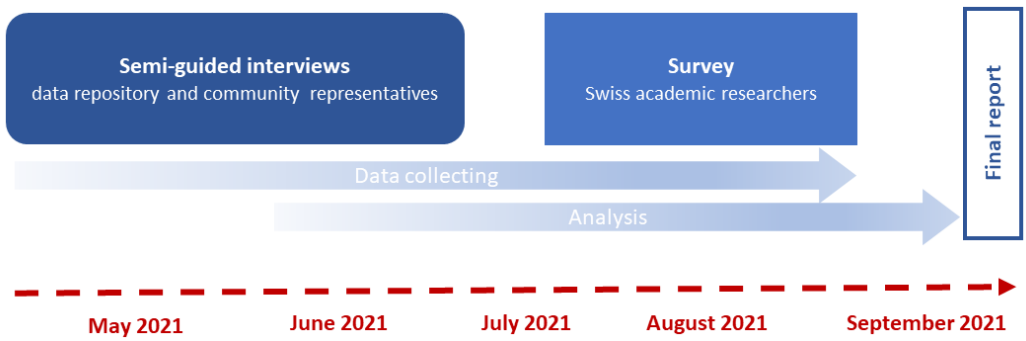
2) Methods and instruments

We used successively two main methods for collecting information:

- semi-structured interviews of data repository and of community representatives
- an on-line survey addressed to the Swiss academic researchers

Figure 1 below illustrates the general timeline of the work, with the successive and somewhat overlapping phases of data collection, analysis and synthesis for the production of the final report.

Figure 1: Timeline of the work



The objectives of both methods were to collect information about the current ORD practices and culture, the needs for improving the practices in ORD and the contributors to involve by research community.

With the semi-structured interviews, we aimed at getting some first inputs on the topic as well as useful information to help with the design of the survey (structure, content). With the online survey, we aimed at getting global and significantly meaningful results at the Swiss academic scale and at the community scale.

Comparative details on the two methods are given in Table 1 below.

Table 1: Comparative details on the semi-structured interviews and the survey

Method	Target audience	Characteristics
Semi-structured interviews	Data repository managers and communities representatives	Flexible: interview guide tailored to each interviewee from a common template; rich information; 100% answer rate
Survey	Swiss academic researchers	Provides enough information to draw significantly meaningful conclusions; little space for recording not anticipated information; answer rate lower than 100%

3) Semi-structured interviews

Details concerning the interviewees and the interview dates are given in Table 2.

Table 2: Semi-structured interviews details

Name	Position	Topics	Date
Pierre-Yves Burgi	OLOS project initiator, DLCM co-director	RD publishing and preservation, RDM services	21/05/21
Daniela Subotic and Rita Gautschy	In charge of Customer Services/Product Management of DaSCH	Disciplinary data repository and services: DasCH (humanities and social sciences)	27/05/21
Torsten Schwede	Vice-president Research University of Basel, Head of research group Computational Structural Biology Group, Swiss Institute of Bioinformatics	Disciplinary data repository and services: FORSbase (social and political sciences)	28/05/21
Georg Lutz	Head of FORS	Disciplinary data repository and services: FORSbase (social and political sciences)	17/06/21

Beate Böckem and Susanne Schumacher	Head of research affairs at Zurich University of the Arts and Co-chair of the digital council of ZHdK, head of the program Digital council	ORD culture, practices and needs of research communities at ZHdK	02/07/21
Claire Clivaz (written contribution sent by e-mail)	Head of DH+ at SIB	ORD culture, practices and needs of the DH+ research group and related research communities	22/08/21

The interview guide was provided to the interviewee(s) a few days before the interview date. This interview guide followed the same structure for all (see Appendix 2) but was tailored according to the specific position of the interviewee. All interviews were done online and were recorded for internal use only, pending the agreement of the interviewees.

The interview report was then produced and sent away a few days after the interview for validation or consolidation with the interviewee's feedback.

Besides those five interviews, Claire Clivaz, head of Digital Humanities at SIB, also sent us her written answers to questions similar to the ones asked during the semi-guided interviews.

Some of the information collected with the interviews was used to design the structure and to formulate the questions of the survey adequately.

The analysis of the collected data was made by ways of a specific table, where appropriate content from each interview final report was pasted and organized by topic (see the list of those topics in Appendix 3).

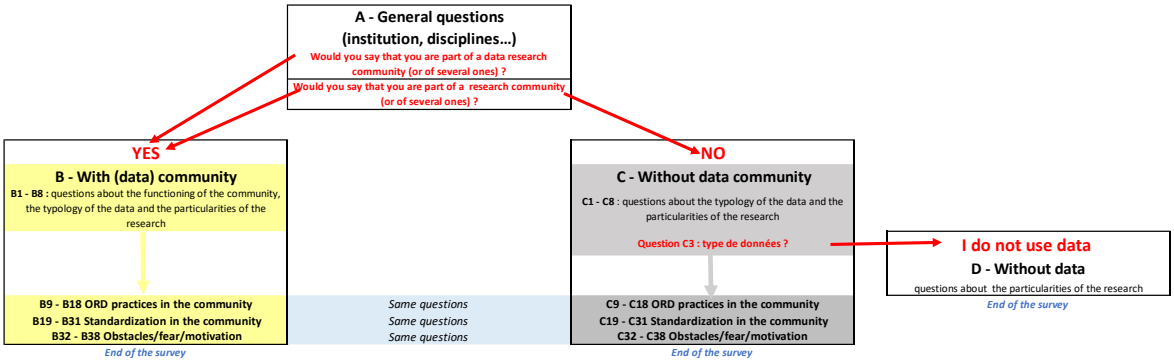
4) Online survey

Content and structure

The survey content is a mixture of closed questions with drop-down lists of possible answers and a few open questions (for additional information). The survey was designed with Qualtrics, which allows a deep analysis of the collected answers.

The structure of the survey is illustrated in Figure 2.

Figure 2: Main structure of the survey



Section A contains questions about the research affiliation and the interviewees research domain. At the end of this section, the researcher is also asked if he/she belongs to a research community, and if yes, to which kind of research community (a community related to data, to a project, to a research lab, to a research discipline or field, to an infrastructure).

If the researcher belongs to a community, he/she is directed to Section B. If not, he/she is directed to section C. The first questions in those two sections then address the characterization of the research community and of the kind of research it performs. Additional questions at the beginning of section B also characterize the community.

The second part of each section (questions from B9 and from C9) is similar and addresses:

- the ORD practices in the community
- the attitude of the community relating to the standards: the possible existence and use of standards
- the obstacles, the concerns and the motivations about ORD practices

The complete content of the survey is given in Appendix 4.

Schedule

The survey was opened on July 21, with a formal call for contributions sent to the swiss-lib list, to the informal RD CH network via his coordinator, and to a list of OS/ORD experts and chief actors in ORD.

A second call for contributions was sent on August 21, with a specific call addressed to researchers from Universities of Applied Sciences (who showed at that time a much lower answer rate than other affiliations). The survey was closed on September 8.

Contributions

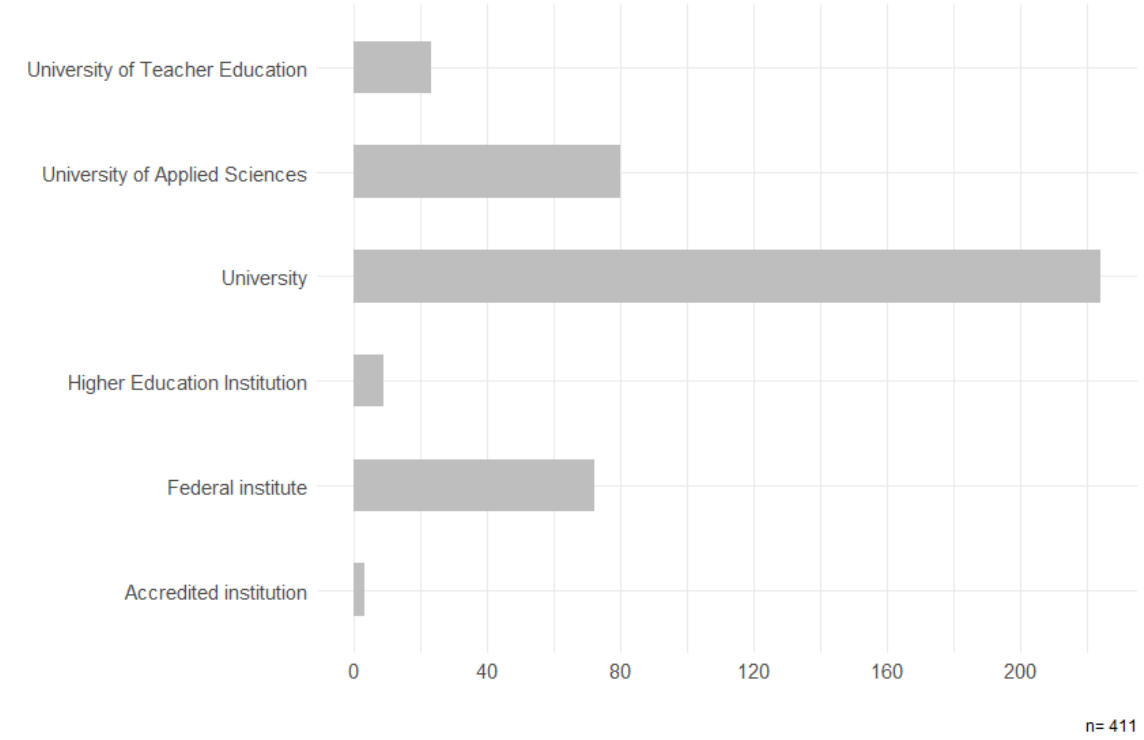
In total 185 complete answers were collected, which we consider a number large enough to draw general results, at least for questions in common for all (A section mainly). Some researchers answered to only a subset of the questions, meaning that the number of answers actually varies depending on the question. For example, this number amounts to more than 416 for the first question about the institution of affiliation, to 356 for the question about the research field or discipline and to 200 for the question relating to the description of the research community. We decided to include anyway those answers in the analysis and the results, providing that the total

number of answers is taken into account in the analysis and is made visible on the result plots if included in the report.

About two thirds of the contributors stated that they belong to a research community and were thus directed to section B of the survey after filling out section A. For the sake of significance concerning the results, we decided to redirect to section C both the contributors stating that they do not belong to a research community and those stating that they do not know (if they belong to a research community). In order to keep it more simple, we designate this group of contributors as people being not part of a community, or non-members of a community.

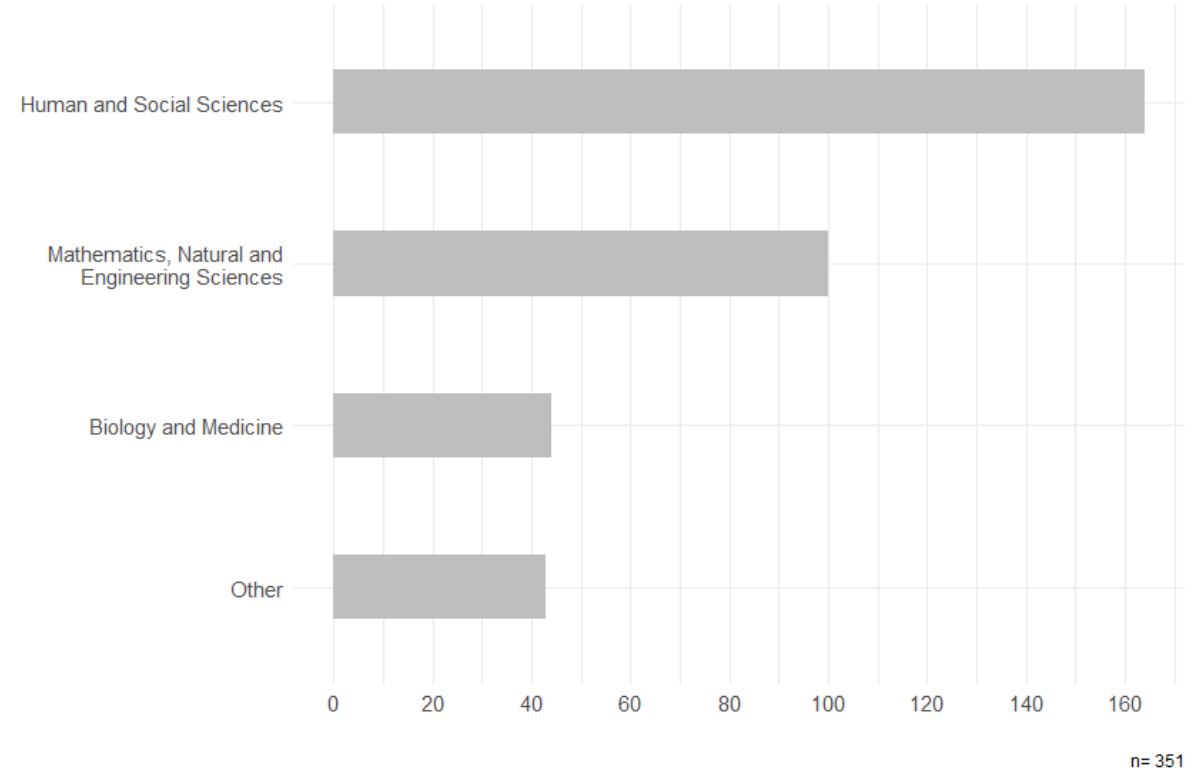
Although this survey addressed the Swiss academic researchers only (this was specified in the introduction part and in the text of the call for contributions), two librarians also contributed. Their answers could be easily identified since they provided their position in the answer to one question.

Figure 3: Institution of affiliation



Most contributions (53% of the answers) come from university researchers (Figure 3). Among those, the University of Fribourg is overrepresented with 76 contributions (18% of the overall number of contributions).

Figure 4: Research areas



Field of interest	n
Human and Social Sciences	164
Mathematics, Natural-and Engineering Sciences	100
Biology and Medicine	44
Other	43

The distribution of researchers by discipline field (Figure 4) is more homogeneous than the institution affiliation though their number is larger in Human and Social Sciences (44% of answers).

Literature review

In its *Swiss National Open Research Data Strategy*, swissuniversities stresses the need to "*establish and develop the ORD landscape in accordance with the needs of researchers and in compliance with the political framework conditions*" (swissuniversities 2021, p. 20).

Berman et al. (2013) point to the importance of community collaboration in the emergence and success of ORD practices. Indeed, "*The role of community in shaping the [Protein Data Bank] and helping it evolve into a vital resource for biological research cannot be underestimated*" (Berman et al. 2013, p.1).

In a similar vein, Copper and Springer explain that "the best policies and standards achieve little without buy-in from the research community" (Cooper, Springer 2019, p. 5)

According to Berman et al (2013), the definition of a data community can evolve over time. In the case of the Protein Data Bank, the profile of users has indeed diversified and "*the user community*

has expanded from the small group of activists who saw the need for an archive to their research to an ever-larger group or users who far outnumber the depositors" (Berman et al. 2013, p.6).

For Cooper and Springer (2019, p. 4), data communities are "*formal or informal groups of scholars who share a certain type of data with each other, regardless of disciplinary boundaries.*"

Regarding the implementation of an ORD practice, Swissuniversities establishes in its *Swiss National Open Research Data Strategy* (2021, pp. 8-10) four objectives:

1. Support for researchers and research communities to develop and adopt ORD practices.
2. Development, promotion and maintenance of financially sustainable basic infrastructure and services for all researchers:
    - Basic infrastructure and services
    - "*Data Stewards*"
3. Development of researchers' skills and exchange of best practices
4. Development of systemic and enabling conditions for research institutions and communities:
    - Reward mechanisms to be modified
    - Access to materials and resources to understand the legal, ethical and societal challenges related to the production, access and use of data
    - Policies and implementation measures must be in line with international standards and actions

According to Cooper and Springer (2019), data sharing and thus ORD practices are inseparable from data communities. The authors highlight three characteristics for their success: bottom-up development, absence or mitigation of technical barriers and standards.

As far as standards are concerned, the *FAIR Data Maturity Model* (Research Data Alliance FAIR Data Maturity Model Working Group 2020) puts forward recommendations for the FAIR use of data, i.e. data that is discoverable, accessible, interoperable and reusable by machines and humans alike. The following standards, among others, can be developed:

- Metadata standards
- PIDs
- Licensing standards
- File format standards
- Documentation standards

Cooper and Springer (2019) also refer to the issue of FAIR standards and principles for successful data sharing: "[Many] *data sharing advocates have embraced the FAIR data principles* [...] *as the standard benchmark for data sharing success*" (Cooper, Springer, 2019, p.2).

I. General comments on "research communities"

As observed in the literature review, research communities are identified as key actors for the development of OS and of ORD in particular. The strategy of swissuniversities is logically aligned with this observation. During the interviews, the importance given to these research communities and the need to involve them in the development of the ORD also seem unanimous.

> *"You need to invest a lot in community buildings. You have to understand all these projects as community-building projects and that is the hard part."* (Lutz 2021)

> *"One of the preconditions for OS to work is that we have understanding for the motivations and the limitations of all players in the communities. Otherwise, there is no motivation to share anything."* (Ackermann, Schwede 2021)

However, despite this shared observation, the perimeter of these communities remains to be defined, particularly in the Swiss environment. Indeed, the research communities seem to correspond to a more transversal idea of research with interdisciplinary or international approaches.

> *"In the interest of establishing and passing on best practices, it is crucial to strengthen knowledge exchange among research communities through the creation of networks between institutions and between data stewards or similar positions within higher education and research institutions."* (Swissuniversities 2021)

While definitions exist in the literature, it is important to agree in practice as well, especially with a view to building on these communities to develop the culture and practice of ORD.

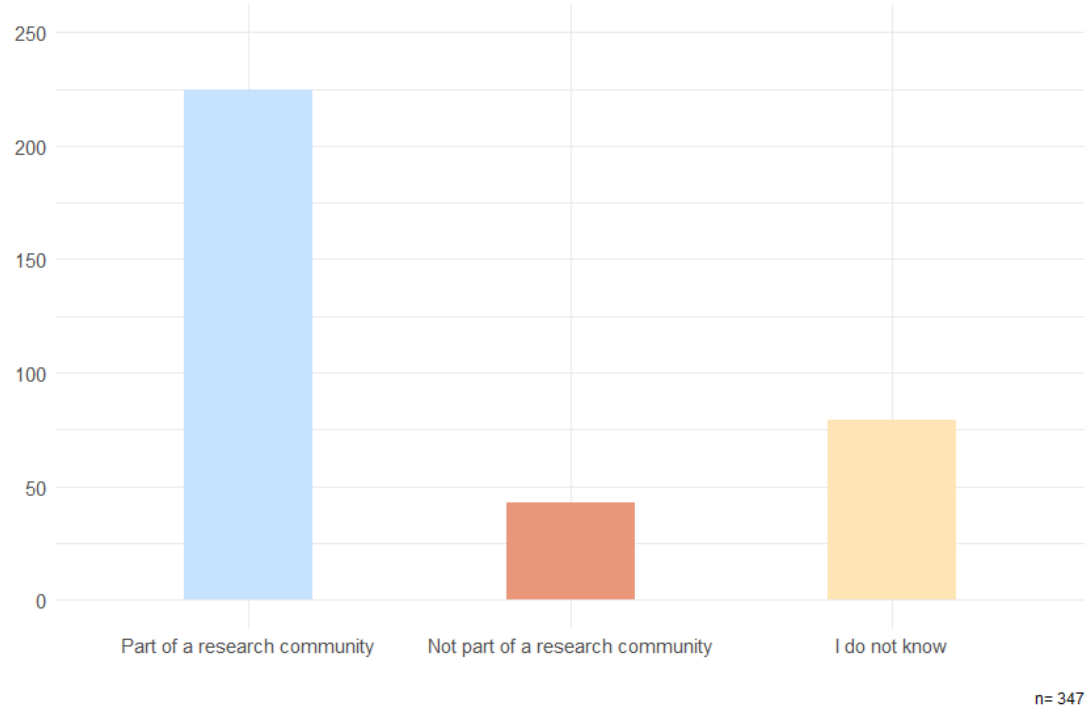
A. Definition of "research communities"

The semantics around the notion of "research community" do not yet seem to be stabilized. One of the challenges of this mandate was to ensure that the discourse between theory and practice was consistent.

1) Being a member or not of a research community

The notion of "research community" is present in the field. During the interviews, this notion was widely used in the exchanges without the tangible existence of these communities in the Swiss research environment being questioned. The survey confirmed this reality with the answers to the question "*Would you say that you are part of a research community (or of several ones)?*"

Figure 5: Membership in a research community



As shown in Figure 5, 230 researchers, or 65% of the researchers surveyed, declared belonging to a research community. The precision on the possibility of belonging to several communities was initially added in order to limit the negative answers based on the singularity of the question. However, it turned out that a very large number of researchers cited several communities when asked to specify to which community or communities they belonged. One could easily conclude that belonging to one research community increases the possibility of belonging to several communities. However, the vision of these research communities seems rather blurred, as shown by the 23% of researchers who answered that they did not know whether they were in a community or not. This blurred vision leads to confusion that limits the interpretation of some of the responses. This semantic problem was identified during the interviews and therefore investigated in the survey.

2) Different types of communities

The interviews offered the opportunity to propose an initial vision of these research communities with a marked focus on data, which logically made sense in a reflection on ORD practices. In the interview guide (Appendix 2), along the question "Among your institution users [or infrastructure users], do you identify specific research data communities?", the following definition was provided:

> *"By research communities, we mean networks of researchers sharing the same (type of) data, the same data sharing tools and infrastructures, the same ORD practices, and the same culture of Open Science, but not necessarily being from the same research discipline."*

The objective was to investigate if this definition corresponded to what was found in the field and if not, to open the discussion. At the end of the five interviews, none of the interviewees really contradicted the definition, but in each case supplements seemed necessary because, as Torsten Schwede points out, research communities have a "broad spectrum" (Ackermann, Schwede 2021). Georg Lutz reminds us, for example, that many researchers do not work on data as digital objects:

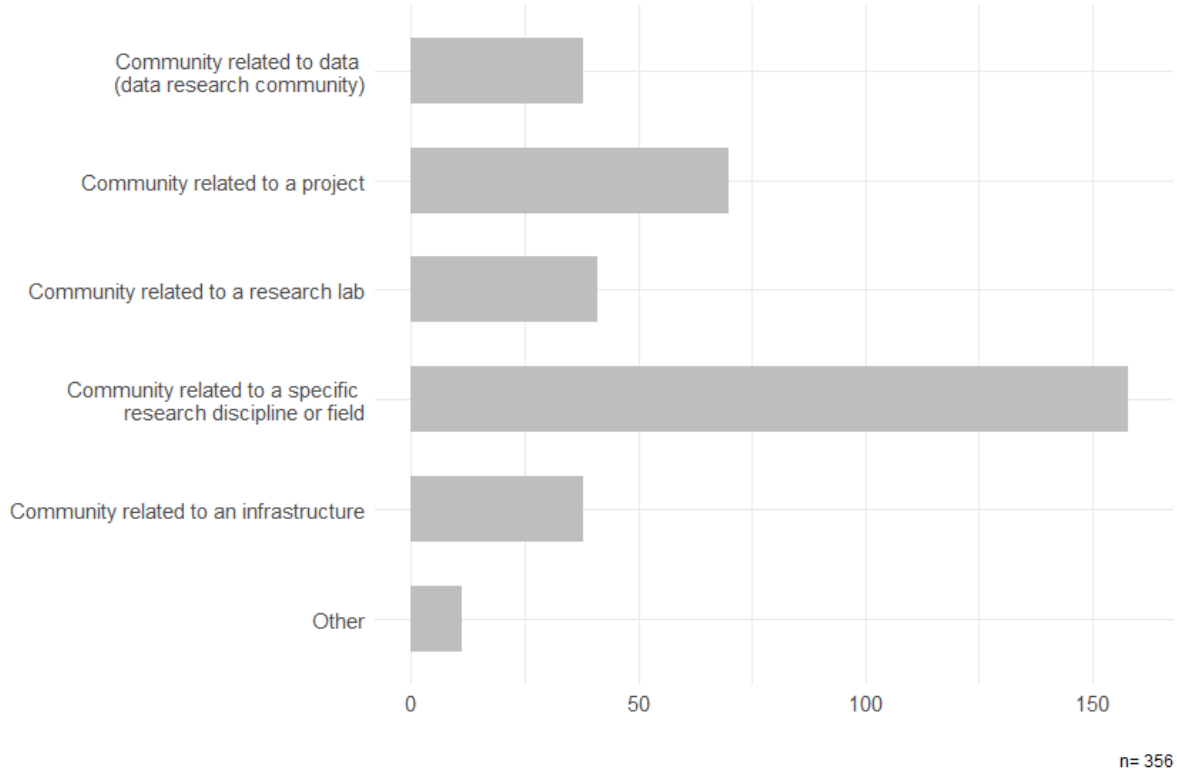
> *"Many researchers do not create data [...]. They use manuscripts, other kind of information than datasets that you can easily store."* (Lutz 2021)

This reality, even if marginal, limits, for example, the initial definition that focused mainly on the use of data. Overall, during the interviews, communities are described as groups of researchers who organize themselves around projects, infrastructures or research methodologies in which data can be similar but without this seeming to be the central point.

> *"They are related to different kinds of infrastructures and to the very different and various kinds of research methodologies... It is very much not only discipline-related but also project-related. "* (Böckem, Schumacher 2021)

This variety in the creation of research communities is reflected in the survey results. Indeed, 65% of the researchers answer that they belong to a research community, but the description they give is variable.

Figure 6: Characterization of the research communities



As shown in Figure 6, the researchers who responded to the survey identify their research community mainly in terms of a discipline or research area. However, the spectrum of research communities remains clearly rather broad. In addition to the definitions proposed in the question, we find in the answers communities related to research methodology, tools (software) or even language (in free text form). According to these different results from the survey, we note that the definition of these communities remains variable.

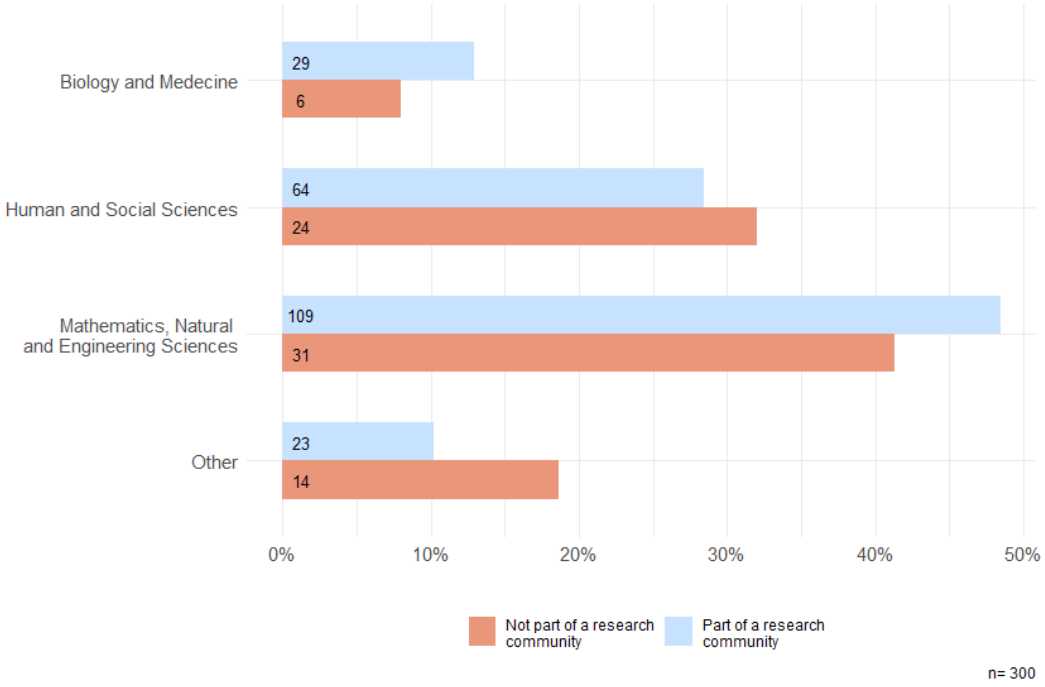
3) An imprecise perimeter

A significant part of the researchers (23%, 80 researchers, see Figure 5) state that they do not know if they belong to a research community. Moreover, among the 13% (44 researchers) who answered that they did not belong to a research community, several explicitly noted that they did not know what a research community was, and among these same 13% who did not belong to a research community, about 15 paradoxically declared that they asked "*their community*" for help with certain ORD practices. All these elements show that the scope of research communities is variable and even rather imprecise.

Although, as in the initial definition, it seems important to specify that communities are not solely disciplinary, in practice, the disciplinary approach remains the majority. Moreover, when interviewees were asked to give examples of research communities and to rank them according to their level of practice in ORD, it was mainly disciplines that were cited.

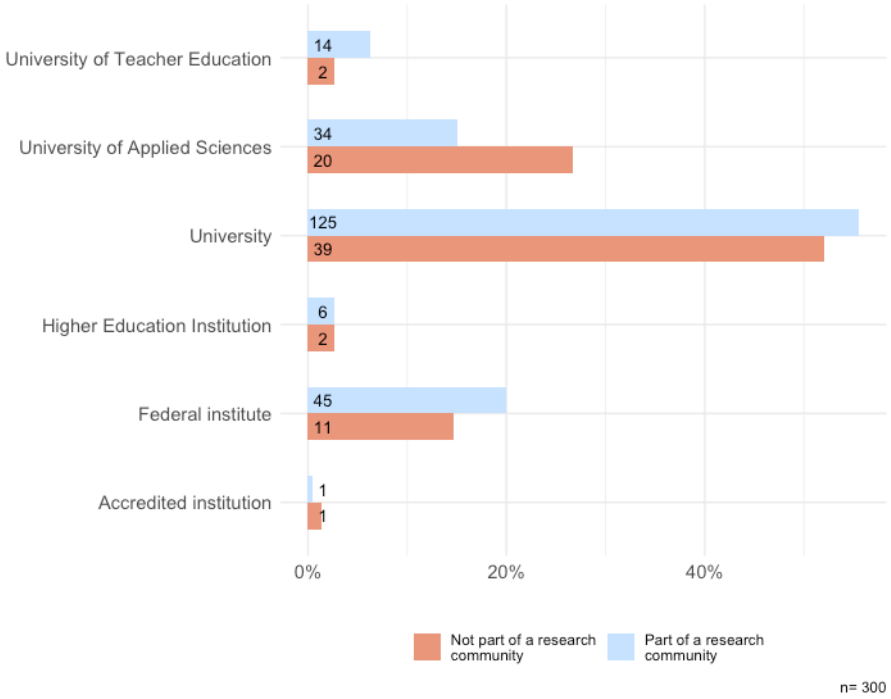
Yet the survey results (Figure 7) show that discipline affiliation is not conclusively linked to participation in a research community.

Figure 7: Research area or discipline



To put it another way, discipline or field of study does not really seem to influence whether or not one belongs to a community. Similarly, the institution of affiliation does not seem to be a conclusive factor, as shown in Figure 8.

Figure 8: Institution of affiliation



It may be noted that researchers from HES say that they are (proportionally) fewer members of a community than those in universities. This slight observation can be counterbalanced by this comment from the survey:

"Being associated with a University of Applied Sciences, with the strong links and expectations between research and an area of professional practice, reinforces belonging to a specific research community, in that area of practice" (response to survey question in free text form)

The attempt to define research communities according to an institutional and/or disciplinary perimeter is therefore not conclusive. However, in order to work with these communities and even to possibly initiate them, it is important to understand how they are organized.

B. Organization of these communities

As discussed above, from the interviews, research communities appear to be quite polymorphous. For this reason, we thought it would be interesting to study whether they have a factual and tangible existence, and thus eliminate the hypothesis of simple informal networks of researchers that would be much more difficult to reach and support in their ORD practices.

Two questions in the survey focus on this aspect, one about how the community is organized and another one about the existence of a website or community pages.

1) Organizational structure

Figure 9: Organization of research communities

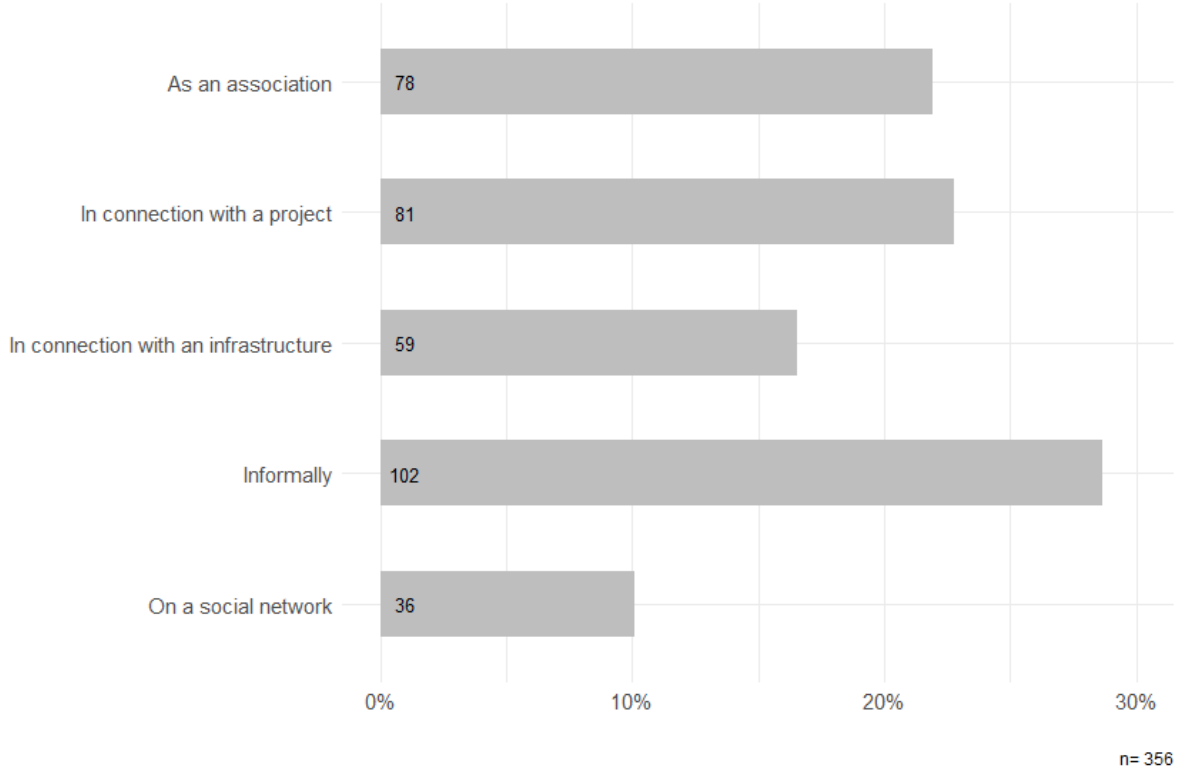


Figure 9 shows that the mode of organization of communities is quite heterogeneous. Logically, it is stated that some communities are organized around projects or infrastructures, which can both define and structure them. With a proportion of responses of 22%, the association is a significant mode of organization. Social networks are also mentioned but do not seem to be the main mode of organization. On the other hand, more than a quarter of these communities are organized informally, which obviously makes them more difficult to reach and mobilize. One researcher comments on this issue as follows:

> *"In law, there is actually no talk of a "scientific community". Nevertheless, there are connections between colleagues in the same field in Switzerland and beyond the national context. These connections are partly informal, partly institutionalized."* (response to the survey question in free text form)

Here appears the idea of belonging to several communities that are more like informal networks with no real boundaries.

The question on the existence of a website or web pages had two objectives. First, we aimed at finding out whether the communities rely on at least one tangible tool (which can already be an indicator of the development of their internal organization) and second, it allowed for the collection of access points to additional information on these communities (Appendix 5).

While 28% of the researchers answered that their communities do not have a web site, the list of web pages mentioned by the researchers illustrates the wide variety of organizational modes. Indeed, many pages are hosted on institutional sites and point directly to departments (e.g. https://www.unifr.ch/ecopol/fr), or sometimes to research centers (e.g. https://www.zne.uzh.ch/en.html). Some communities, especially those organized in associations, have their own site (e.g. https://www.manep.ch/). Others use social networks (e.g. https://www.linkedin.com/groups/13763674/).

It is therefore difficult to find "typical" structures for these communities. Even the number of members ranges from 5 to "*enormous to the point we can't count*." (response to survey question in free text form).

Yet, while communities are difficult to delineate, some informal exchanges seem to shape certain behaviours with tacit agreements.

2) Sharing culture

Indeed, some communities seem to have informal agreements:

> *"Moreover, no structural biologist will review a paper which describes a structure not in the archive. No editor will even accept the paper."* (Ackermann, Schwede 2021)

Torsten Schwede defines them as "gentlemen agreements" and for him :

> *"It is also good for the disciplines to have a gentleman agreement (or other) in the community saying that nobody publishes before the first study author (the one who generated data) has published [...]. Otherwise, nobody would share. And that works surprisingly well."* (Ibid.)

Some communities would be more inclined to share their data than others with a more developed "culture of sharing".

> *"Certaines communautés (comme les sciences de l'environnement) sont beaucoup plus en avance sur le partage des données. Cela est aussi lié à la difficulté d'acquérir les données (efforts, temps, budget nécessaires). En physique des particules, par exemple, les machines comme les accélérateurs sont tellement chers que la tradition est de partager pour ainsi ne pas privilégier uniquement certains chercheurs. "* (Burgi 2021)

The example given by Pierre-Yves Burgi highlights that this habit of sharing data is also linked to a specific context and to specific needs. Other examples from the interviews show that this can happen over time:

> *"Our group] does all steps around the Data life cycle [quite general in bioinformatics, and has been going on like that for a long time]: reuses, generates, processes, preserves and gives access"*
>
> *"This is quite general in the domain of bioinformatics. This use and practice of the whole data lifecycle has been going on for a long time"* (Ackermann, Schwede 2021)

Or be present in newer communities as well:

> *"The younger communities are among the newly established communities. They rely stronger on infrastructures, tools or digital practices than communities like musical historians."* (Böckem, Schumacher 2021)

On the other hand, one researcher indicated in the survey that he was not part of a research community because his research area was new ("*partly new research areas"*). The recency of the community or the discipline does not necessarily imply this culture of sharing. It is rather the tools and infrastructures used that seem to be at the origin of this willingness to share. To this, we can also add the research methodology which is based on the reuse of data, as Torsten Schwede explains that "all his career has been based on reusing open data". Another example is given by Beate Böckem and Susanne Schumacher:

> *"Other forms for reusing data, related to the very strong developed collaborative mode that is practiced in the arts: you are using an artistic method, or design method, and you take data, continue to write on them, to draw on them, to program on them, to bring a new view on what is there and to document this, to bring this to a piece of art or a piece of programming "* (Böckem, Schumacher 2021)

As a result of this culture of sharing, the tools can be designed to be interoperable and open.

> *"In the field of computer music/electronic music: researchers at ZHdK are quite famous to contribute to community, to deploy patches or plug-ins for programs. They are always open source code that will be brought into community and reused in the community."* (Böckem, Schumacher 2021)

We note that the formation of a community and the culture of sharing that results from it seem to be linked to initial needs in the use of certain tools or the choice of methodology. Other factual elements linked to the typology of research (scope, typical time scale and types of data types) may appear to be correlated with the fact of belonging to one or more communities.

3) Link between research typology and community organization

Several aspects related to the typology of research appeared repeatedly in the examples given during the interviews: the scope, the duration of the research and the typology of the data. They seem to be related to whether one belongs to a community or not. Therefore, these elements were used in the survey to investigate their possible link to the membership in a community.

Scope of research

The issue of the scope of research is linked to the nature of research projects which increasingly involve partnerships across borders. Researchers involved in these European or international projects must comply with the ORD principles of their partners, as confirmed, for example, by Beate Böckem :

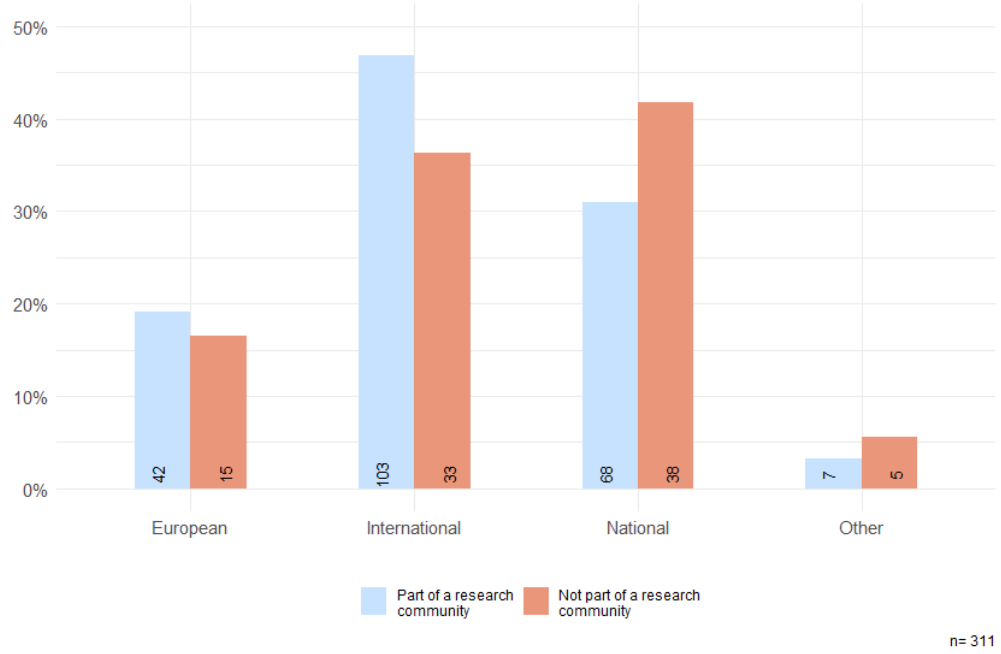
> *"In the national and international base, we are strongly related to the DORA declaration[1], to the Florence principles[2], to the Frascati manual[3], so to all the international/national standards for quality and research "* (Böckem, Schumacher 2021)

Certain practices follow logically from these partnerships:

> *"Practices that are related, fixed and rooted down to the national/international background, like the research catalogue [4]."* (Ibid. 2021)

The scope of the research seems to influence the practices of the researchers, but more broadly the membership in a community. Thus, Sonia Ackermann emphasizes the international scope of communities: "*it is an international game because research communities are international.*" (Ackermann, Schwede 2021). Indeed, the results of the survey confirm this observation as shown in Figure 10:

Figure 10: Scale of research



Researchers who report doing research at the international level are more likely to report being members of communities. This result also extends to respondents doing research at the European level.

[1] San Francisco Declaration on research evaluation: https://sfdora.org/read/
[2] The Florence Principles: https://elia-artschools.org/page/Florence-Principles-On-the-Doctorate-of-the-Arts
[3] Frascati Manual: https://www.oecd.org/fr/sti/manuel-de-frascati-2015-9789264257252-fr.htm
[4] Research catalogue (an international database for artistic research): https://www.researchcatalogue.net/

The time scale of the research

The duration and timing of some research projects were mentioned during the interviews as an aspect that makes some collaborations difficult:

> *"Reward issue: who gets the paper? When can you publish ? Who gets credit? Hard to solve if communities [collaborating on the same research project] work on different time scales."* (Ackermann, Schwede 2021)

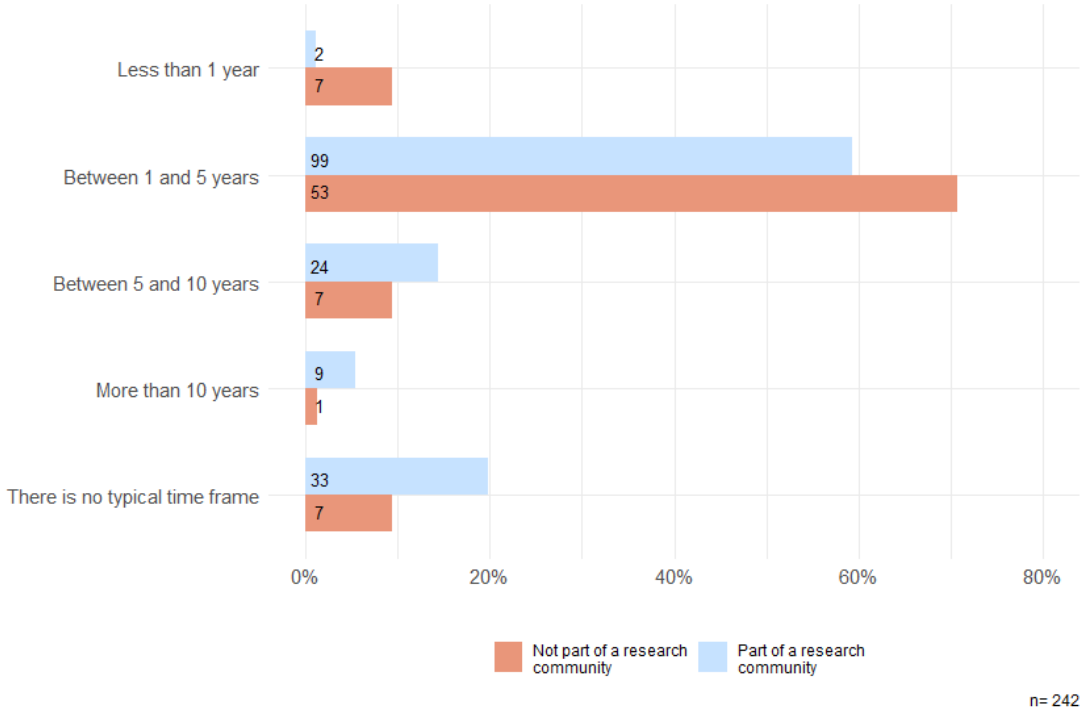
The time scale is also mentioned by Rita Gautschy for archaeological research as a factor influencing the willingness to share:

> *"With Archaeology, there are very long time projects. The projects are often never-ending. When a project is really finished after a few years, people are more willing to share their data because it is finished for them too."* (Gautschy, Subotic 2021)

According to Georg Lutz, this could also be an excuse not to share one's data (regardless of the research area).
*This example concerning archaeology illustrates the fact that a very long research time can limit the willingness to share. The results of the survey, on the other hand, tend to show that researchers who state that their research projects typically last at least five years are better represented in the "*member of a research community*" category (*

Figure 11). For researchers whose research typical time scale is shorter than five years, it is the opposite.

Figure 11: Research typical time scale



However, in the comments to the survey, a Pollination Ecology researcher remarked:

> *"Data may not be public until all the research groups involved have published, which may take years".* (response to survey question in free text form).

As with archaeology, the duration of the project makes it difficult to envisage a real culture of sharing.

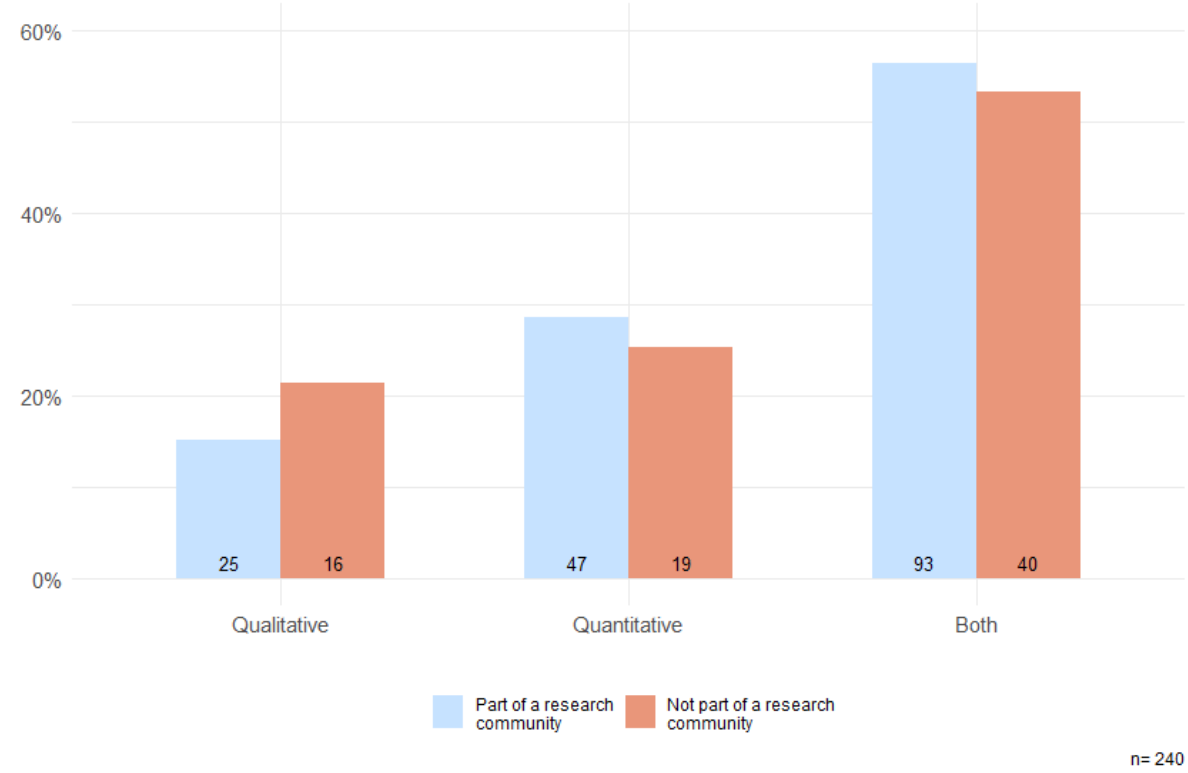
Type of research and data typology

The type of research (qualitative, quantitative) and the typology of data used are other elements discussed in the interviews and in the survey.

> *"In qualitative research, people may not yet have a culture of data sharing (and data are super hard to anonymize)"* (Lutz 2021)

In the survey, community members more often describe their research more often as quantitative than as qualitative. It is the opposite for non-community members. That being said, almost or more than half of the respondents say that their work falls under both qualitative and quantitative research.

Figure 12: Nature of Research



Two survey questions on the type of data handled and the typical size of the datasets provide a more detailed characterization of the data used.

Figure 13: Data type

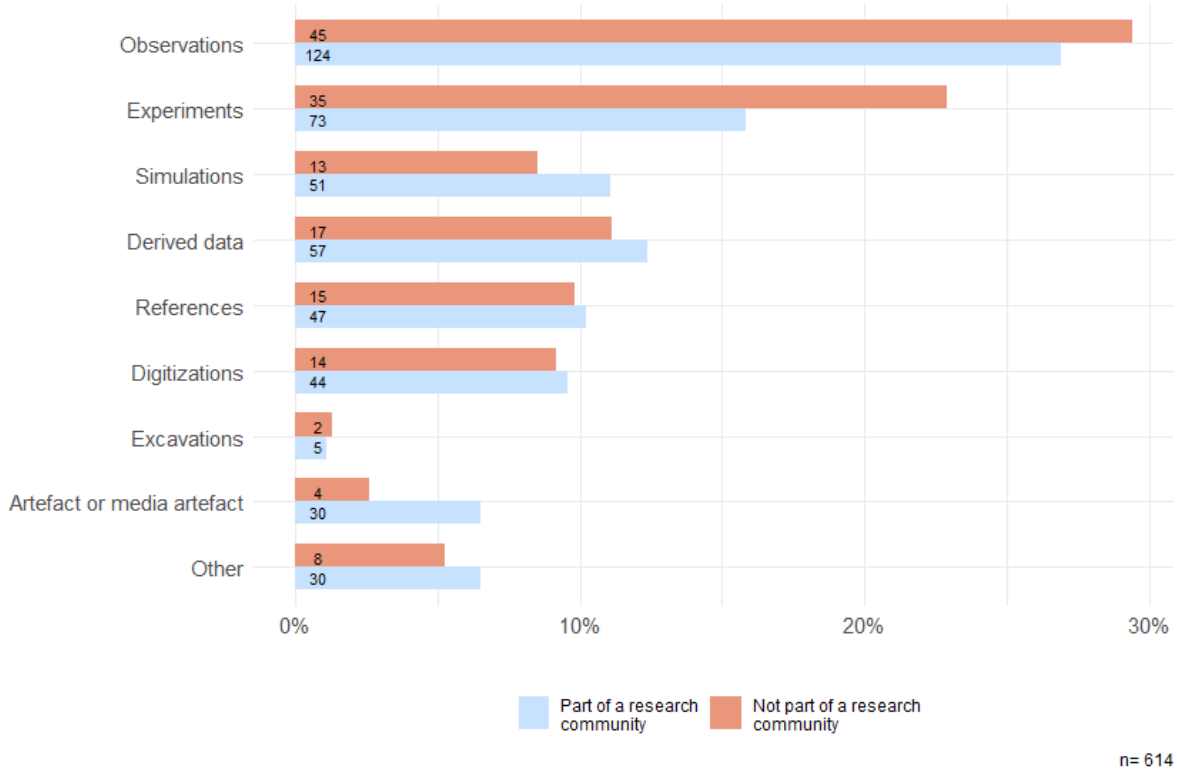
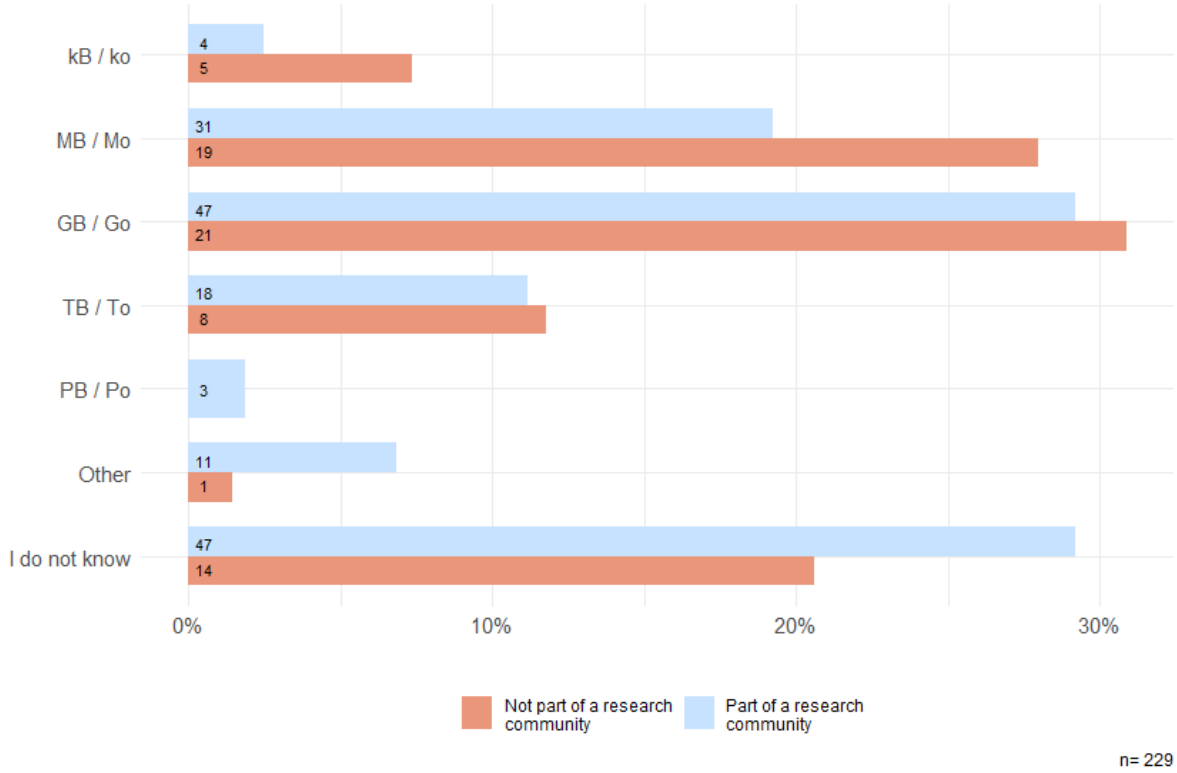


Figure 14: Data size



The data types most cited by the researchers are observational data and data resulting from experiments (Figure 13). Non-members of research communities declare that they proportionally handle more of these two types of data than members of communities. The data of the latter are more spread over the various categories.

The typical size of the data files most frequently selected is GB / Gb, followed by MB / Mb (proportionally especially by non-members), and the answer "I don't know" in a proportion close to 30% for community members. It is possible that this high proportion for community members comes from the greater variety of data typology, which also translates into a greater variety of typical file sizes, and thus to the more frequent choice of this answer.

The results in this section on the organization of communities highlighted that the European or international scope of the research, a typical time scale of at least five years and, to a lesser extent, the more quantitative type of research are, in particular, characteristics which seem to be correlated with the researchers' affiliation with one or more communities.

C. Focus on two particular profiles: data communities and "data-less" researchers

1) Data communities

Our interview with Torsten Schwede and the article by Berman et al. (2013) describe in detail the development and functioning of the "*Protein Data Bank*" community (presented and discussed later in this document). Based on this information, one of our goals was then to investigate whether the concept of a "*data community*" is found elsewhere in the Swiss academic research landscape and to characterize its practices. Our survey also seeks to characterize, at the other end of the spectrum of this polymorphous typology of communities, researchers (and their individual practices) who state not using data.

Out of the 230 researchers who said they belonged to a research community, only 38 said they were in a data community. In addition to this low representation rate, the variety of responses and some inconsistencies blur the picture of this type of community. In addition, like many other respondents, these data community members report belonging to several communities, which makes it difficult to read the results in a linear way. One of the elements of relevance chosen for this analysis is simply the ability to clearly name the community or communities mentioned and/or to give a link to a relevant web page. With these criteria, we obtain less than 20 exploitable answers that allow us to identify about twenty potential data communities (Appendix 6).

All three major disciplinary areas (humanities and social sciences; mathematics, natural sciences and engineering; biology and medicine) are represented. The scale of the research is at least national and mostly international (11 researchers). The data handled are quantitative or quantitative and qualitative (but not only qualitative). We also observe that about a quarter of these communities is informal and another quarter is project-related. These elements are more to be seen like indications for possible future studies of these data communities, since the small answer rate for this kind of category and the vagueness of the definition of research communities among the interviewees limit the possibility of a more in-depth characterization.

2) Researchers "without data"

At the other end of the spectrum, during the interviews and some discussions prior to this mandate, profiles of researchers who would not use data regularly emerged. For example:

> *"Many researchers do not create data [...]. They use manuscripts, other kind of information than datasets that you can easily store."* (Lutz 2021)

The survey was therefore designed to identify potential research practices without any data use. In the end, only three researchers stated that they did not use data in the survey. Two of them are from the field of economics and law and the third from musicology. Two of them declare that they follow individual research practices. Two of these researchers declare that they publish in Open Access, thus demonstrating that at least they are familiar with and practice this older component of Open Science (than ORD). From these observations, the "myth" of the data-less researcher does not seem to correspond to a broad reality, despite this comment from a musicology researcher:

> *"Most of my research is philological/historical so it is expressed in texts and not in "data""* (response to the survey question in free text form)

Another aspect to consider is that the survey, because of its digital format, may not have reached enough representants of this profile of researchers.

II. Role of organizing into "research communities" for ORD practices

Having attempted to define 'research communities' more precisely, the aim is now to find out whether and how those communities may influence the behaviour and the practices of researchers in relation to ORDs. To do this, we looked at three aspects:

- Can belonging to a community alleviate concerns, minimize obstacles and increase the motivational factors felt by the researchers?
- Does community membership play a role in the use of standards by researchers?
- Can belonging to a community be beneficial to the development of researchers' skills in ORD?

A. Motivating factors and barriers for researchers in the development of ORD

The interviews and the literature review allowed us to identify various elements that could positively or negatively interfere in the development of ORD practices. These elements were divided into three groups:

- Concerns for less tangible elements related to more subjective fears;
- Barriers to tangible items directly impeding data sharing;
- Incentives for encouraging (or even forcing) researchers to share their data.

As Torsten Schwede points out:

> *"One of the precondition for OS to work is that we have understanding for the motivations and the limitations of all players in the communities. Otherwise, there is no motivation to share anything"* (Ackermann, Schwede 2021)

It seems essential to take these factors into account to better understand how to handle them.

1) Researchers' concerns in ORD practices

Interviewees often mentioned 'fears' that seemed to hinder the development of the researchers' ORD practices. The synthesis of these fears mentioned in the interviews allowed setting up the series of proposed answers for the part of the survey on concerns. Table 3 illustrates these proposals using examples from the interviews.

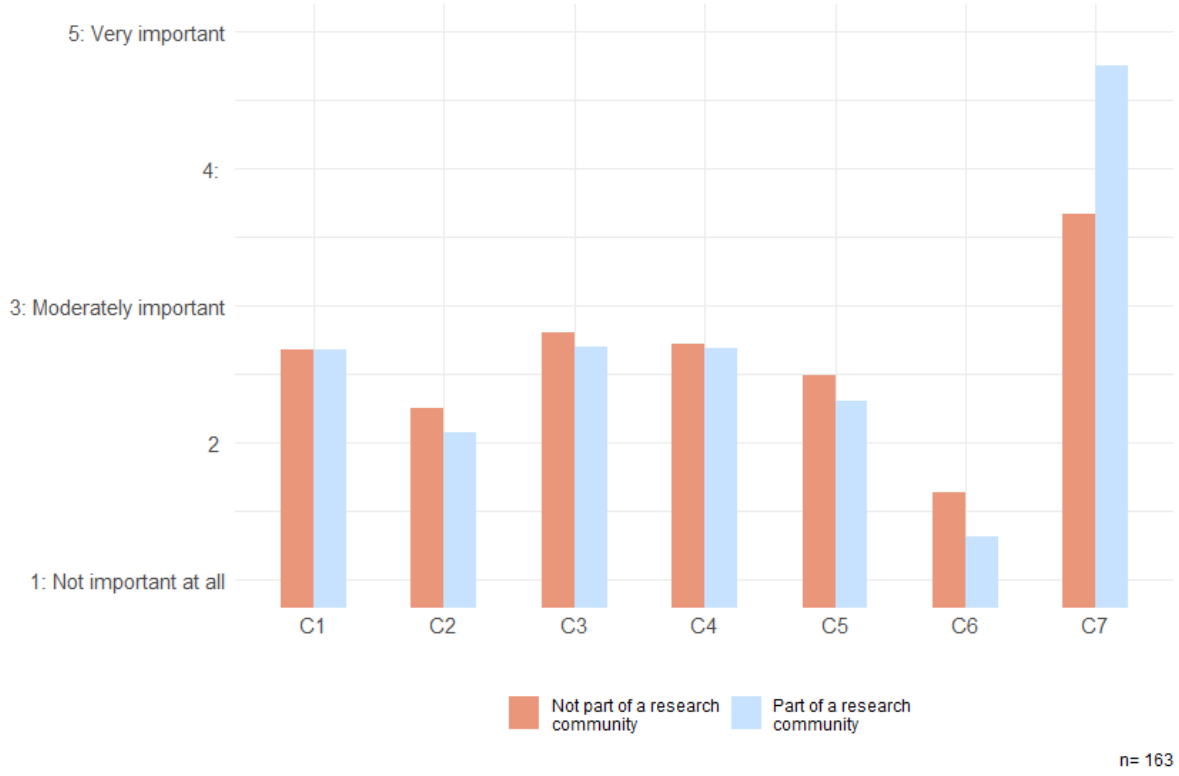
Table 3: List of concerns with excerpts from the interviews

Excerpts from the interviews	**Proposed answer in the survey**
"Craintes sur la réutilisation des données et le risque que ses données soient mal utilisées" (Burgi 2021)	I am concerned about my data being not well handled and used.
"They want time for their publications; make sure they can use the data their own way; to keep control on their data" (Gautschy, Subotic 2021)	I feel that I might lose control on my data I am concerned not to be properly acknowledged as the data producer.
"In the medical domain (personalized health data network): if "I share my data": • *It would take something away from me*	I am concerned that someone else publishes results from my data before I do.

• *people may see errors in it and that would be embarrassing for me* • *people may reap benefits from my data without me getting anything out of it" (*Ackermann, Schwede 2021)	
"L'acquisition des données représente le gros du travail scientifique et les donner gratuitement serait équivalent à travailler pour rien" (Burgi 2021)	I think that the benefit would be too low as compared to the efforts.
"Si les données peuvent être facilement produites, il y a moins de motivation à les partager. " (Burgi 2021)	I do not see the point.

Researchers could rate the importance of their answers on a scale of 1 to 5.

Figure 15: Concerns about compliance with ORD practices



C1: I am concerned about my data being not well handled and used
C2: I feel that I might lose control on my data
C3: I am concerned that someone else publishes results from my data before I do
C4: I am concerned not to be properly acknowledged as the data producer
C5: I think that the benefit would be too low as compared to the efforts
C6: I do not see the point
C7: Other

Figure 15 shows that none of the items mentioned are considered really important, but they all appear as existing concerns. Only C6 "*I do not see the point*" seems less relevant. The category C7

"Other" was left as free text to allow researchers to indicate concerns that would not appear in the list. Logically, researchers who took the time to complete this free text rated their concern as very important. In all, twelve additional answers were given, which can be grouped under five main categories:

- Problems related to the particularities of the data that make them impossible to share
  "Context information is crucial for ethnographic projects; if data is anonymised/masked; it loses its potential use value both for me and further researchers"
- Problems of data anonymization, especially in health
- Lack of support for data management
  "Main concern: how to make sensitive data accessible in an appropriate way without having enough time and financial or technical support from the university"
- Technical issues
  "My concerns are mostly technical: current ORD platforms are too primitive for my needs"
- Time issue

Those answers are very factual and echo the statements on the obstacles listed in the following question. The concerns, although they exist, do not seem to have very strong effects on ORD practices. Above all, there is no significant difference between the concerns of researchers inside and outside the communities.

2) Obstacles to the development of the ORD

Based on the same principle as the concerns, a list of barriers was developed from the literature review and from the interviews.

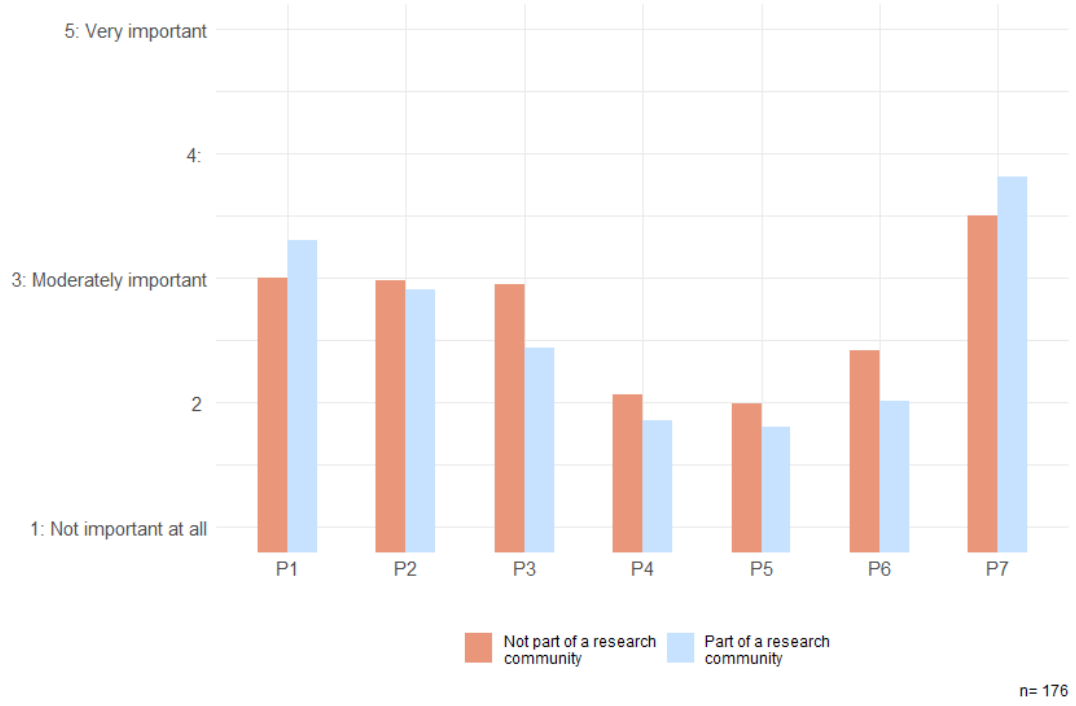
Table 4: List of obstacles with excerpts from the interviews

Excerpts from the interviews	**Proposed response in the survey**
"As soon as they are individual data it gets very tricky (in education sciences, data collected from kindergartens). Video tapes: almost impossible to anonymize. How to make available sensitive data and secure access to data and documentation?" (Lutz 2021)	There are restricted conditions on the use of the data that I handle (and thus on sharing them).
"Medical domain: not much time to do research. They are afraid that if they make their data available, scientists with more time for research just steal their glory and treat them as someone delivering samples and observations." (Ackermann, Schwede 2021)	I do not have enough time to prepare the data and to make them accessible and reusable.
*"If the time scale is long, people are less willing to share their data, they wait for the project to be finished. "(*Lutz 2021)	The typical time scale of my research is very long (and I will not share my data before the project is finished).
"Sometimes data are extremely complex, and preparing or arranging them for re-use takes some skills. They miss the knowledge, the skills" (Gautschy, Subotic 2021)	I feel that I do not have the right skills to prepare my data in a way that makes them accessible and reusable.

"Magnitude of the data. Some disciplines produce terabytes of data. How to handle them?" (Ackermann, Schwede 2021)	I think that no appropriate technical tool is available for sharing my data and making them reusable.
"People need more skills and a better education at the individual scale. The education part is the most important. You can always offer more, but without education, providing more services would be useless." (Gautschy, Subotic 2021)	I do not have access to the necessary support or coaching that I would need to make my data accessible and reusable.

Table 4 includes the more factual elements that were added in the concerns. The importance of these elements was considered more consistent in the survey results as shown in Figure 16.

Figure 16: Practical barriers to data sharing and reuse



P1: There are restricted conditions on the use of the data that I handle (and thus on sharing them)
P2: I do not have enough time to prepare the data and to make them accessible and reusable
P3: The typical time scale of my research is very long (and I will not share my data before the project is finished)
P4: I feel that I do not have the right skills to prepare my data in a way that makes them accessible and reusable
P5: I think that no appropriate technical tool is available for sharing my data and making them reusable
P6: I do not have access to the necessary support or coaching that I would need to make my data accessible and reusable
P7: Other

The two least important barriers are the issue of skills and of access to appropriate tools. Most answers (especially the ones from community members) state the restricted conditions on data use

as an important obstacle. These restrictions have been detailed in the rest of the question and are mostly legal, to a lesser extent ethical and, for some researchers, commercial.

The additional obstacles stated in free text form for the *Other* box are more numerous than those related to concerns. In all, about thirty responses were completed, which can be grouped around these five categories:

- Lack of institutional support
  "There is no institutional awareness about providing infrastructure for sustainability of research results not just for simple preservation"
- Fear that there are errors in the data
- Reluctance of other people involved in the research (colleagues, supervisor, participants...)
  "Participants (e.g., journalists, communicators, strategists) would most likely not participate in our studies, if they knew the data was published"
- Lack of interest in sharing and especially lack of financial incentives
  "I don't care much about sharing my data (at least for free)"
- Lack of interest in the data
  "I am not convinced that my qualitative data may be of interest for other researchers and that they can be re-used"

All of these issues appear once or twice, only the financial issue is cited three times. Taking these obstacles into account seems unavoidable in order to improve ORD practices, especially since they affect all researchers without any significant distinction between researchers in and outside the community.

3) Motivating factors for the development of the ORD

Motivating factors echo barriers in the sense that removing them can be a motivating factor. An example of this is the financial barrier identified in the researchers' answers, which can become a motivating factor as in the following examples.

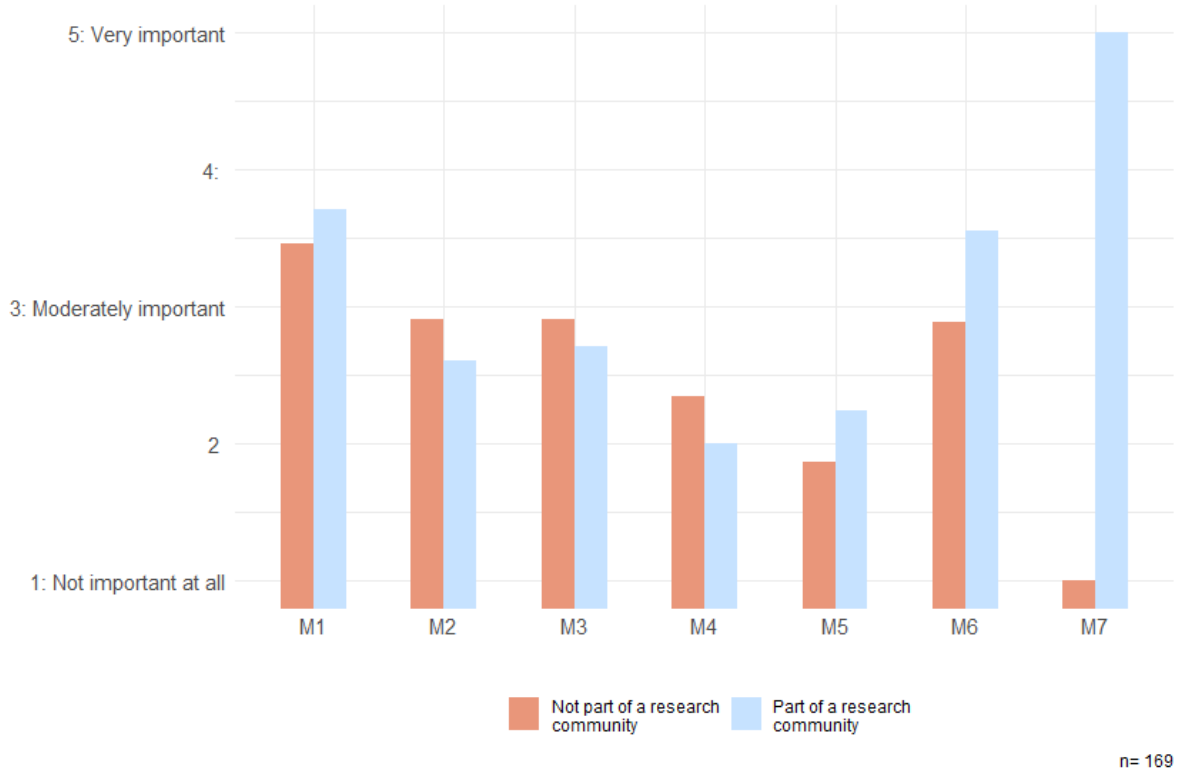
Table 5: List of motivational factors with excerpts from the interviews

Excerpts from the interviews	**Proposed answer in the survey**
"There is a change of culture, with a push from the funders and the journals (ask for replication material). current "stick" approach by funders and journals (e.g.: your next project will be funded only if you share your data from the current one)" (Lutz 2021)	It complies with the requirements from (funding agencies or) publishers.
"There is an understanding in the communities that you also have to deposit the raw data just to prove that you are doing high quality science. If you want some credibility for your field, you have to play it open" (Ackermann, Schwede 2021)	It brings me academic recognition (evaluation criterion for my academic career).
"We do not do it because FNS tells us to, but I want to do it because I get benefit from it, personally or indirectly (people see my research, or I get more	It brings me scientific recognition.

visibility, ...)." (Ackermann, Schwede 2021)	
"What is really missing is the understanding by the funding agencies that once a community comes together with an agreement for standards and archive, they need to finance this archive [to support] the effort driven by the community for the high quality curation of the data, to make them admit that this kind of service adds value to the community. »	I would like to benefit from the financial incentives existing for encouraging ORD practices.
"... an agreement in the community that you cannot publish a paper if you do not deposit your data into that archive. "(Ackermann, Schwede 2021)	It is very much practiced in my community.
"you identify people who are open-minded and willing to engage, from a good distribution of universities (including universities of applied sciences) and of disciplines." (Lutz 2021)	I commit to ORD because it corresponds to a persistent belief.

The survey results on these motivating factors are more diverse as shown in Figure 17.

Figure 17: Motivational factors



M1: It complies with the requirements from funding agencies or publishers
M2: It brings me academic recognition (evaluation criterion for my academic career)

M3: It brings me scientific recognition
M4: I would like to benefit from the financial incentives existing for encouraging ORD practices
M5: It is very much practiced in my community
M6: I commit to ORD because it corresponds to a persistent belief
M7: Other

The influence of funders on motivation (and here we find the financial issue) appears to be important. At a similar level of importance, we also note the idea of a kind of "persistent belief" which could have a link with the notion of "sharing culture". The "Other" part of the motivations is the most selected one with an evaluation at a "very important" level. In reality, only eight additional answers were given but all of them were evaluated as essential:

- Solidarity between researchers
  "To make data accessible to researchers with no/less funding" (free text response to survey question)
- Willingness to advance science and society (stated four times)
- Use of public funds (stated twice)
  "Public money is used to produce such data and they should be exploited / I would never be able to exploit the data at the 100% of their potential myself not will my research team" (response to free text survey question)
- Follow-up to UNESCO's recommendations for OS

The added motivations mostly relate to societal values and echo personal convictions that also appear to be drivers for the development of ORD practices. In Figure 17, we can see that these convictions are more important for researchers in research communities than for those outside. However, as with concerns and obstacles, this difference is not significant.

All motivating factors mentioned must be understood and activated as part of an ORD strategy. In this sense, Beate Böckem brings a more pragmatic vision that highlights the relevance of a bottom-up approach that is more effective than a top-down approach:

> *"Our demand, and one of the incentives of our researchers, is to not having to change ourselves to fit in, but to add to the field of data management. "(Böckem, Schumacher 2021)*

Concerns, barriers and motivating factors should thus be all taken into account. There is no real difference between researchers in and outside the community. Rather, there is a form of consensus on these subjects with similar degrees of importance.

B. The use of standards

The issue of standardization was very present in the interviews, with a shared opinion on the need to move towards maximum standardization.

> *"Standardization is crucial. Need of at least one better data scheme per discipline, acceptance of how things should be documented and described, in a disciplinary way. "It is where the community-building aspect comes into place. Not wise to impose standards. Create a frame, a structure where the dialogue happens."* (Lutz 2021)

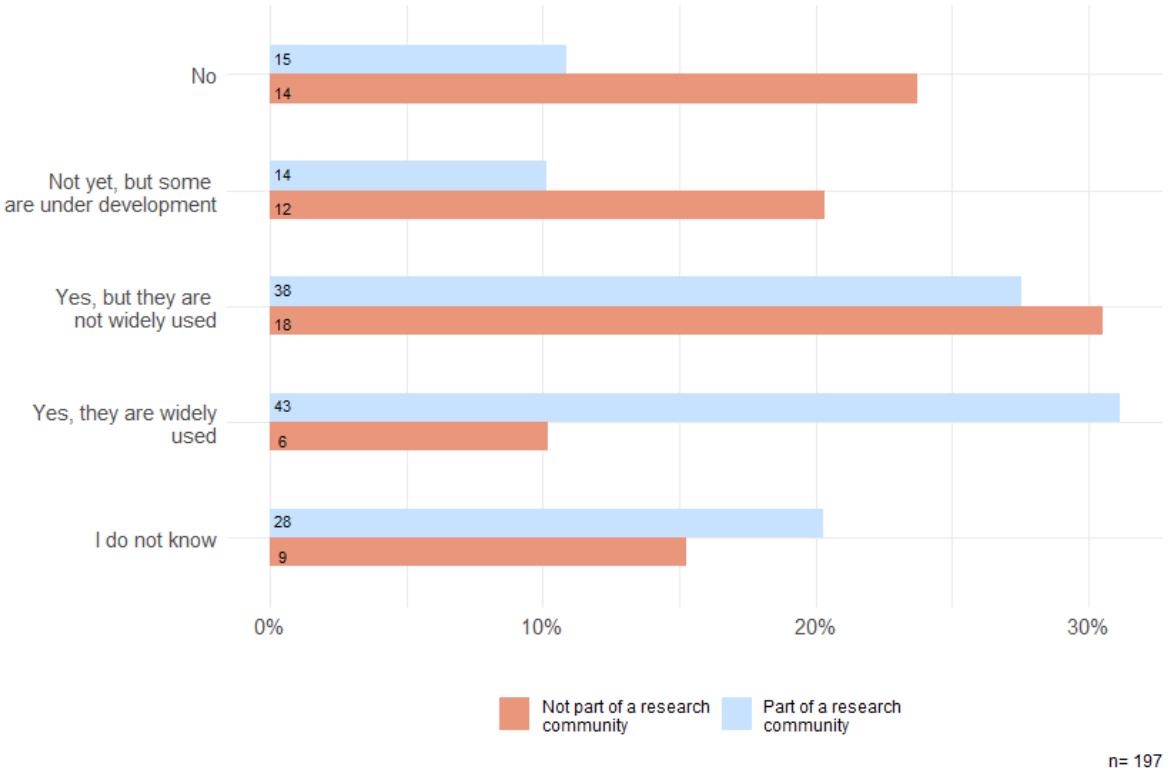
> *"Unless the data is based on common standards and understanding (on how to annotate the data, the metadata, formatting, interoperability,...), it is a waste of disk space - Community driven semantic standardization type of efforts"* (Ackermann, Schwede 2021)

> *« [C'est aux communautés] d'amener les standards. Sans contrainte forte de standardisation, il est difficile d'envisager cette réutilisation. »* (Burgi 2021)

For Pierre-Yves Burgi, if standardization is necessary, the standards must be brought by the communities to fit at most as possible with the realities of the field.

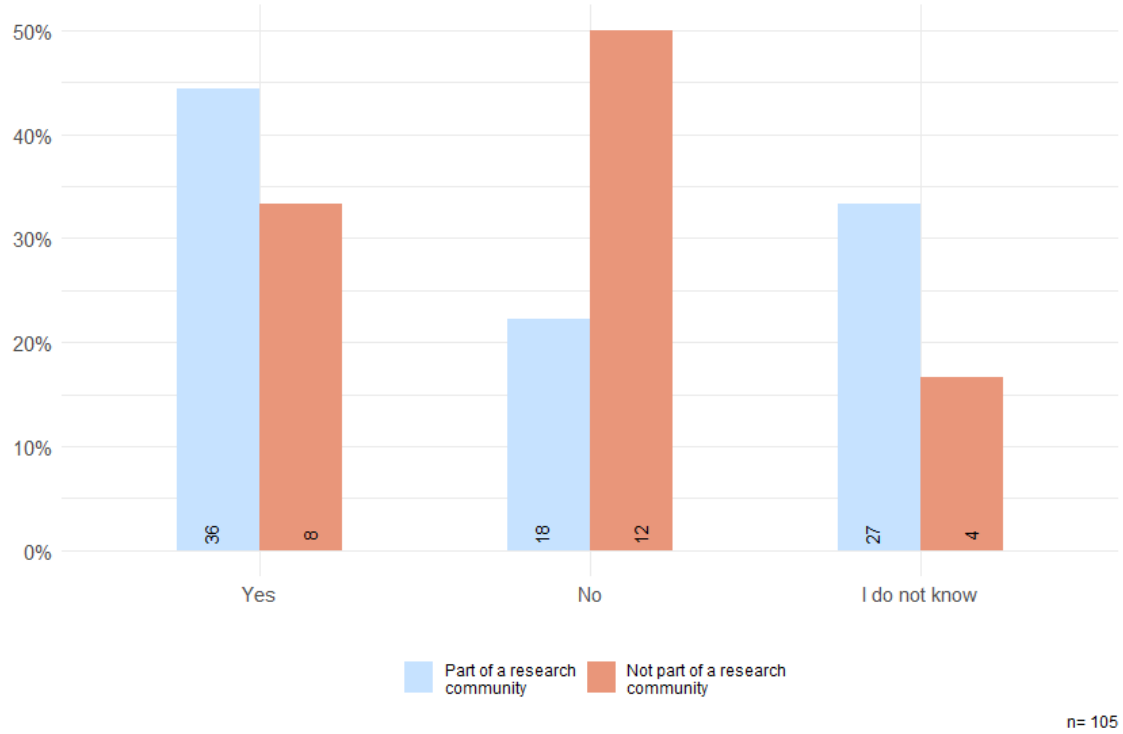
The need for standardization is therefore clearly present in the interviews. On the other hand, this need is more difficult to detect from the results of the researchers' survey.

Figure 18: The use of data standards



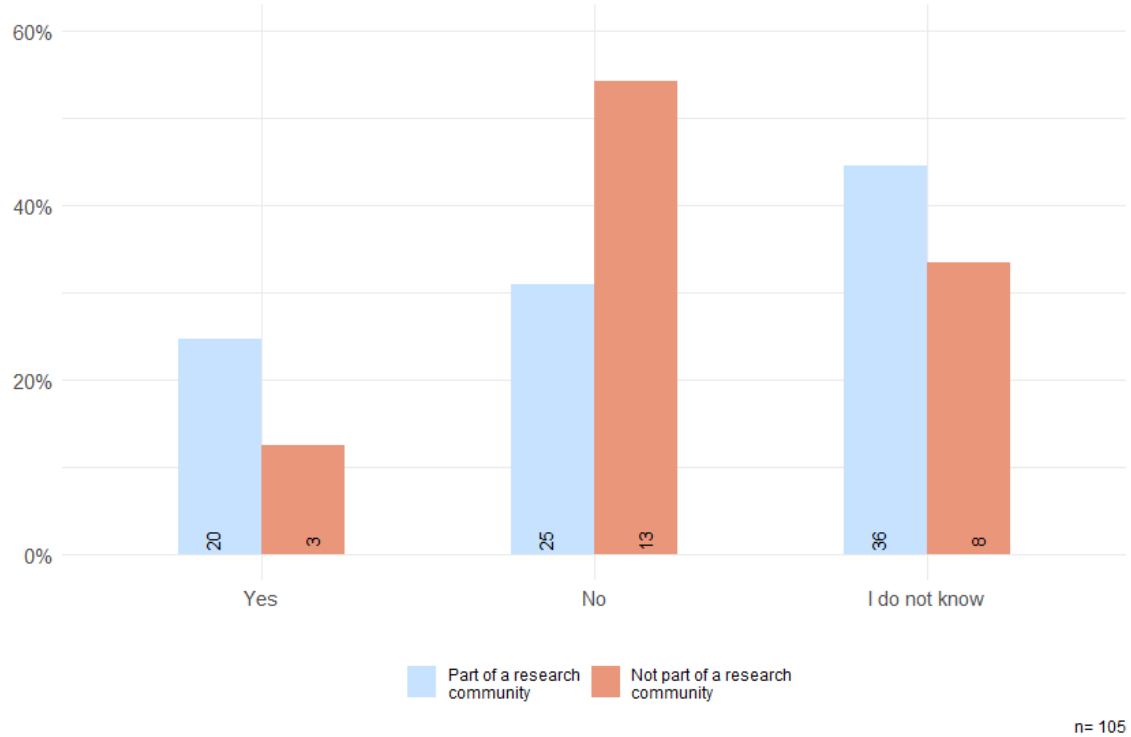
It can be seen (Figure 18) that proportionally, standards in general are used much more frequently by researchers who are members of communities than by researchers who are not. However, it is also these profiles that state in a significant proportion (around 20%) that they "do not know" whether standards exist in their community. Among non-members of the community, a significant proportion of the responses indicate that they are aware of standards that are being developed or that exist but are not yet widely used.

Figure 19: The use of metadata standards



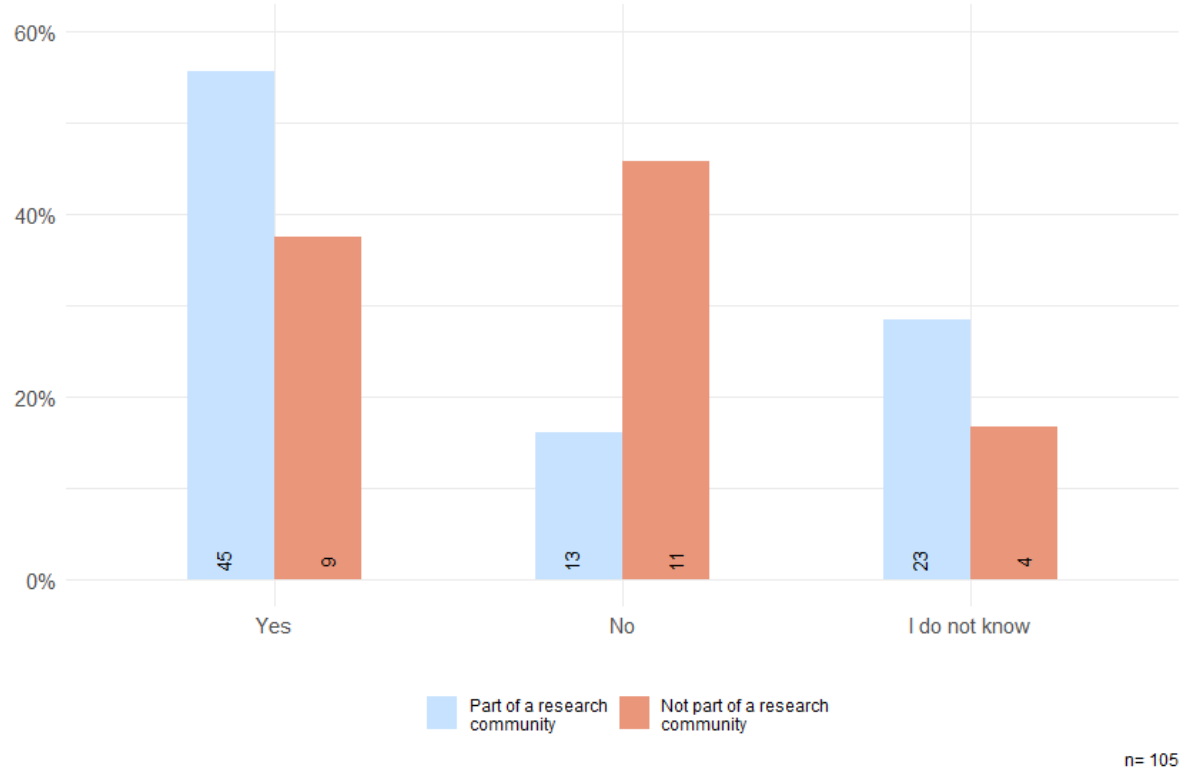
In particular, metadata standards are used proportionally more by community members, although about a third of this category also say that they do not know if those standards are used within the community (Figure 19)s. It should be noted that among the non-members, a majority answers that they do not use such standards, but also that quite a large part of them (about 30%) answers that they do use them.

Figure 20: The use of PIDs



PIDs are not as commonly used as other standards (Figure 20). Nearly half of non-community members say they do not use them, and in both categories, a significant proportion of researchers say that they do not know.

Figure 21: The use of file format standards



The use of file format standards is rather strong among community members (more than half of the answers in this category). Most of the responses from non-members are split between use and non-use, with a slight predominance of the latter (Figure 21).

Responses on the use of documentation standards are mostly negative for non-members but also (although in a lower proportion) for community members (Figure 22). This result can be explained in part by the possible ambiguity of the term "documentation standards" used in this question: standards on metadata (addressed in a previous question) or standards on the text file formats that document data (also addressed in another question).

Figure 22: The use of documentation standards

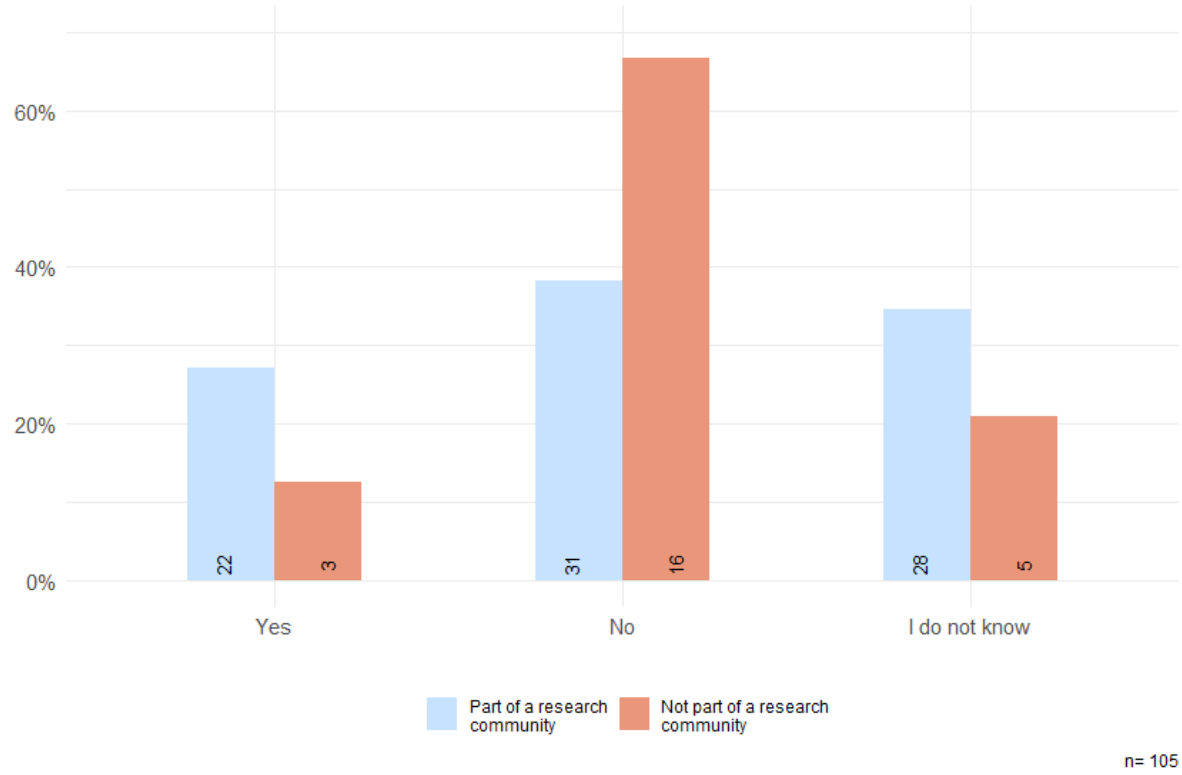
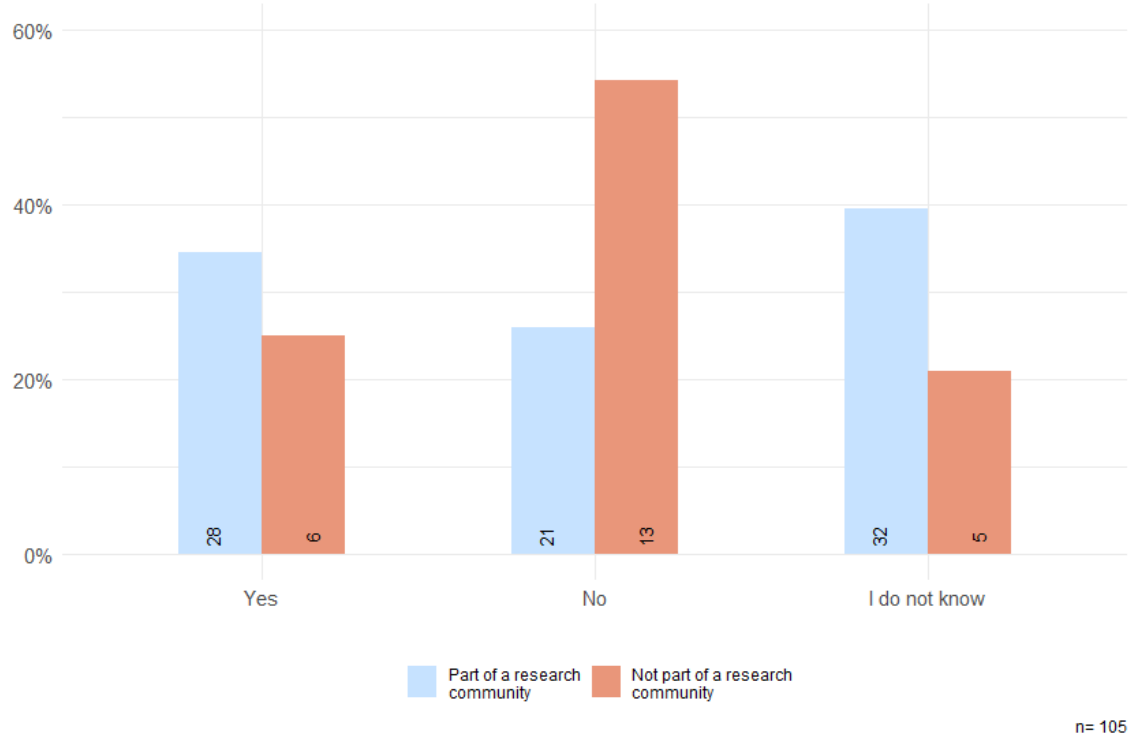
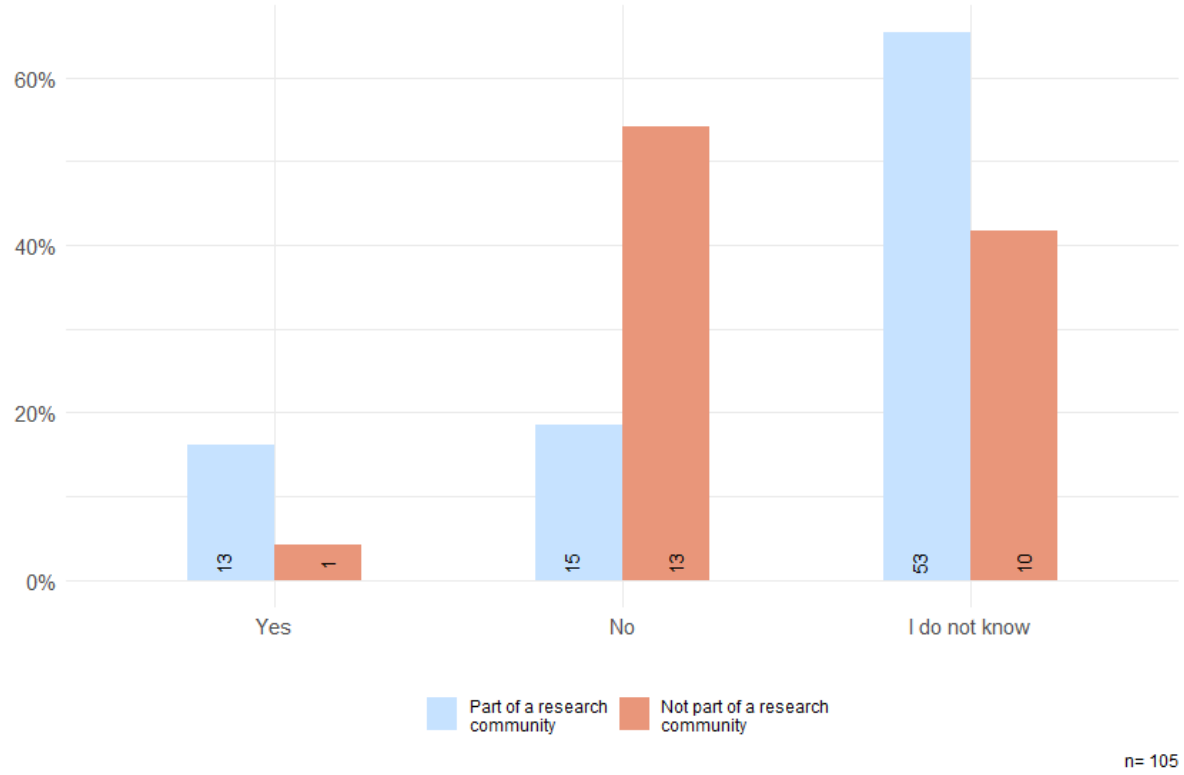


Figure 23: The use of license standards



Regarding license standards, the majority of non-members say that they do not use them (Figure 23). On the other hand, community members are most likely to say they do not know if they are used in their community. This is followed by the number of "yes" answers, then (but in a similar proportion) the number of "no" answers.

Figure 24: The use of other standards



Finally, about the possible use of other standards than those proposed above, the majority of community members state that they do not know if some are used or not (with a minority of "yes" answers), and a majority of non-members say that they do not use other standards (Figure 24).

The results in this section on the knowledge and the use of standards highlighted that community members use standards more commonly than non-community members. These results are confirmed in particular for metadata and file format standards. A significant proportion of community members (between 20% and more than 40% depending on the question) declare for several of these questions that they do not know whether standards (in general or specific standards) are used in their community.

C. The level of skills in ORD

During the interviews, the data management cycle was used as a basis for discussion of ORD practices and their level of development. For the survey, aimed directly at researchers, it was more relevant to address this aspect from practical elements and tasks that researchers could fulfill without knowing much about RDM theory and approaches.

The activities covered were as follows:

- Producing a Data Management Plan (DMP)
- Preparing your data for uploading them into a repository

- Uploading research data into a repository
- Sharing research data via a repository
- Searching research data produced by others
- Re-using research data produced by others
- Reproducing research results produced by others

For each activity, the same range of response options was provided, in order to assess the researcher's level of skill and autonomy in completing the task.

- I am able to do it alone
- I do not know if I am able to do it
- I am able to do it with help from a librarian
- I am able to do it with help from my community
- I am able to do it with help from my institution
- I am able to do it with help from the data infrastructure I use
- I am aware of the process but I do not know how to do it

For the seven tasks proposed, the majority of community members are able to perform the task independently (see Figure 25 to Figure 31). The rest of the answers are distributed in a variable way. Support from the community, the institution, or the data infrastructure used plays a role in several cases. Community support is particularly important for discovering, reusing data and replicating research results. In contrast, the results show that for all tasks, support from librarians is lower than support from other means or individuals.

The proportion of non-members of communities who perform a task independently is also significant. As with members, it is the majority for all tasks, except for producing a DMP, reproducing research results and, in a lesser extent, sharing data via a repository. In these two cases, the majority of researchers say they do not know if they are capable of performing the task.

The analysis of the results for these questions was then restricted to the responses of data community members only. It shows a behavior quite similar to that of the community members as discussed above.

Figure 25: Developing a Data Management Plan (DMP)

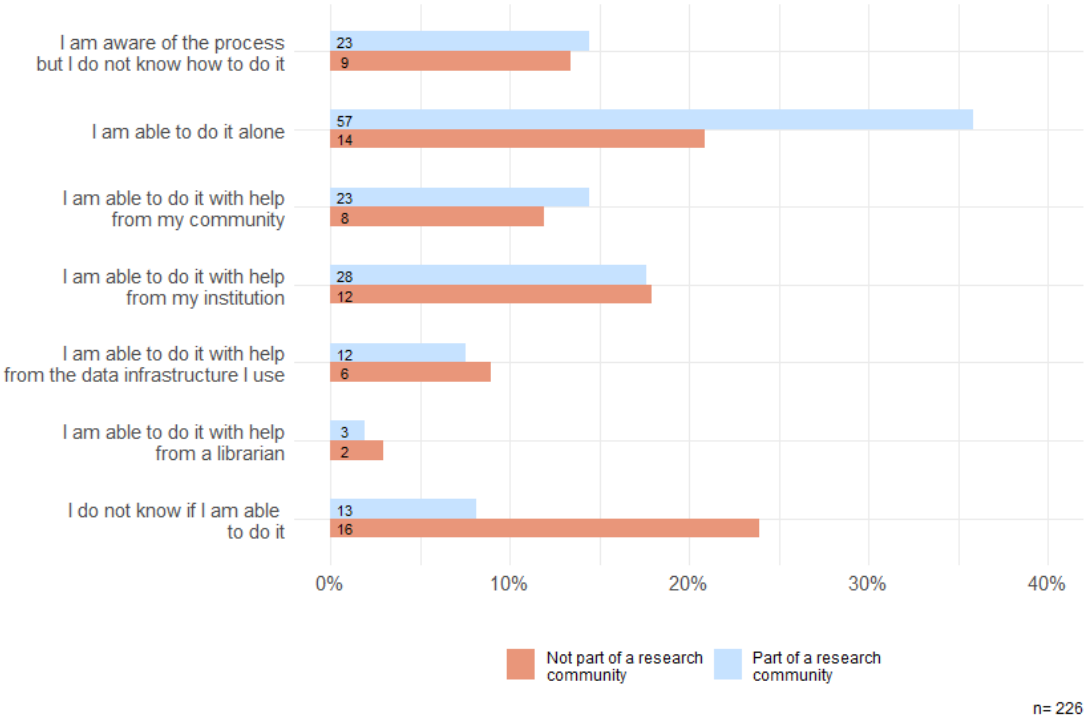


Figure 26: Preparing your data for upload to a repository

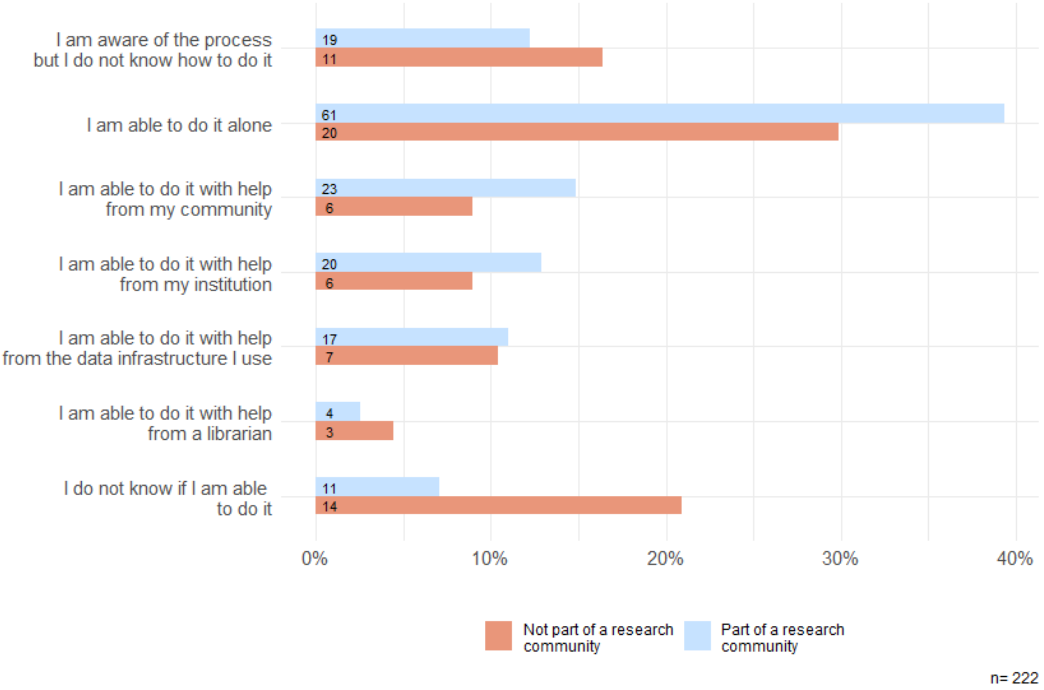


Figure 27: Downloading search data to a repository

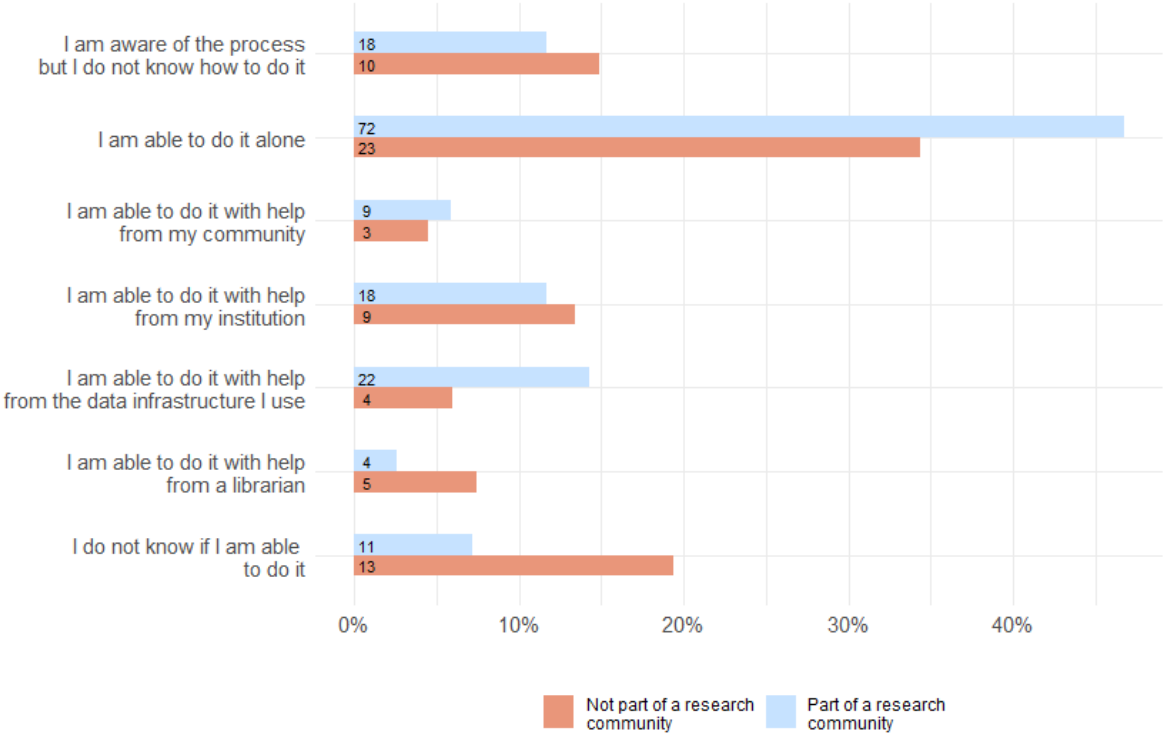


Figure 28: Sharing research data via a repository

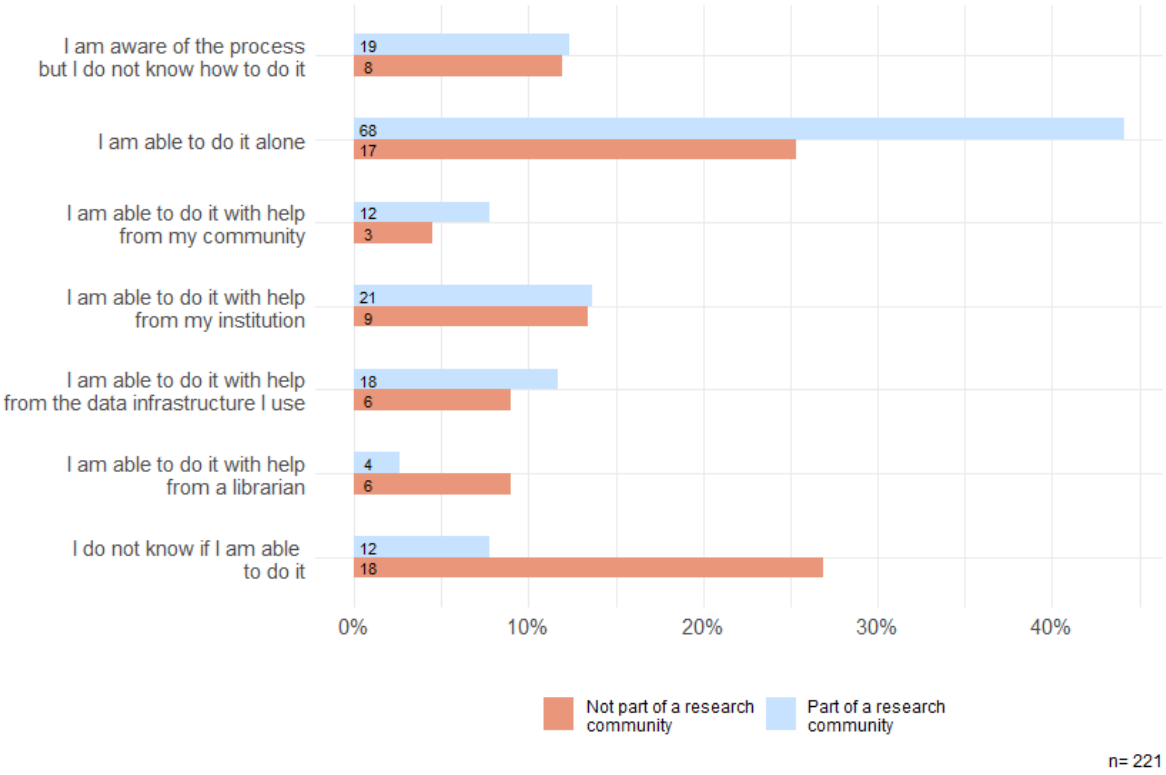


Figure 29: Searching for research data produced by others

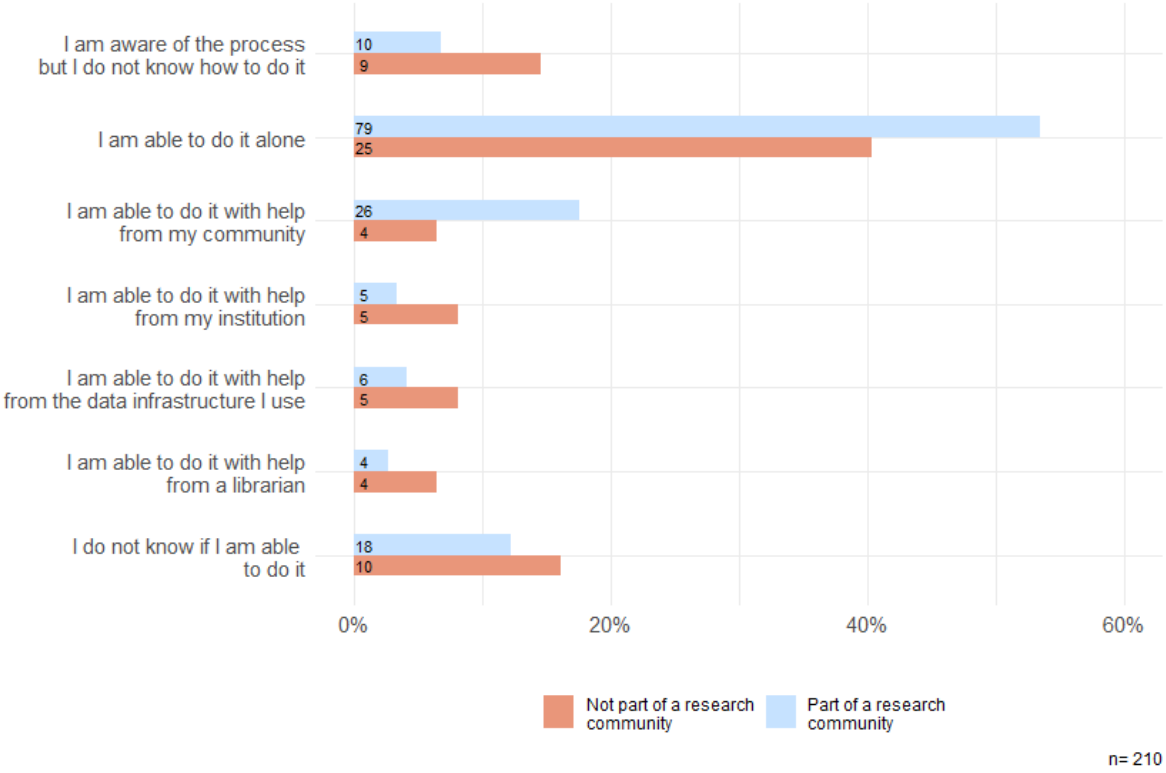


Figure 30: Reusing research data produced by others

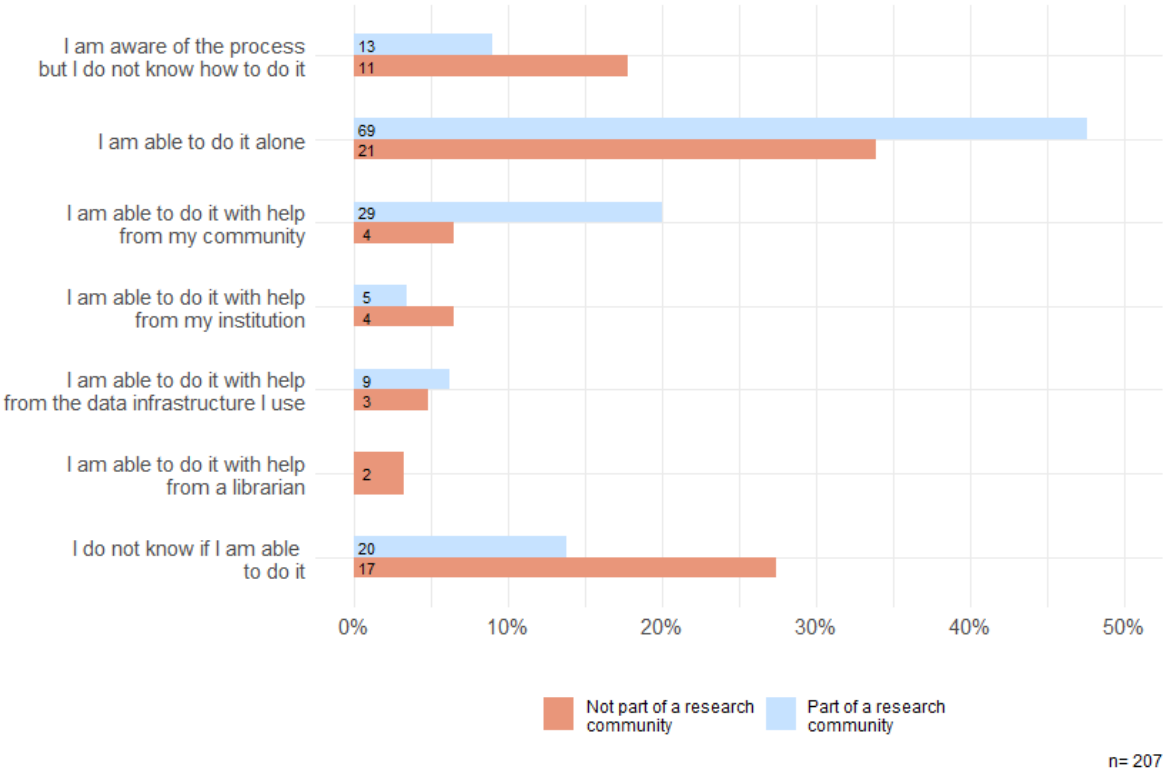
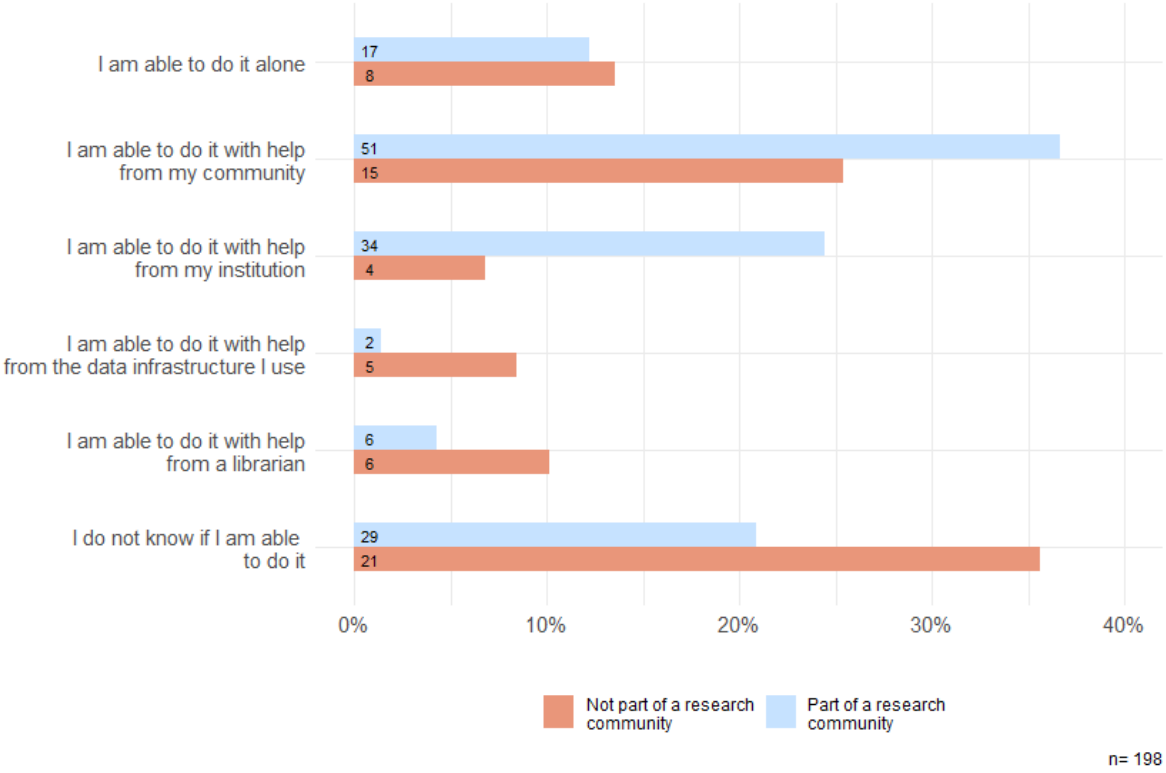


Figure 31: Replicating research results produced by others



To summarize this section, we can say that, according to these results, belonging to a community is correlated with a greater degree of autonomy in carrying out the proposed tasks. Non-members may also show some degree of autonomy, but to a lesser extent and for fewer tasks. The support provided by librarians is insignificant, in contrast to that provided by the community, the institution, or the data infrastructure.

We find that membership in a community is correlated with the use of standards and more developed ORD skills. These two elements seem to be good candidates for evaluating the level of maturity of ORD practices by a community. They are also potential levers for the development of ORD practices for each community.

III. Possible levers to contribute to the development of the ORD

A. Key players

The survey also included a free text question on the possible actors who could contribute to the improvement of the ORD practices. The responses were varied. They are presented by category in Table 6, along with some details, where appropriate, on the possible ways in which these actors could be involved.

Table 6: Key players for the ORD

Category and subcategories	Proposed action to involve them	Possible contribution
Institutional framework		
Research colleagues, senior researchers	• Raising awareness	
Researchers from qualitative disciplines (e.g. anthropologists, humanities)	• Take into account their different needs; involve them in the discussion of approaches, solutions and perspectives from their point of view	
Professors, teachers	• Raising awareness	
PIs		• Promote ORD values in the training of young researchers.
All institutional leaders (rectors, directors, etc.) and heads of departments, laboratories or research groups	• Stimulate their commitment	• To involve their employees in the dynamics.
Students	• Make them aware as early as possible	
Universities and research institutions at the leadership level		• Get more involved, provide more HR for support - with designated people • Provide better tools for data management at the research group level (data repository), e.g. radically different institutional solutions than those currently offered by libraries (and allowing dynamic queries for discoverability) • Provide support in the form of funding allocation and overtime, technical support and clarification of legal aspects. • Acknowledge and reward data sharing - by allocating dedicated funding, particularly to large, completed empirical projects. • Take into account the contribution of researchers to the ORD in their hiring,

		evaluation and career development procedures (follow DORA principles[5]).
Technology transfer offices		
IT services, data and IT experts		• Provide tools and portals that meet the highest quality standards
Librarians		
Funding agencies and research umbrella organizations		
Federal authorities (funding and long-term planning); SNSF, CUSO; government agencies, federal departments rectorates		• Funding data managers for their contribution to the ORD • Impose constraints • Provide secure and centralized data repositories
University accreditation bodies		• Forcing universities to implement ORD policies • Make funding conditional on the production of an ORD-compliant DMP • Developing policies more adapted to the humanities and social sciences
Deposits	• Involve research infrastructure staff	• Improve and simplify tools for data submission and metadata curation
Publishers (In particular those of the largest scientific journals (Elsevier...)		
Archivists		
Museums		
Information Managers		
Developers		• Develop effective sharing methods for data sharing

More detailed and contextualized information is available in the interviews, as discussed below.

Research communities

Researchers need to be involved, for example at key stages of managing their data and sharing it (on the DaSCH infrastructure): "it helps if the researchers are part of the whole process, and get to know what is happening with their data". (talking about the data curation on DaSCH, Subotic and Gautschy 2021*).*

But more generally, the research communities must be acknowledged and given their central place, and it is they who must provide the impetus for developing and improving ORD practices:

> *"About stakeholder, there is a pressure coming from existing archive or service providers who are positioning themselves as drivers in OS and standardization process. But it is difficult because OS and the driver forces come from the research communities. We have to work on the communities and the international communities. The initial force always comes from the*

[5] San Francisco Declaration on research evaluation: https://sfdora.org/read/

researchers. If we do not get the researchers on board, we will not get there" (Ackermann, Schwede 2021).

Once initiated, the rest of the process relies on key people:

> *"You identify people who are open-minded and willing to engage, from a good distribution of universities (including universities of applied sciences) and of disciplines. Try to make them reach consensus on metadata, data sharing, organizing training"* (Lutz 2021)

The involvement of researchers in the process is done through training and peer stimulation:

> *"It is much easier to get researchers enthusiastic by getting them in contact with other scientists than if told to do things because it is recommended by a librarian - the peer-to-peer effect (possibly someone from another field)."* (Ackermann, Schwede 2021)

Funding agencies

Research funding agencies were also mentioned several times. They should financially support data archive infrastructures and the necessary associated services:

> *"What is really missing is the understanding by the funding agencies that once a community comes together with an agreement for standards and archive, they need to finance this archive [to support] the effort driven by the community for the high quality curation of the data, to make them admit that this kind of service adds value to the community."* (Ackermann, Schwede 2021).

But they should also go beyond by supporting the training of researchers in their research data management skills:

> *"Funding agencies think they have solved the problem: there are now enough infrastructures and services. They are not aware of the whole process, what it really means. What is needed is to train the researchers and provide them with the skills needed to manage data."* (Gautschy, Subotic 2021).

They can also be drivers for the implementation of key principles (DORA Declaration[6]):

> *"Swissuniversities could work on implementing DORA in a better way and making ORD practices part of the evaluation criteria."* (Lutz 2021).

Consistently with the discussion above on the scope of research (if European or international, it is more favourable to the research communities), there is also an emphasis on the international cooperation of funding agencies to ensure the sustainability of data archiving infrastructures and an appropriate level of associated services:

> *"Funding and sustainability of ORD repositories]: international research/national funding agencies: all funding agencies come together to create a global fund for Open data repositories. We need this different kind of mechanism to keep the archives alive."* (Ackermann, Schwede 2021).

> *"It is an international game with international communities. And one country (a national funding agency) will not be the curator of all structures."* (Ackermann, Schwede 2021)

[6] San Francisco Declaration on research evaluation: https://sfdora.org/read/

And are mentioned more generally "all actors in the international context". (Ackermann, Schwede 2021)

Librarians

The lesser role of librarians compared to other actors in supporting researchers to perform some of the tasks characteristic of MDM and ORD has already been discussed above.

T. Schwede and S. Ackermann do see a role for them, but it is more focused on setting up and organizing awareness-raising activities and training for researchers. The content would be provided by representatives of communities that are already very advanced in ORD:

> *"Education of the research communities about the advantages of OS and sharing...The library could organize the process [of training the researchers] ... [but not be] the one providing the content. Libraries often see the archives in a traditional way. It should be complemented by examples of best practices from other communities and fields that use different approaches than a library would do... Librarians may be asked to organize the process and to pull in researchers from communities to deliver the content. Librarians have lot of experience in the humanities field but when you talk about genomics or integrative modeling, you need people of the same field."* (Ackermann, Schwede 2021)

Infrastructure

The infrastructure representatives interviewed (or associated structures, such as FORS for FORSbase) obviously mention them for the services they provide to researchers (in addition to data archiving and data access). These services take the form of standardized functionalities and mechanisms (OLOS), or services available during the whole data life cycle (creation, processing, analysis of data by FORS staff):

> *"Develop services along all the data cycle for mandates or research projects who ask for support: creating, processing, analyzing data. A significant part of the funding and resources of FORS ... Staff of about 60, among them about 30-35 work on those services ... FORS is well positioned to provide services of interest for a lot of research institutions in Switzerland ... Our role is to make sure that the technical platform and services needed to support Open Research Data are in place and easy to use for the researchers to make the data available."* (Lutz 2021)

DaSCH also contributes to the awareness and training of researchers.

> *"DaSCH tries to cover some part of it [improvement of skills and better education at the individual scale] (with information, training, education...)"* (Gautschy, Subotic 2021)

Institutions

Institutional policies can play a complementary role in supporting communities.

> *[Reuse:] "Ce sont plutôt des pratiques qui doivent être développées dans les communautés de chercheurs. [Ce] sont plutôt les politiques institutionnelles qui peuvent agir et doivent mettre l'accent sur ce sujet.* " (Burgi 2021)

Beate Böckem and Susanne Schumacher specifically refer to the research department of their institution, which directs researchers to the right person (in this case ZHdK).

> *"The research affairs department helps with all that (the first contact point is in the department of BB). There is a contact person and he/she can help with questions, establish contact to the responsible person in the archive, in the library, or in the technical"* (Böckem and Schumacher 2021).

Other actors were mentioned: developers, publishers, academies, the general public.

Developers: in one example cited by T. Schwede and S. Ackermann, they have been used to develop a tool for the automatic generation of metadata, allowing to limit the errors made manually on the metadata and strongly increasing the quality of these.

> *"Discussed with the original software developers (only 10-15 software tools are used in the field) to agree on a mechanism with which the software automatically writes the metadata, without the user manipulating it."* (Ackermann, Schwede 2021).

> *"It also allowed to involve more the software community as an ally."* (Ackermann, Schwede 2021)

Publishers: In some research fields, material useful for the reproduction of research results is sent to publishers. There is a need to establish and maintain a dialogue between repositories and publishers so that this system does not threaten the sharing and reusability of data.

> *"The FNS policy only asks for the replication material to be available, but FORS pushes for the original data and the documentation to be made available. The material which goes to publications is always a snapshot of it. FORS has set up a lighter service for the deposit of replication material: lighter metadata scheme, lighter documentation, shorter procedure, quality control. So that all replication material does not go to the publishing houses*." (Lutz 2021).

> *"Put services in place, keep the dialogue with editors, push for reproducibility but not only in the publishers' system."* (Lutz 2021).

Academies: in some research areas, they can be an important actor (coordination of large-scale projects, initiation of DaSCH, ...)

> *"Academies can play a role and do this in some extent in the medical field. The biggest project is the SPHN (Swiss Personalized Health Network) (besides SIB) that is administered by the Academy of Medical Science. Academy of Social Sciences and humanities also played a role in humanities and in building DaSCH."* (Lutz 2021).

The general public: it is the "final actor" in the process of developing a culture and practices of ORD quoted by T. Schwede and S. Ackermann. But its precise involvement has yet to be made explicit:

> *"An open data or open science repository with a certain value also becomes relevant for the general public and for education. Open science data = valuable to the society as a whole. We should not forget those stakeholders in that context."* (Ackermann, Schwede 2021).

B. Implementation strategy

The creation and development of communities is therefore acknowledged as necessary (possibly at the cost of some effort):

> *"You need to invest a lot in community buildings. You have to understand all these projects as community-building projects and that is the hard part."* (Lutz 2021)

To this end, two types of approaches were mentioned during the interviews.

A bottom-up approach

With regard to the creation and development of a community, a bottom-up approach is advocated. It is mentioned for different aspects:

Triggering the process

It is acknowledged that the impetus must necessarily come from the researchers themselves:

> *"OS and the driver forces come from the research communities. We have to work on the communities and the international communities. The initial force always comes from the researchers. If we do not get the researchers on board, we will not get there."* (Ackermann, Schwede 2021).

Operation

Governance must also take place at the community (or sub-community) level:

> *"The community-building needs to be steered and governed within each of the sub-community."* (Lutz 2021)

The definition of standards must also come from the communities or eventually be done by taking into account all the needs of the community:

> *"Standards [et c'est aux communautés de les amener]." (Burgi 2021)*

> *"Metadata schemes, standardization of formats, ... Standardization is crucial. Need of at least one better data scheme per discipline, acceptance of how things should be documented and described, in a disciplinary way. It is where the community-building aspect comes into place. Not wise to impose standards. Create a frame, a structure where the dialogue happens."* (Lutz 2021)

> *"[Data harvesting]: it was really worthwhile bringing the community together and agreeing on standards, going to the source of the software and not relying on the individual scientists." (Ackermann, Schwede 2021).*

> *[Referring to Helen Berman's paper on building the PDB community]: "To agree on common standards on data archiving and sharing. The PDB made sure to pick up all the needs of the communities to further develop the archive."* (Ackermann, Schwede 2021).

By discipline

A disciplinary approach can be mentioned for some aspects:

> *"The community-building and the service development must happen in a disciplinary way. But the technical infrastructure does need to be disciplinary focused. Swissuniversities does not have this disciplinary focus, SNSF has it much more, but swissuniversities could formulate the needs of the universities."* (Lutz 2021)

> *"Services that need to be discipline-specific (metadata schemes), but not necessarily tools and technical infrastructure."* (Lutz 2021)

> *"It is also good for the disciplines to have a gentleman agreement (or other) in the community saying that nobody publishes before the first study author (the one who generated data) has published."* (Ackermann, Schwede 2021)

A top-down approach

It is also advocated to foster the community functioning and development at the strategic level:

> *"Aux USA, Obama avait imposé que tout soit en accès libre au bout de tant d'années et maintenant les grandes infrastructures de recherche l'appliquent. Mais il ne faut pas trop compter sur l'engagement au niveau des chercheurs. Il faut toujours penser cela en top-down."* (Burgi 2021)

and for the application of the FAIR principles:

> *"The big topic is not so much as developing tools as connecting tools. We need to make the whole system of data infrastructure FAIR."* (Lutz 2021)

Several quotes mention the need to establish and maintain the connection between different actors, outside the strict framework of the research community, with possible difficulties:

> *"Put services in place, keep the dialogue with editors, push for reproducibility but not only in the publishers' system."* (Lutz 2021)

> *"It might be difficult for swissuniversities to play a coordinating role if the funding mechanism is completely disconnected from the coordination role."* (Lutz 2021)

> *"Swissuniversities could work on implementing DORA in a better way and making ORD practices part of the evaluation criteria. Mandate from the SERI to develop a strategy in the national plan, and to have coordination. But coordination aspects are not clear. SNSF put up a reflection group on data infrastructure, but without swissuniversities or representatives of academies. (Lutz 2021).*

Two quotes highlight limitations to this top-down approach:

> *"Constant lack of coordination in Switzerland overall, very little top-down coordination, a big issue."* (Lutz 2021)

> *"Top-down model is quite expensive and hardly extendable at a large scale."* (Clivaz 2021)

C. Awareness, education and training

After the identification of key actors and of the most relevant strategy for the establishment and development of research communities, a third aspect for facilitating and developing the culture and practices of the ORD emerged during the interviews:

> *"Educate people - education - education - education"* (Gautschy, Subotic 2021)

This "education" encompasses both awareness of the benefits of the ORD:

> *"Education also includes making the researchers fully aware that sharing and re-using data (the possibilities of the semantic web, ...) is in his/her own interest."* (Gautschy and Subotic)

> *"Educate more the research communities about the advantages of OS and sharing and what it means and how it could work. Librarians may be asked to organize the process and to pull in researchers from communities to deliver the content (peer-to-peer effect)."* (Ackermann, Schwede 2021)

> *"Training workshops will allow large discussion in Humanities and convince scholars to produce data with a good balance between flexibility and standardization"* (Clivaz 2021)

and training for researchers to improve their skills for good ORD practices:

> *"People need more skills and a better education at the individual scale. The education part is the most important. You can always offer more, but without education, providing more services would be useless."* (Gautschy, Subotic 2021)

> *"What is needed is train the researchers and provide them with the skills needed to manage data."* (Gautschy, Subotic 2021)

> *"Cela passe par la formation au niveau suisse sur la GDR"* (Burgi 2021)

> *"SNSF/ORD joint workshop would be wishable"* (Clivaz 2021)

Details are given on the timing and target audience of these awareness-raising and training activities.

It is important to get involved as early as possible in the research process, before the data collection stage:

> *"You can optimize by teaching the researchers before they start collecting data"* (Lutz 2021)

> *"Training of researchers: Try to capture the people as early as possible [as soon as the project is funded]. It should be a joint effort between research funders and universities to put support in place and help the researchers"* (Lutz 2021)

but also to train future researchers while they are still students:

> *"Training of students"* (Lutz 2021)

> *"Universities can educate students because they need it in their future"* (Gautschy, Subotic 2021)

> *"Whatever the disciplines, at the Master level (or even at the Bachelor level), there should be mandatory basic training. The students there are the future application writers."* (Gautschy, Subotic 2021)

> *"About older researchers: some of them are interested; for other ones, it is hopeless. They tend to delegate the management tasks to their students."* (Gautschy, Subotic 2021)

IV. Recommendations for developing the level of ORD practices within communities

The recommendations presented in this section are based on the findings of the interviews and on the results of the survey. In order to bring the strategy as close as possible to the practices, they follow a bottom-up and top-down logic as advised.

1) Talking the same language and starting from the beginning

// Define the concept of research community

The term "research community" is the central concept of this work, as requested by the mandator. It is also mentioned in the literature review. The results of this work confirm the potential of these communities in the development of the researchers ORD skills. However, our results also highlight that there is still little understanding and consensus on this term among OS/ORD experts and among researchers. The interview and survey data show that the term "research community" is often understood as the community of the research discipline itself, even though we proposed other definitions during the data collection. The term "data community" discussed in the literature is even less used by researchers, as the results of the massive survey underline.

Therefore, before seeking to strengthen the culture of ORD and improve the ORD practices of researchers, efforts should first be made to understand what a research or data community is and how it helps researchers improve their ORD practices; and from there, build and strengthen these communities.

Recommendations :

- **1a** Agree on a definition of research communities that includes their variety but specifies the importance of formal organization (e.g., by projects, in the form of associations, around infrastructures, etc.)
- **1b** Rely on formal organizational models that identify points of contact (people in charge, institutions, etc.)
- **1c** Identify the communities and the contact points for connecting to them

2) Supporting the construction and consolidation of research communities

Initiating the formation of a community, consolidating it and supporting it in acquiring maturity in ORD is a two-way process. Our results highlight that there is no point in forcing it if there is no strong motivation from the researchers themselves. However, it requires a stable governance framework, connected to the wider environment, adapted to different contexts, and long-term planning.

Recommendations :

- **2a** Acknowledge the need for a bottom up and top down approach
- **2b** Put in place tools that allow researchers to bring forward their needs or questions and involve them in the process

3) Develop incentives (reduce barriers)

- **3a** Acknowledge the effort, time and skills that researchers invest in ORD good practices: take these aspects into account for institutional hiring and academic career advancement

- **3b** Promote the use of informal community agreements for data sharing ethics: no publication of an article before the first author of the study or the first data producer has published

4) Involve the different actors and formalize their contribution

Several actors play different roles in this context and should be involved at specific stages of the process. Section III presents these potential actors and their possible involvement based on the results of the interviews and the survey. What also emerges from this mandate is the strong interest of many people in this topic and a willingness to share their knowledge.

Recommendations

- **4a** Identify people who are particularly active on the subject (specialists or community members who are advanced in their practices and in ORD culture) and who are likely to attract and bring together other members
- **4b** Formalize a network of these people (adapt the Data Champions model into a "Community Champions" model)

5) Use timely communication and training

Communication is essential to raise awareness of the benefits of ORD practice. As such, it should be a priority when trying to support the formation of a new community.

Training is a fundamental part of the process. The objectives are not the same at each step of the process, but it must be part of the process throughout, and beyond the consolidation of the community, if it is to exist in the long term.

However, the lack of time is a real problem that appears in the researchers answers. Offering full days of training is contradictory when lack of time is already an obstacle in itself.

Recommendations

- **5a** Offer very short and easily accessible training formats (micro-learning)
- **5b** Train the actors of the ORD in this training methodology
- **5c** Involve representatives of the Community Champions

6) Provide the right level of support for the definition and adoption of standards

Standardization is an essential aspect of ORD maturity. The existence and the use of standards by communities is an indicator of the maturity of their ORD practices. The adoption of standards can also accelerate the motivation of researchers to follow good ORD practices if they see the benefits. However, there is no point in imposing these standards. They must be designed and implemented by the community members themselves, otherwise they are unlikely to be fully adopted by the community. Support must be provided in another way: by identifying community members who can bring others together around the definition of standards and provide them with adequate support throughout the process.

Recommendations :

- **6a** Encourage projects that aim at standardization with a bottom-up approach
- **6b** Communicate and promote the use of standards used and/or developed

7) Planning for sustainable funding

The lack of time has been widely discussed during this work and ultimately needs to be addressed, especially from the financial point of view. This funding must be sustainable enough to allow the infrastructure, the services and the awareness-raising activities (communication, training and support) to be available throughout the process of developing the community's maturity in ORD, and beyond, to sustain the community and its good practices. In supporting the early stages of a new community, funding should be provided on a case-by-case basis to meet specific needs as they arise. In the next stages, it should be provided on a regular basis to ensure the long-term sustainability of the infrastructure and services available to communities.

Recommendations :

- **7a** Evaluate the financial and human resources required for each stage of the research process (percentage to be established in relation to the typology of the research)
- **7b** Take this assessment into account when estimating the funding needed for research projects

Decision matrix

We assigned an ID number to each recommendation (16 in total) and grouped them in Table 7.

Table 7: ID number of recommendations used in the decision matrix

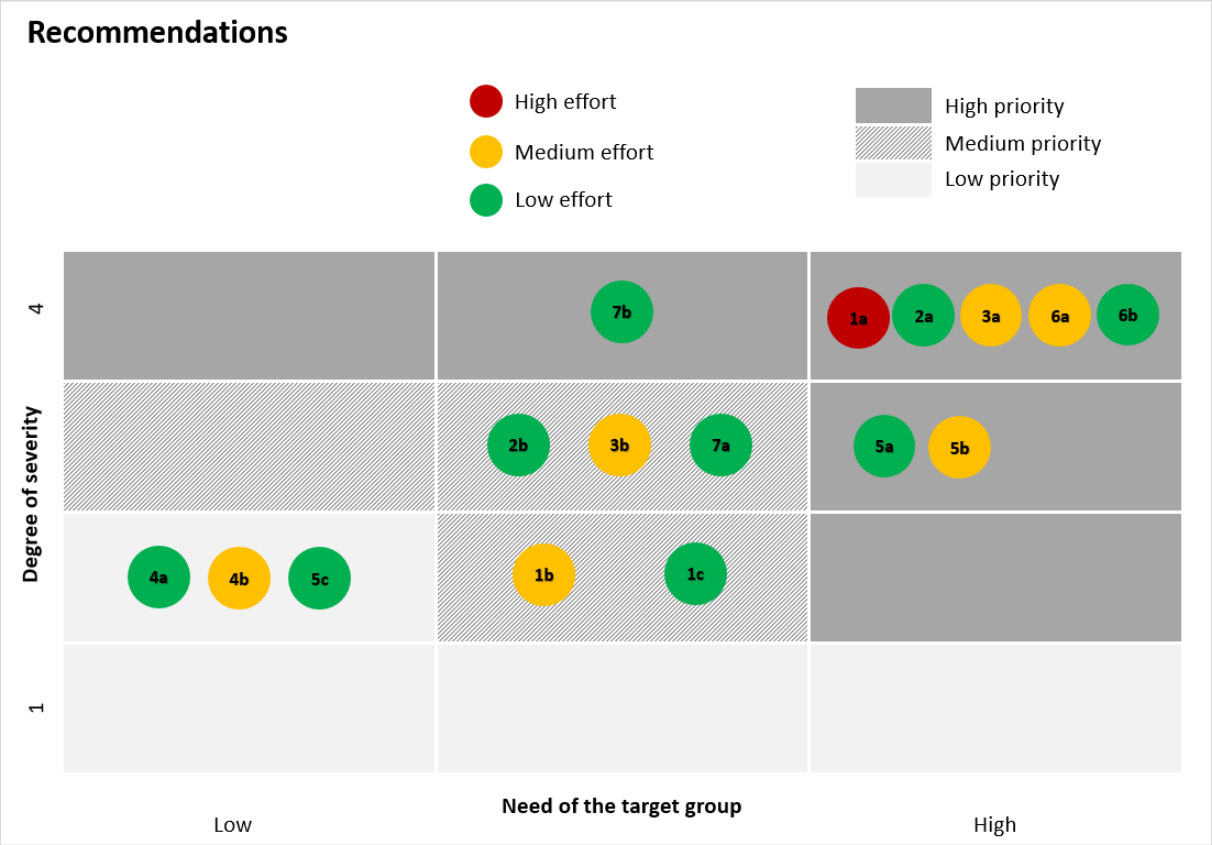
Identifier	Recommendations
1a	Agree on a definition of research communities that includes their variety but specifies the importance of formal organization (e.g., by projects, in the form of associations, around infrastructures, etc.)
1b	Rely on formal organisational models that allow for the identification of contact points (persons in charge, institutions, etc.)
1c	List the communities and the points of contact that allow you to exchange with them
2a	Acknowledge the need for a complementary approach (bottom up and top down)
2b	Put in place tools that allow researchers to bring forward their needs or questions and to involve them in the rest of the process
3a	Acknowledge the effort, time and skills that researchers invest in good practice ORD: take these aspects into account for institutional commitment and academic career development
3b	Promote the use of informal community agreements for data sharing ethics: no publication of an article before the first author of the study or the first producer of the data has published
4a	Identify people who are particularly active on the subject (specialists or community members who are advanced in their practice and in the ORD culture) who are likely to bring together other members
4b	Formalize a network between them (transpose the Data Champions model to "Community Champions")
5a	Offer very short training formats (micro-learning) that are easily accessible
5b	Train the actors of the ORD in this training methodology
5c	Involve representatives of the Community Champions
6a	Encourage projects that aim at standardization with a bottom-up approach
6b	Communicate and promote the use of standards used and/or developed

7a	Evaluate the financial and human resources required for each stage of the research process (percentage to be established in relation to the typology of the research)
7b	Take this assessment into account when estimating the necessary funding for research projects

The decision matrix shows the priority level of each recommendation, according to its degree of severity and the need of the target group (Figure 32). The level of effort required to implement each of them is also assessed:

- Priority: low, medium and high
- Recommendation severity: the level of importance of the recommendation (which therefore influences the impact of the recommendation). 1 being the lowest and 4 the highest degree of severity.
- Need for the target group: low, medium and high
- Effort: low, medium and high

Figure 32: Decision matrix



Conclusion

Swissuniversities, as part of its National Strategy for Open Research Data initiated by its Open Science Delegation, has commissioned the HEG and the FHGR to carry out two complementary studies. The first one aims at characterizing the open research data (ORD) landscape in Switzerland and the second one, carried out by the HEG, aims at characterizing the practices and needs of researchers in ORD by focusing on research communities.

To carry out this work, two successive phases were performed. First, semi-structured interviews were conducted with representatives of data sharing infrastructures and research communities. Second, Swiss academic researchers were asked to answer an online questionnaire. These two steps aimed at collecting information on the current culture and practices in ORD, as well as the needs for improvement of these practices, and the actors to be involved.

The report is divided into four main parts, each with its own sections.

The first part focuses more globally on "research communities" and discusses the different definitions of these communities, their organization as well as two particular profiles identified, the Data Communities members and the "data-less" researchers.

The second part focuses on the role of organization of research by "research communities" for ORD practices and discusses the obstacles as well as the motivating factors for ORD development. It sheds light on the use of standards as well as on the level of ORD skills. In the perspective of a future development of a maturity level model, these two aspects would be essential elements to take into account.

The third part highlights the possible levers to contribute to the development and advancement of ODR practices. These include the creation of research communities based on simultaneous bottom up and top down approaches, awareness raising, education and training, as well as key players. It was found that the community, the institution, and the data infrastructure provide significant support to researchers for good ORD practices. Librarians could be more involved in awareness raising and in the organization of dedicated training.

Finally, the fourth part proposes recommendations for developing the level of ORD practices within communities. They are divided into seven topics, each of which highlights specific recommendations. A decision matrix is used to plan their strategic implementation.

The results of this study confirm the hypothesis that making part of a research community has a positive impact on the development level of researchers ORD practices. Those results also highlight relevant concepts and elements for the possible development of a detailed model for evaluating the ORD maturity level of research communities. This model would be a useful tool for the monitoring and the strengthening of ORD and OS development in Swiss academic research.

Bibliography

BERMAN, Helen M., KLEYWEGT, Gerard J, NAKAMURA, Haruki and MERKLEY John L, 2013. How Community Has Shaped the Protein Data Bank. *Structure* [online]. 3 September 2013. 21(9), pp. 1485-1491. [Accessed 11.06.2021]. Available at: https://doi.org/10.1016/j.str.2013.07.010[access by subscription].

BÖCKEM, Beate, SCHUMACHER, Susanne, 2021. Interview with Ms. Böckem, Head of research affairs at ZHdK, and Ms. Schumacher, Co-chair of the digital council of ZHdK, videoconference, 2 July 2021.

BURGY, Jean-Yves, 2021. Interview with Pierre-Yves Burgy, Deputy Director of the UNIGE Information Systems Division, initiator of the OLOS project and co-director of the DLCM, videoconference, 21 May 2021.

COOPER, Danielle, SPRINGER, Rebecca, 2019. Data Communities : Data Communities: A New Model for Supporting STEM Data Sharing. Ithaka S+R [online]. 13 May 2019. Available at https://doi.org/10.18665/sr.311396

COOPER, Danielle, SPRINGER, Rebecca, 2020. Data Communities: Empowering Researcher-Driven Data Sharing in the Sciences. *International Journal of Digital Curation* [online]. 30 December 2020. Vol. 15, No. 1. [Accessed 11.06.2021]. Available from: https://doi.org/10.2218/ijdc.v15i1.695

LUTZ, Georg, 2021. Interview with Georg Lutz, Director of FORS, videoconference, 17 June 2021.

Research Data Alliance FAIR Data Maturity Model Working Group, 2020. FAIR Data Maturity Model: Specification and Guidelines. June 1, 2020. Available at: https://doi.org/10.15497/rda00050

SCHWEDE, Torsten, 2021. Interview with Dr. Torsten Schwede, Vice President for Research, Head of the Computational Structural Biology Group, University of Basel, Swiss Institute of Bioinformatics, videoconference, May 28, 2021.

Swissuniversities, 2021. Swiss National Open Research Data Strategy. July 2021. Available at: https://www.swissuniversities.ch/fr/themes/digitalisation/open-research-data

V. Appendixes

Appendix 1: Key Concepts

The concepts used in this work are defined as follows:

Data communities: according to Danielle Cooper and Rebecca Springer: "[Formal] or informal groups of scholars who share a certain type of data with each other, regardless of disciplinary boundaries" (Cooper, Springer, 2019).

Research communities: networks of researchers sharing the same (type of) data, the same data sharing tools and infrastructures, the same RDO practices and the same Open Science culture, but not necessarily belonging to the same research discipline.

Research data: according to the Swiss national open research data strategy: "Research data are factual elements collected, observed or generated that are commonly accepted in the scientific community as necessary to enable, document or validate research results." (swissuniversities 2020).

Survey (online): A method of collecting information in writing. It allows you to target a large number of recipients and receive their responses within a limited time frame.

- (Semi-structured) interview: a method of collecting information verbally. In semi-structured interviews, the list of topics to be discussed and certain questions to be asked are fixed. The other questions and the order in which they are asked remain flexible, allowing for adaptation to the interviewee (as long as the discussion follows a natural order).

Infrastructure: Data sharing infrastructure, i.e. data storage and sharing solutions. Beyond the minimum, they very often provide specific features and tools to assist the repository process, make the data FAIR and archive it. By default, the term "infrastructures" will also include these features and tools.

Support services: services provided to researchers to adopt and improve their research data management practices. These services may include information, training, coaching, etc.

Appendix 2: Template of the semi-structured interview guides

Scope of the interview

Introduction and short reminder of the objectives - 5 mn

I. Infrastructures and support services for ORD - open questions - 15 mn
II. Maturity level of ORD practices - card sorting - 15 mn
III. Identified needs for ORD - open questions - 15 mn

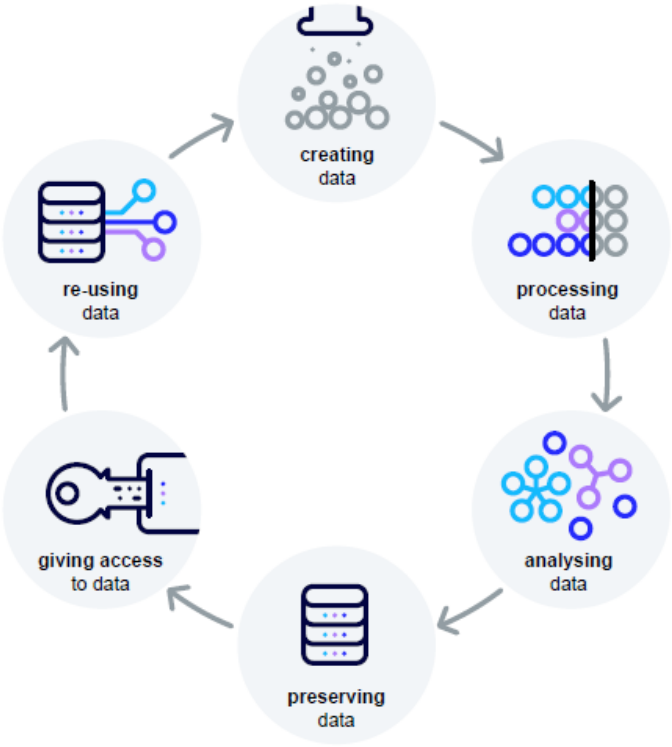
Conclusion (validation procedure of the interview report and reminder about the survey to come) - 5 mn

Detailed content

Infrastructures and support services for ORD - open questions

1. What kind of infrastructure (data sharing repository) and support services does your research group use for managing and sharing its data ? Why ?

2. In the Data Lifecycle Management Cycle illustrated as below, which steps is more relevant to your research group, which steps does it practice more?



3. About "re-using data" (according to the answer given at the previous question) :

Option 1 - if the "re-using" step is stated (besides giving access): in which way do you think your research group facilitates research reproducibility (besides sharing data)?

Option 2 - if the "re-using" step is not stated: are there plans to contribute more to research reproducibility, and if yes, in which way?

4. Within research in bioinformatics or at your institution, do you identify specific research data communities ? If yes, which ones ? Are there researchers in those communities that we could contact on your behalf to investigate more about their practices and their needs (in terms of ORD)?

> *Note: By research communities, we mean networks of researchers sharing the same (type of) data, the same data sharing tools and infrastructures, the same ORD practices, and the same culture of Open Science, but not necessarily being from the same research discipline. We are interested in getting your views on this definition, if you agree with it or if you have suggestions for another one.*

Maturity level of ORD practices - card sorting

Guidelines :

- Sort out the communities according to how you evaluate their ORD level: low, intermediate or high.
- On which criteria do you base for sorting the cards?

Identified needs for ORD - open questions

1. About research data, can you think of other types of data than the following ones:

- comments
- experiences
- simulations
- derived data
- references
- digitalizations (see details in Appendix 1) ?

2. If yes, which types ? In which disciplines or research communities ? Does the management of these data follow different steps from the ones: "preserve, give access, re-use" ?

> *Note : we are trying here to identify research communities with less common views or behaviour in terms of data management. Better understanding their practices may allow to address their needs with more adapted support (infrastructures, services, ...).*

3. In a general way of speaking, according to you, what are the main obstacles for ORD (and more specifically for data re-using) ? Are they different from one research community to the other ?

4. What are the main stimulating factors for ORD (and data re-using)? Are they different from one research community to the other ?

5. In view of the current offer of infrastructures, services, incentives available to the researchers, what are their needs to more actively engage in ORD and to improve their practices (and more specifically relating to data re-using)? Do those needs differ by research data community?

6. Do you have other suggestions to overcome the current shortcomings or caveats and to further contribute to better practices and more commitment into ORD?

7. In a general way, which shareholders could be more involved?

Bibliography

RITZE, Dominique, ECKERT, Kai and PFEFFER, Magnus, 2013. *Forschungsdaten*. In: DANOWSKI, Patrick, (Open) Linked Data in Bibliotheken. Berlin: De Gruyter Saur, 2013, pp. 122-138. ISBN 978-3-11-027634-3

Typology of Research Data

Art	Characteristik	Beispeil
Beobachtungen	Daten werden in Echtzeit erfasst meistens unersetzbar	Sensordaten Umfragedaten
Experimente	meist im Labor erstellt reproduzierbar aber teuer	Gensequenzen Chromatogramme
Simulationen	von Testmodellen generiert Model und die Metadaten wichtiger als Ausgabe	Klimamodelle Wirtschaftsmodelle
Abgeleitete Daten	aus anderen Daten abgeleitet oder kompiliert reproduzierbar	Textmining 3D-Modelle
Referenzen	Sammlung kleinerer Datensätze Meist publiziert	Gensequenzdatenbank Primäre Textquellen
Digitalisierungen	Digitale Versionen von analogen Objekten, reproduzierbar, solange die Originale verfügbar sind	Manuskripten

(Adapted from Ritze et al. 2013)

Appendix 3: Topics and sub-topics of the analysis grid of the data collected with the interviews

Topic	**Sub-topic**
Disciplines and communities	
	Individual/Communities
	Difference between disciplines
	Difference between disciplines - reasons
Organization of communities	
	Culture of reusing - quote
	Culture of reusing - topic
Typology of research and consequences on ORD	
	Types of data
	Methodology
	Research time scale
	Research level (national, international, ...)
	Publication
	ORD - bottom up, building communities
	ORD - top down
Concerns - obstacles - incentives	
	Concerns - quote
	Concerns - topic
	Obstacles - quote
	Obstacles - topic
	Incentives - quote
	Incentives - topic
Solutions for ORD development	
	Standardization - quotes
	Standardization - topics
	Education
	Skills needed
About infrastructure	
	Motivation and goal
	Cycle
	Functionality
	Standard
	Statistic
Other topics	
	Licenses
	Metadata
	"False" obstacles
	Actors
	Public open repository - commercial structures

Appendix 4: Detailed content of the online survey

OpenScience survey

Introduction

This survey is performed within the framework of a swissuniversities mandate and in the context of the National Open Research Data Strategy initiated by the Open Science Delegation of swissuniversities. The mandate specifically aims at characterizing the practices and needs in ORD within Swiss research communities and **the survey has been designed for researchers** based in Switzerland.

If you are not a researcher yourself but in a position where you have access to a network of researchers in Switzerland (e.g. OA/ORD/OS officer or librarian at Swiss higher education institutions), please forward it via this network.

This survey will stay open until 29 August 2021. **It should take about 15-20 minutes to complete**. The data collected will be used only for this mandate
We thank you in advance for your contribution to our work.

HEG mandate team
Dir : Prof. René Schneider
Haute Ecole de Gestion de Genève
Filière Information documentaire
Rue de la Tambourine 17 - 1227 Carouge
gaia.bongi(at)hesge.ch
marielle.guirlet(at)hesge.ch
elise.pelletier(at)hesge.ch

End of block: Introduction

Beginning of block: Question for all

A1 What is your institution of affiliation?

- Federal institute (1)
- University (2)
- University of Applied Sciences (3)
- University of Teacher Education (4)
- Higher Education Institution (5)
- Accredited institution (6)

Post this question:

If A1 = Federal institute

A2 - 1 Please specify your answer given to the last question:

- Ecole polytechnique fédérale de Lausanne EPFL (1)
- Eidgenössische Technische Hochschule Zürich ETHZ (2)
- Swiss Federal Institute for Vocational Education and Training, SFIVET (3)
- Swiss Federal Institute of Sport Magglingen, SFISM (4)
- Paul Scherrer Institute, PSI (5)
- Swiss Federal Institute of Aquatic Science and Technology, Eawag (6)
- Swiss Federal Laboratories for Materials Science & Technology, Empa (7)
- Swiss Federal Institute for Forest, Snow and Landscape Research, WSL (8)

Post this question:

If A1 = University

A2 - 2 Please specify your answer given to the last question:

- Universität Basel (1)
- University of Bern (2)
- University of Freiburg (3)
- University of Geneva (4)
- University of Lausanne (5)
- University of Lucerne (6)
- University of Neuchâtel (7)
- University of St. Gallen (8)
- Università della Svizzera italiana (9)
- University of Zurich (10)

Post this question:

If A1 = University of Applied Sciences

A2 - 3 Please specify your answer given to the last question:

- BFH - Bern University of Applied Sciences (1)
- FHGR - University of Applied Sciences of the Grisons (8)
- FHNW - University of Applied Sciences and Arts Northwestern Switzerland (2)
- HES-SO - University of Applied Sciences and Arts Western Switzerland (3)
- HSLU - Lucerne University of Applied Sciences and Arts (5)
- Kalaidos FH - University of Applied Sciences (4)
- OST - Eastern Switzerland University of Applied Sciences (6)
- SUPSI - University of Applied Sciences and Arts of Southern Switzerland (7)
- ZFH - Zurich University of Applied Sciences and Arts (9)
- ZHAW - Zurich University of Applied Sciences (11)
- ZHdK - Zurich University of the Arts (12)
- PHZH - Zurich University of Teacher Education (13)
- HWZ - Zurich University of Applied Sciences in Business Administration (14)

Post this question:

If A1 = University of Teacher Education

A2 - 4 Please specify your answer given to the last question:

- University of Teacher Education of the Cantons of Bern, Jura and Neuchâtel (1)
- Pedagogical University of the Canton of Vaud (2)
- Pädagogische Hochschule Wallis | Haute école pédagogique du Valais (3)
- Pädagogische Hochschule Freiburg (4)
- Interkantonal Hochschule für Heilpädagogik Zürich (5)
- Pädagogische Hochschule Graubünden | Alta scuola pedagogica dei Grigioni | Scola auta da pedagogia dal Grischun (6)
- Pädagogische Hochschule Bern (7)
- Pädagogische Hochschule Luzern (8)
- Pädagogische Hochschule Nordwestschweiz (FHNW) (9)
- Pädagogische Hochschule St. Gallen (10)
- Pädagogische Hochschule Schaffhausen (11)
- Pädagogische Hochschule Schwyz (12)
- Pädagogische Hochschule Thurgau (13)
- Pädagogische Hochschule Zürich (14)
- Pädagogische Hochschule Zug (15)
- Schweizer Hochschule für Logopädie Rorschach SHLR (16)
- SUPSI - Dipartimento formazione e apprendimento (Ticino, SUPSI) (17)

Post this question:

If A1 = Higher Education Institution

A2 - 5 Please specify your answer given to the last question:

- Graduate Institute of International and Development Studies, Geneva, IHEID (1)
- Stiftung Universitäre Fernstudien Schweiz, Brig / Foundation University Distance Learning Switzerland in Brig (2)

Post this question:

If A1 = Accredited institution

A2 - 6 Please specify your answer given to the last question:

- Facoltà di Teologia di Lugano (1)
- Franklin University Switzerland (2)
- Staatsunabhängige Theologische Hochschule Basel (3)
- Theologische Hochschule Chur (THC) (4)

A3 To which research field(s) or discipline(s) does your research relate?

- Theology & religious studies, history, classical studies, archaeology, prehistory and early history (1)
- Linguistics and literature, philosophy (28)
- Art studies, musicology, theatre and film studies, architecture (29)
- Ethnology, social and human geography (26)
- Psychology, educational studies (27)
- Sociology, social work, political sciences, media and communication studies, health (24)
- Economics, law (25)
- Mathematics (22)
- Astronomy, Astrophysics and Space Science (23)
- Chemistry (30)
- Physics (2)
- Engineering Sciences (3)
- Environmental Sciences (4)
- Earth Sciences (5)
- Basic Biological Research (6)
- General Biology (7)
- Basic Medical Sciences (8)
- Experimental Medicine (9)
- Clinical Medicine (10)
- Preventive Medicine (Epidemiology/EarlyDiagnosis/Prevention) (11)
- Social Medicine (32)
- Other (33)

A4 Please specify your answer given to the last question:

_____

_____

_____

A7 Would you say that you are part of a **research community** (or of several ones)?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If A7 = Yes

A8 Which one(s)?

_____

_____

_____

Post this question:

If A7 = Yes

A9 How would you describe this "research community"?

- Community related to data (data research community) (1)
- Community related to a project (2)
- Community related to a research lab (3)
- Community related to a specific research discipline or field (4)
- Community related to an infrastructure (5)
- Other: (6) _____

End of block: Question for all

Start of block: With (data) communities (A5 or A7 = YES)

Characterization of community and research

B1 How is your community organised?/How are your communities organised?

- informally (1)
- in connection with an infrastructure (2)
- in connection with a project (3)
- as an association (4)
- on a social network (5)
- Other: (6) _____

B2 Is there a website for one or more of your communities?

- No (2)
- Yes (1)

Post this question:

If B2 = Yes

B3 Please give the website address. For the rest of the survey, please focus on one community in which you are the most active (enter the website of this community in first position).

B4 What is the size of your community (approximate number of researchers)?

B5 What is the typical scale of research in your community?

- National (1)
- European (2)
- International (3)
- Other (4) _____

B6 What is the typical time frame of research in your community?

- less than 1 year (1)
- between 1 and 5 years (2)
- between 5 and 10 years (3)
- more than 10 years (4)
- there is no typical time frame (5)

B7 What is the main nature of research data used in your community?

- Quantitative (1)
- Qualitative (2)
- Both (3)

B8 What is the main type(s) of research data used in your community?

- Comments (1)
- Experiences (2)
- Simulations (3)
- Derived data (4)
- References (5)
- Digitizations (6)
- Excavations (7)
- Artifact or media artifact (9)
- Other: (8) _____

B8 + What is the typical size of datasets in your community?

- kB / ko (1)
- MB / Mo (2)
- GB / Go (3)
- TB / To (4)
- PB / Po (5)
- Other (6) _____
- I do not know (7)

ORD activities/practices

B9 Which answer seems to describe your skills at best in **producing a Data Management Plan (DMP)**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

B9 + Which answer seems to describe your skills at best **in preparing your data for uploading them into a repository**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

B10 Which answer seems to describe your skills at best **in uploading research data into a repository**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

B11 Which answer seems to describe your skills at best **in sharing research data via a repository**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Post this question:

If B11 = I am able to do it alone

Or B11 = I am able to do it with help from my community

Or B11 = I am able to do it with help from my institution

Or B11 = I am able to do it with help from the data infrastructure I use

Or B11 = I am able to do it with help from a librarian

B12 Where does your community usually share its data?

- On a generalist repository (4)
- On an institutional repository (5)
- On a disciplinary or specialized repository (6)
- Other: (8) _____
- I do not know (7)

Post this question:

If B11 = I am able to do it alone

Or B11 = I am able to do it with help from my community

Or B11 = I am able to do it with help from my institution

Or B11 = I am able to do it with help from the data infrastructure I use

Or B11 = I am able to do it with help from a librarian

And if

B12 != I do not know

B12 + Please indicate the name of the repository you use:

_____

Post this question:

If B11 = I am able to do it alone

Or B11 = I am able to do it with help from my community

Or B11 = I am able to do it with help from my institution

Or B11 = I am able to do it with help from the data infrastructure I use

Or B11 = I am able to do it with help from a librarian

B13 To which kind of data do you usually give access?

- Active data (4)
- Final data (preserved at the end of the project) (5)
- Both (6)

B14 Which answer seems to describe your skills at best **in searching research data produced by others**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

B15 Which answer seems to describe your skills at best **in re-using research data produced by others**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Post this question:

If B15 = I am able to do it alone

Or B15 = I am able to do it with help from my community

Or B15 = I am able to do it with help from my institution

Or B15 = I am able to do it with help from the data infrastructure I use

Or B15 = I am able to do it with help from a librarian

B16 Which data do you re-use?

- Active data (1)
- Final data (preserved at the end of the project) (2)
- Both (3)

Post this question:

If B15 = I am able to do it alone

Or B15 = I am able to do it with help from my community

Or B15 = I am able to do it with help from my institution

Or B15 = I am able to do it with help from the data infrastructure I use

Or B15 = I am able to do it with help from a librarian

B17 Where does your community usually get the data it re-uses?

- From a generalist repository (4)
- From an institutional repository (5)
- From a disciplinary or specialized repository (6)
- Other: (8) _____
- I do not know (7)

Post this question:

If B15 = I am able to do it alone

Or B15 = I am able to do it with help from my community

Or B15 = I am able to do it with help from my institution

Or B15 = I am able to do it with help from the data infrastructure I use

Or B15 = I am able to do it with help from a librarian

And if

B17 != I do not know

B17 + Please indicate the name of the repository you use:

_____

B18 Which answer seems to describe your skills at best **in reproducing research results produced by others**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from my librarian (6)
- I do not know if I am able to do it (7)

Standardization

B19 Do standards exist in your community?

- No (1)
- Not yet, but some are under development (2)
- Yes, but they are not widely used (3)
- Yes, they are widely used (4)
- I do not know (5)

Post this question:

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B20 Do the researchers of your community use **metadata standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If B20 = Yes

B21 Which one(s)?

_____

Post this question:

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B22 Do the researchers of your community use **PIDs standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If B22 = Yes

B23 Which one(s)?

Post this question:

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B24 Do the researchers of your community use **file format standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If B24 = Yes

B25 Which one(s)?

_____

Post this question:

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B26 Do the researchers of your community use **documentation standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If B26 = Yes

B27 Which one(s)?

_____

Post this question:

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B28 Do the researchers of your community use **licence standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If B28 = Yes

B29 Which one(s)?

_____

Post this question:

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B30 Do the researchers of your community use **other standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If B30 = Yes

B31 Which one(s)?

Obstacles/concerns/motivations

B32 What are **your practical obstacles** for sharing and reusing data? (please specify a value on the scale according to their importance)

	Not important at all		Moderately important		Very important
	1	2	3	4	5

There are restricted conditions on the use of the data that I handle (and thus on sharing them) ()	
I do not have enough time to prepare the data and to make them accessible and reusable ()	
The typical time scale of my research is very long (and I will not share my data before the project is finished) ()	
I feel that I do not have the right skills to prepare my data in a way that makes them accessible and reusable ()	
I think that no appropriate technical tool is available for sharing my data and making them reusable ()	
I do not have access to the necessary support or coaching that I would need to make my data accessible and reusable ()	
Other: ()	

Post this question:

If B32 [There are restricted conditions on the use of the data that I handle (and thus on sharing them)] >= 2

B34 Which **restricted conditions** are there on the reuse of your data?

- legal restrictions (1)
- ethical restrictions (2)
- commercial restrictions (3)
- security restrictions (5)
- Other: (4) _____

B35 What are **your concerns** about following ORD practices? (please specify a value on the scale according to their importance)

Not important at all		Moderately important		Very important
1	2	3	4	5

I am concerned about my data being not well handled and used ()	
I feel that I might lose control on my data ()	
I am concerned that someone else publishes results from my data before I do ()	
I am concerned not to be properly acknowledged as the data producer ()	
I think that the benefit would be too low as compared to the efforts ()	
I do not see the point ()	
Other: ()	

B37 What are **your motivations** to follow ORD practices? (please specify a value on the scale according to their importance)

	Not important at all		Moderately important		Very important
	1	2	3	4	5
It complies with the requirements from funding agencies or publishers ()					
It brings me academic recognition (evaluation criterion for my academic career) ()					
It brings me scientific recognition ()					
I would like to benefit from the financial incentives existing for encouraging ORD practices ()					
It is very much practised in my community ()					
I commit to ORD because it corresponds to a personal belief ()					
Other: ()					

B39 Which actors could be more involved for better practices in ORD?

_____

End of block: With (data) communities (A5 or A7 = YES)

Start of block: Without (data) communities (A5 and A7 = No/?)

Characterization of research

C1 Why?

- There is no community for my discipline (1)
- I mainly follow individual research practices (2)
- Other: (3) _____

C2 In a general way, to whom do you show your research data?

- Some colleagues in my lab (1)
- Some colleagues from the same research project or programme (2)
- Some colleagues from the same institution (regardless of discipline) (3)
- Some colleagues in the same discipline (4)
- General public (citizen science) (5)
- No one (6)
- Other: (8) _____
- I do not know (7)

C3 From whom the research data may be of interest to you?

- Some colleagues in my lab (1)
- Some colleagues from the same research project or programme (2)
- Some colleagues from the same institution (regardless of discipline) (3)
- Some colleagues in the same discipline (4)
- General public (citizen science) (5)
- No one (6)
- Other: (8) _____
- I do not know (7)

C5 What is the typical scale of your research?

- National (1)
- European (2)
- International (3)
- Other (4) _____

C6 What is the typical time frame of your research?

- less than 1 year (1)
- between 1 and 5 years (2)
- between 5 and 10 years (3)
- more than 10 years (4)
- there is no typical time frame (5)

C7 What is the main nature of your research data?

- Quantitative (1)
- Qualitative (2)
- Both (3)
- I do not use data (4)

End of block: Without (data) communities (A5 and A7 = No/?)

Start of block: Without (data) communities (C7 != 'I do not use data')

C8 What is the main type(s) of your research data?

- Comments (1)
- Experiments (2)
- Simulations (3)
- Derived data (4)
- References (5)
- Digitizations (6)
- Excavations (7)
- Artifact or media artifact (9)
- Other: (8) _____

C8 + What is the typical size of your datasets?

- kB / ko (1)
- MB / Mo (2)
- GB / Go (3)
- TB / To (4)
- PB / Po (5)
- Other: (6) _____
- I do not know (7)

ORD activities/practices

C9 Which answer seems to describe your skills at best **in producing a DMP**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

C9 + Which answer seems to describe your skills at best **in preparing your data for uploading them into a repository**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

C10 Which answer seems to describe your skills at best **in uploading research data into a repository**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

C11 Which answer seems to describe your skills at best **in sharing research data via a repository**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Post this question:

If C11 = I am able to do it alone

Or C11 = I am able to do it with help from my community

Or C11 = I am able to do it with help from my institution

Or C11 = I am able to do it with help from the data infrastructure I use

Or C11 = I am able to do it with help from a librarian

C12 Where do you usually share your data?

- On a generalist repository (4)
- On an institutional repository (5)
- On a disciplinary or specialized repository (6)
- Other: (8) _____
- I do not know (7)

Post this question:

If C11 = I am able to do it alone

Or C11 = I am able to do it with help from my community

Or C11 = I am able to do it with help from my institution

Or C11 = I am able to do it with help from the data infrastructure I use

Or C11 = I am able to do it with help from a librarian

And if

C12 != I do not know

C12 + Please indicate the name of the repository you use:

_____

Post this question:

If C11 = I am able to do it alone

Or C11 = I am able to do it with help from my community

Or C11 = I am able to do it with help from my institution

Or C11 = I am able to do it with help from the data infrastructure I use

Or C11 = I am able to do it with help from a librarian

C13 To which kind of data do you usually give access?

- Active data (4)
- Final data (preserved at the end of the project) (5)
- Both (6)

C14 Which answer seems to describe your skills at best **in searching research data produced by others**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

C15 Which answer seems to describe your skills at best **in re-using research data produced by others**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Post this question:

If C15 = I am able to do it alone

Or C15 = I am able to do it with help from my community

Or C15 = I am able to do it with help from my institution

Or C15 = I am able to do it with help from the data infrastructure I use

Or C15 = I am able to do it with help from a librarian

C16 Which data do you re-use?

- Active data (1)
- Final data (preserved at the end of the project) (2)
- Both (3)

Post this question:

If C15 = I am able to do it alone

Or C15 = I am able to do it with help from my community

Or C15 = I am able to do it with help from my institution

Or C15 = I am able to do it with help from the data infrastructure I use

Or C15 = I am able to do it with help from a librarian

C17 Where do you usually get the data you re-use?

- From a generalist repository (4)
- From an institutional repository (5)
- From a disciplinary or specialized repository (6)
- Other: (8) _____
- I do not know (7)

Post this question:

If C15 = I am able to do it alone

Or C15 = I am able to do it with help from my community

Or C15 = I am able to do it with help from my institution

Or C15 = I am able to do it with help from the data infrastructure I use

Or C15 = I am able to do it with help from a librarian

And if

C17 != I do not know

C17 + Please indicate the name of the repository you use:

_____

C18 Which answer seems to describe your skills at best **in reproducing research results produced by others**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Standardization

C19 Are you aware of data standards?

- No (1)
- Not really, but I know that some of them are under development in my field (2)
- Yes, but they are not widely used in my field (3)
- Yes, they are widely used in my field (4)
- I do not know (5)

Post this question:

If C19 = Yes, they are widely used in my field

Or C19 = Yes, but they are not widely used in my field

C20 Do you use **metadata standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If C20 = Yes

C21 Which one(s)?

_____

Post this question:

If C19 = Yes, but they are not widely used in my field

Or C19 = Yes, they are widely used in my field

C22 Do you use **persistent identifiers data (PIDs) standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If C22 = Yes

C23 Which one(s)?

_____

Post this question:

If C19 = Yes, they are widely used in my field

Or C19 = Yes, but they are not widely used in my field

C24 Do you use **file format standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If C24 = Yes

C25 Which one(s)?

_____

_____

Post this question:

If C19 = Yes, but they are not widely used in my field

Or C19 = Yes, they are widely used in my field

C26 Do you use **documentation standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If C26 = Yes

C27 Which one(s)?

_____

_____

Post this question:

If C19 = Yes, but they are not widely used in my field

Or C19 = Yes, they are widely used in my field

C28 Do you use **license standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If C28 = Yes

C29 Which one(s)?

_____

_____

Post this question:

If C19 = Yes, but they are not widely used in my field

Or C19 = Yes, they are widely used in my field

C30 Do you use **other standards**?

- No (2)
- Yes (1)
- I do not know (3)

Post this question:

If C30 = Yes

C31 Which one(s)?

_____

Obstacles/concerns/motivations

C32 What are your **practical obstacles** for sharing and reusing data? (please specify a value on the scale according to their importance)

	Not important at all 1	2	Moderately important 3	4	Very important 5
There are restricted conditions on the use of the data that I handle (and thus on sharing them) ()					
I do not have enough time to prepare the data and make them accessible and reusable ()					
The typical scale time of my research is very long (and I will not share my data before the project is finished) ()					
I feel that I do not have the right skills to prepare my data in a way that makes them accessible and reusable ()					
I think that no appropriate technical tool is available for sharing my data and making them reusable ()					
I do not have access to the necessary support or coaching that I would need to make my data accessible and reusable ()					
Other: ()					

Post this question:

If C32 [There are restricted conditions on the use of the data that I handle (and thus on sharing them)] >= 2

C34 Which **restricted conditions** are there on the reuse of your data?

- legal restrictions (1)
- ethical restrictions (2)
- commercial restrictions (3)
- security restrictions (5)
- Other: (4) _____

C35 What are **your concerns** about following ORD practices? (please specify for them a value on the scale according to their importance)

	Not important at all		Moderately important		Very important
	1	2	3	4	5
I am concerned about my data being not well handled and used ()					
I feel that I might lose control on my data ()					
I am concerned that someone else publishes results from my data before I do ()					
I am concerned not to be properly acknowledged as the data producer ()					
I think that the benefit would be too low as compared to the efforts ()					
I do not see the interest ()					
Other ()					

C37 What are **your motivations** to follow ORD practices? (please specify a value on the scale according to their importance)

	Not important at all		Moderately important		Very important

	1 2 3 4 5
It complies with the requirements from funding agencies or publishers ()	
It brings me academic recognition (evaluation criterion for my academic career) ()	
It brings me scientific recognition ()	
I would like to benefit from the financial incentives existing for encouraging ORD practices ()	
It is very much practised in my community ()	
I commit to ORD because it corresponds to a personal belief ()	
Other: ()	

C39 Which actors could be more involved for better practices in ORD?

_____

_____

_____

_____

_____

End of block: Without (data) communities (C7 != 'I do not use data')

Start of block: Without (data) communities (C7 = 'I dot not use data')

D1 What are the main particularities of your research?

_____

_____

_____

_____

_____

D4 Who are the main funders of your research?

_____

_____

_____

_____

_____

D2 Do you publish in Open Access?

- Never (3)
- Sometimes (2)
- Frequently (1)
- I am not sure about what it means (4)

Appendix 5: List of websites from answers to the survey

ala.org	https://www.euro-online.org/websites/bor/
aforgen.wsl.ch/	https://www.europeansociology.org/
atlas.cern	https://www.gcb.uzh.ch/en.html
dada.zahnd.be	https://www.geo.uzh.ch/en/units/rsl
dhd-blog.org	https://www.globalbuddhism.org/jgb/index.php/jgb/
dkf.unibas.ch	https://www.hes-so.ch/accueil
exoplanets.eu	https://www.hets-fr.ch/fr/recherche/accueil
github.com	https://www.hetsl.ch/laress/
hfh.ch/sure	https://www.hevs.ch/fr/rad-instituts/institut-sante/
historikerinnen.ch	https://www.hevs.ch/fr/rad-instituts/institut-sante/activites-instituts/physiolab-11160
hssuisse.ch	https://www.iassidd.org/
http://aclals.net/	https://www.iavceivolcano.org
http://arthobservations.org	https://www.isanet.org
http://contemporarydrama.de/	https://www.isle-linguistics.org/
http://enfin.info/	https://www.linkedin.com/groups/13763674/
http://p3.snf.ch/Project-180350	https://www.manep.ch/
http://swsa.semanticweb.org	https://www.monitoringdemocracy.eu/
http://www.airhm.net/delegations/delegation-suisse/	https://www.nccr-catalysis.ch/
http://www.comparativecandidates.org/	https://www.postcolonialstudiesassociation.co.uk/
http://www.fremdsprachendidaktik.org/	https://www.sagw.ch/fr/seg/, ,
http://www.irscl.com/	https://www.saute.ch
https://aaac.world/	https://www.saute.ch/en/saute/
https://aom.org	https://www.sgg-ssl.ch/sgg/
https://blog.unifr.ch/carbonatelab/	https://www.sgr-sssr.ch/sgr-sssr/
https://brain.ieee.org	https://www.sgs-sss.ch/
https://cigev.unige.ch/	https://www.sgs-sss.ch/en/
https://citizensciences.net/	https://www.sib.swiss/
https://claire-ai.org	https://www.spectralbiology.org/
https://conbio.org	https://www.springer.com/journal/146
https://cses.org/	https://www.uni-due.de/biwi/politische-bildung/forschungskolloquium.php
https://data-innovation.org/	https://www.unifr.ch/cgf/en/research/research.html
https://dh-ch.ch/	https://www.unifr.ch/ecopol/fr/
https://dh-ch.ch/	https://www.unifr.ch/english/en/projects/civility-project/
https://dig-hum.de/	https://www.unifr.ch/med/de/research/groups/nordmann/
https://digitale-kunstgeschichte.de/	https://www.unige.ch/fapse/decisionlab/
https://digitalpsychotherapylab.ch	https://www.unige.ch/sciences/astro/variability/en/
https://droit.cuso.ch/accueil	https://www.unige.ch/sciences-societe/socio/fr/recherche/irs/index/
https://eaclals.eu/	https://www.unil.ch/cemep/fr/home.html
https://essenglish.org/	https://www.unil.ch/samemes/en/home.html

https://forum.openhardware.science/	https://www.unilu.ch/fakultaeten/tf/
https://gleon.org/	https://www.wsl.ch/de/ueber-die-wsl/programme-und-initiativen/forschungsinitiative-blue-green-biodiversity.html
https://hia4sd.net/	https://www.wsl.ch/en/about-wsl/programmes-and-initiatives/wsl-biodiversity-center.html
https://hummingbird.bio/	Https:/dasch.swiss
https://ibam.swiss/fr/home-2-main-fr/	infoclio.ch
https://iiif.io	many. Ei. infoclio.ch
https://institut-mehrsprachigkeit.ch/deit	nccr-onthemove.ch/
https://linked.art/	safmed.ch
https://lives-nccr.ch	scala-lang.org
https://mitglied.scnat.ch/sgw-ssbf	scta.info
https://musewiki.wikidot.com/musestmeetings	sites.google.com/fmach.it/g-bike-genetics-eu/home
https://nccr-onthemove.ch/	sjdm.org
https://nccr-robotics.ch/	stackexchange.com
https://pragmatics.international/default.aspx	stackoverflow.com
https://research.cs.wisc.edu/dbworld/browse.html	unige.ch/adp
https://researchdata.unibas.ch/en/	www.eahil.eu
https://sigchi.org/	www.ecargument.org
https://sslarch.github.io	www.esag.swiss
https://swissbias.github.io/	www.evoltree.eu
https://swissforestlab.wsl.ch/de/index.html	www.federalism.ch
https://unfccc.int/topics/land-use/workstreams/land-use--land-use-change-and-forestry-lulucf	www.gentree-h2020.eu
https://www.4sonline.org/	www.ighg.org
https://www.aric-interculturel.com/	www.lavater.com
https://www.asianstudies.org/	www.mlanet.org
https://www.aslo.org/	www.ophen.org
https://www.bioinspired-materials.ch/en/	www.pancare.eu
https://www.centre-lives.ch/	www.pancarefollowup.eu
https://www.dgff.de/	www.publicpolicyargument.eu
https://www.earli.org	www.Rai.uk
https://www.earlychildhoodresearch.ch/	www.selects.ch
https://www.eawag.ch/en/research/water-for-ecosystem/biodiversity/blue-green-biodiversity-research-initiative/	www.sphn.ch
https://www.ecolelasource.ch/vieillissement-sante/	www.spog.ch
https://www.egu.eu/cr/home/	www.zne.uzh.ch
https://www.elmi2021.org/	

Appendix 6: List of web sites or pages related to data communities

From the researchers' answers in the survey

https://dh-ch.ch/
https://musewiki.wikidot.com/musestmeetings
https://dig-hum.de/
https://nccr-robotics.ch/
https://www.sib.swiss/
http://arthobservations.org
http://enfin.info/
https://www.gcb.uzh.ch/en.html
https://www.elmi2021.org/
https://hummingbird.bio/
https://iiif.io/
www.sphn.ch
atlas.cern
https://www.earlychildhoodresearch.ch/
https://www.aric-interculturel.com/
http://www.comparativecandidates.org/
https://www.unige.ch/urbanhub/bienvenue/
https://forum.openhardware.science/
infoclio.ch
www.esag.swiss
https://unfccc.int/topics/land-use/workstreams/land-use--land-use-change-and-forestry-lulucf
https://cses.org/
https://www.monitoringdemocracy.eu/
https://www.geo.uzh.ch/en/units/rsl
https://www.spectralbiology.org/
https://swissforestlab.wsl.ch/de/index.html