Data Literacy, Collaboration and Sharing of Research Data Among Academics at the University of Iceland

Ágústa Pálsdóttir[(✉)]

Department of Information Science, University of Iceland,
Gimli v/Sæmundargötu, 101 Reykjavík, Iceland
agustap@hi.is

Abstract. Data literacy and data management have received growing attention in the past years, with open access and data sharing plans as a focus point. The study is part of an international research and the same questionnaire was used by all countries. The focus here is on data practices, collaboration and data sharing among academics at the University of Iceland. Response rate was 15%. Collaborating and sharing data with others was found to be a normal practice in the academics' work. This was done within a rather closed circle of researchers, whom they knew. Thus, their relationship is built on trust that has developed through the collaboration and partnership and it may be assumed that this forms the basis for data sharing. Offering open access to data to everyone was rare. Legal and ethical issues were the main causes of concern for sharing data but other reasons were also identified.

Keywords: Research collaboration · Data literacy · Data sharing
Open access

1 Introduction

Research data literacy and data management have received growing attention in the past years. As research funders have started to put emphasis on the valuables of the data and the possibility of reusing it, open access to research data and data sharing plans have become a focus point [1–3]. Data sets are increasingly considered part of the scholarly output and are being distributed as such [4].

This has started to create new demands on researchers, who are faced with complex challenges that they have not previously had to deal with. As a result, there is a growing interest among information specialists in how they can offer support to the research community on how to address these demands. In particular, information specialists who work in academic and research libraries need to be aware of these trends. It is expected that an important function of information service in the coming years will be to offer expertise service at developing data management plans, data description and preservation strategies [5].

Research data literacy has been identified as a sub-discipline of information literacy [6]. In a review article of the importance of data literacy for researchers and the

possible roles of academic and research libraries in research data management, Koltay [5] defined data literacy as "a specific skill set and knowledge base, which empowers individuals to transform data into information and into actionable knowledge by enabling them to access, interpret, critically assess, manage, and ethically use data" (p. 10). He furthermore pointed out that the definition corresponds strongly with the joint definition by UNESCO and IFLA (International Federation of Library and Information Association) of Media and Information Literacy, which allows individuals to "…access, retrieve, understand, evaluate and use, create, as well as share information and media content in all formats…" [7].

Koltay [5] further noted that data literacy should act as an incentive for researcher to share their data with others. In an examination of the data sharing practices among scientists, Tenopir et al. [8] found that if 85% of participants in their study were interested in using datasets from other researchers if it could be easily accessed. The researcher's willingness to share their own data, however, was not as great, because 46% do not make their data electronically available to others, while a similar share of them claimed to make at least some of their data available. Other studies have indicated an even lower interest among researchers for making their data openly available to everyone, with the share of those who do so ranging from 25% [9] to 15% [10]. It seems that data is mainly being shared with researchers own network of collaborators [9, 10].

The fact that so many researchers do not share their data openly may be caused by a variety of reasons. Managing research data for sharing is for example a time consuming activity that involves many steps [11], and it may reduce the researchers interest if they feel that they are not being rewarded for the effort that is needed. Tenopir et al. found that the main reasons reported for not sharing data with others were that the researchers did not have sufficient time for it, and a lack of funding. Other barriers reported were not having a place for the data, lack of standards, and that the sponsor does not require data sharing [8]. The fear of researchers regarding legal issues and the misuse of their data has also been described as major barriers for sharing data [9]. Thessen and Patterson have identified the need for both cultural changes and technical advances to encourage and promote data sharing. While the latter is considered easier to solve, cultural differences across disciplines and sub-disciplines call for better understanding [9]. Hatter et al. have added ethical challenges to these, and noted that researchers must also ensure that the ethical standards of academia are being met [12]. It has, furthermore, been pointed out that the data may be a vital asset for the researchers, which they use to promote their career [10].

The aim of the present study is to investigate the data literacy and attitude towards the sharing of research data, among academics at the University of Iceland. To do so, answers to the following research question will be sought: (1) What characterises the data sharing practices of the academics? (2) What challenges are the academics facing regarding data sharing? Recognition of the academics' attitudes towards the issue of collaboration and data sharing, and the problems and challenges that they are facing, is critical in order to design relevant support for them. The study results may provide important understanding about these issues. The findings can be used by universities to develop data policies and services that assist the research community.

2   Method

This is the Icelandic part of an international online survey that investigates research data literacy and data management, conducted among academic researchers and doctoral students at the University of Iceland. The same questionnaire was used by all participating countries. The online system Lime Survey, which is an open source software, was used for data collection. The survey was sent by email in February and March 2017, to a sample of 225 doctoral students and 567 academic researchers. The response rate for doctoral students was 25% (n = 56) and for academic researchers the response rate was 15% (n = 85). However, the current paper involves only results from the academic researchers. Of those academics who replied to the survey, 51.76% are from the field of social science, 35.30% from science and 12.94% from humanities. A total of 52.94% were men and 47.06 were women.

The questionnaire was translated from English to Icelandic. The measurement instrument consisted of 26 questions in all. A total of 18 questions emphasized data literacy, awareness of data management issues, the extent of collaboration and sharing of research data and the researchers' attitude towards this. In addition, the questionnaire consisted of seven background questions, and one open question.

As previously noted, the content of the survey is data literacy and research data management. It is, however, not possible within the constraints of this paper to investigate all the questions in the survey. Therefore, it was decided to focus on six questions about the following items: the file type of data that the participants normally use for their research, how they usually get the data for their research, how they usually use data that they get from others/outside sources; if they collaborate with other researchers and share data with them, what kind of access to their data they allow others, and finally, if they have any concerns for sharing data with others. All the questions allow multiple answers. The analysis of the data is descriptive.

3   Results

The chapter starts by presenting results about what file type of data the participants use for their research work. This will be followed by results about how they get data for research and what is needed by them to be able to use data from others. After that, results about collaboration with other researchers, data sharing practices, as well as open access and concerns regarding it will be presented.

Figure 1 shows that standard office documents are being used by the overwhelming majority of the participants, or 88.24%. After that come internet and web-based data (64.70%), images (58.82%) and structured scientific and statistical data (52.94%). What is by far the least common to use is configuration data (3.53%) and structured graphics (4.71%). In addition, very few of the participants claim to use software applications (8.24%) and encoded text (9.41%).

Results about how the participants usually get data for their research are presented in Fig. 2. The question contained five statements and the participants were asked to give an answer to all that applied to them.
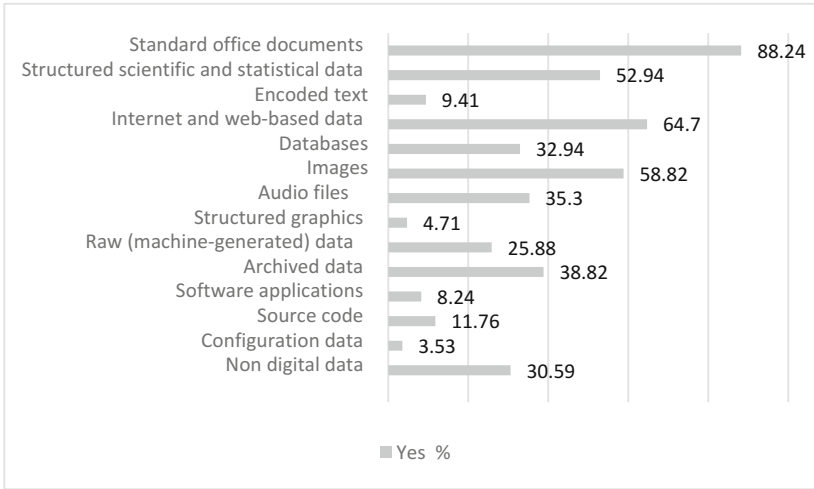
Fig. 1. Please indicate the file type of data that you normally use for your research
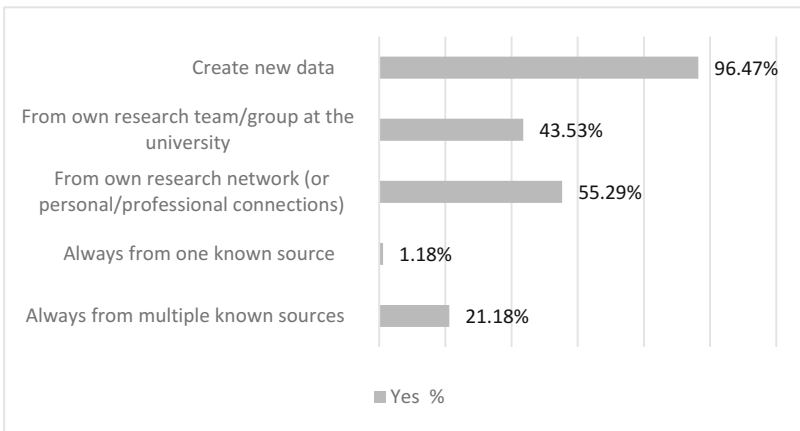

Fig. 2. How do you usually get the data for your research?

Creating new data for research work is what is most common among the participants. As can be seen from Fig. 2, the vast majority of them claim that they usually get data by doing so. Over half of them claim to get data from their research network, while less than half of the participants obtains it from their own research team or group at their university. Getting data always from one known source is almost unheard of among them and the majority of participants do not get data from multiple know sources.

The participants were asked how they usually use data that they get from others/outside sources. Four statements were presented and they were asked to give an answer to all statements that apply to them. The results are presented in Fig. 3.
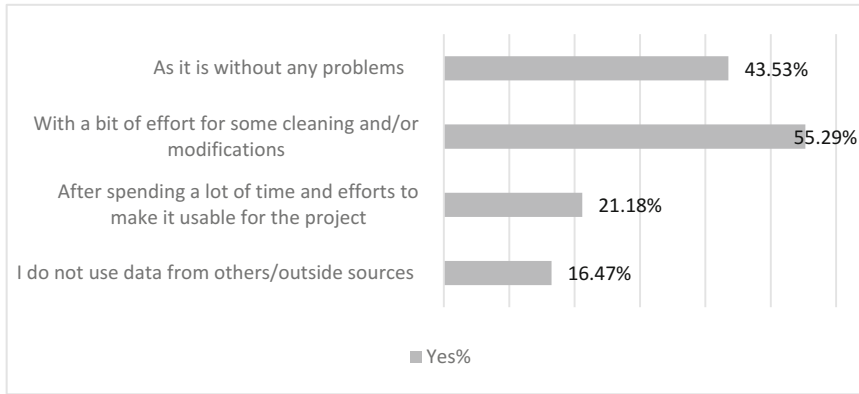

As it is without any problems — 43.53%
With a bit of effort for some cleaning and/or modifications — 55.29%
After spending a lot of time and efforts to make it usable for the project — 21.18%
I do not use data from others/outside sources — 16.47%
■ Yes%

Fig. 3. How do you usually use data that you get from others/outside sources?

From Fig. 3 it can be seen that when data from others/outside sources is being used, over half of the participants needs to put in some effort for cleaning and/or modifying it before the data can be used. Less than half of the participants state that they have no problems using the data as it is. Over twenty percent claim that a lot of time and effort may be needed for making the data usable.

When the participants were asked if they collaborate with other researchers and share their data with them, most of the participants, or 88.24%, claimed that they do so. It is most common (61.18%) to collaborate and share data with researchers that belong to the same research team as they do. In addition, more than half of the participants replied that they collaborate and share data with other researchers in the same university (56.47%) and half of them (50.59%) that they do so with researchers in other institutions.

Figure 4 presents results from two questions. The first one asked what kind of access to their research data the participants provide others with. The second question asked if they have any concerns for sharing their research data with others. At each question, five statements were presented and the participants were asked to give answers to all statements that apply to them.

As can be seen from Fig. 4, the great majority of participants, or 85.88%, do not allow open access to their research data to everyone. Nevertheless, a similar rate of participants (83.53%) do not agree with the statement that their data is not available to anyone else. What is most common is for the participants to have the data openly available to members of their research team (32.94%). A similar share of participants claim to allow access to their data upon request as those who allow access to only some parts of the dataset while other parts have restricted access.
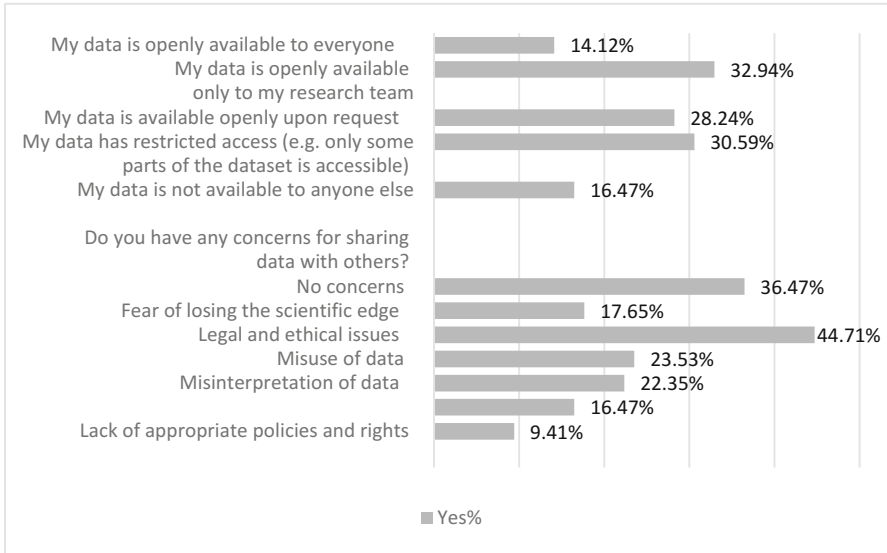
Fig. 4. Open access to research data – Concerns for sharing research data with others

The majority of the participants denied the statement that they had no concerns about sharing their data with others, hence they were worried about it, while over one third of them claimed that they had no concerns. What is most common is concerns about legal and ethical issues. A similar rate, or less than a quarter, of participants claimed that they had concerns about the misuse of shared data as was concerned about the misinterpretation of the data. On the other hand, only less a small rate of participants, or less than 10%, is concerned about the lack of appropriate policies and rights protection in relation to the sharing of data with others (Fig. 4).

4   Discussion

The study sought to contribute to current research into data literacy and the sharing of research data by presenting findings about academic researchers at the University of Iceland. For that purpose the study investigated various aspects related to the data practices, collaboration, open access to data, as well as the challenges and concerns that the academics facing.

The results show that an overwhelming majority of the participants (96.47%) create new data for their research work. When asked what file types they use for their data, the results show that standard office documents is by far the most commonly used (88.24%), which is consistent with the results from the PARSE Insight study (94%). In addition, the results that the majority of participants use web-based data (64.70%) and images (58.82%) is in line with results from the PARSE study [9]. The current study, however, also found that it was common to use structured scientific and statistical data (52.94%).

Besides, creating new data for their research, the majority of participants (55.29%) claim to get data from their research network or through personal/professional connections. Getting data from their own research team or group at their university is also rather common (43.53%). Using data from others however, requires some work on behalf of the participants, as over half of them (55.29%) need to put in some effort for cleaning and/or modifying it before the data can be used, while 43.53% say that they can use it without any problems.

The great majority of the participants claim to collaborate and share data with other researchers (88.24%), which is consistent with the results that a slightly less share of them (83.53%) disagrees with the statement that their data is not available to anyone else. Data is mainly being shared with members of the participants own research teams (61.8%), a finding that is in line with results from the PARSE Insight study (58%) [9], whereas the study by Research Information Network (RIN) found that a lower rate of their participants (37%) did so [10].

The current study, furthermore, shows that more than half of the participants (56.47%) collaborate and share data with other researchers in the same university, and that doing so with researchers in other institutions is only slightly less common (50.59%). This share is quite high compared with the results by Research Information Network (RIN) that 19% of researchers share data openly within their research community [10].

Open access to the research data, on the other hand, is not something that the participants want to offer to everyone, as the great majority of them, or 85.88%, claim that they do not allow it. This is the same result as in the study by Research Information Network (RIN), where 15% of researchers provided open access to their data [10], while the PARSE Insight study found that a higher share of researchers, or 25%, offered open access to data [9]. Results from the study by Tenopir et al. [8], however, show that 46% of those who replied to the question did not make their data electronically available to others.

What is most common is for the participants to have the data openly available to members of their own research team (67.06%). One way to allow others to make use of the research data, although it is not completely open, is to grant restricted access to it. This can be done, for example by granting admission to some parts of the datasets, or by allowing access to the data upon request. However, only about one third of the participants offered these options.

Prior studies have identified various barriers for researchers sharing data with others [8–12]. In this study, the majority of participants (63.53%) claimed to be worried about sharing data with others. Legal and ethical issues were the main cause of concern (44.71%), which is similar to the results found by the PARSE Insight study that 41% of researchers consider legal issues to be a barrier for sharing research data [9]. When other barriers are compared, the results in the current study show that concerns about the misuse of shared data (23.53%) and fear of losing the scientific edge (17%) was much less common and not comparable to the findings of the PARSE Insight study (41% and 27% respectively) [9].

It needs to be kept in mind that the response rate in the study was very low, only 15%. Therefore, the findings cannot be generalized to all academic researchers at the University of Iceland. Nevertheless, the results may provide important understanding

about their research data procedures, in particular what characterises their data sharing practices as well as the challenges that they face regarding it.

Taken together, the results indicate that it is usual for the academics in the study to collaborate and share data with others and that it may be considered a normal practice in their work. However, the results also indicate that this is done within a rather closed circle of other researchers whom they already know and have established a relationship of collaboration. It may be assumed that their relationship is built on trust that has developed through the partnership and that this forms the basis for data sharing. The great majority of the participants claims to have concerns for sharing data with others and does not offer open access to their data to everyone. Although legal and ethical issues were reported as the main cause of worry, other reasons were also identified. Thus, it is well possible that it is the combined effect of reasons, rather the individual causes, that has influence on the data sharing practices of the academics.

References

1. Haendel, M.A., Vasilevsky, N.A., Wirz, J.A.: Dealing with data: a case study on information and data management literacy. PLoS Biol. **10**(5), e1001339 (2012). https://doi.org/10.1371/journal.pbio.1001339
2. Koltay, T.: Data literacy: in search of a name and identity. J. Doc. **71**(2), 401–415 (2015)
3. Martin, E.R.: What is data literacy? J. eSci. Librariansh. **3**(1), 1–2 (2014)
4. McMillan, D.: Data sharing and discovery: what librarians need to know. J. Acad. Librariansh. **40**, 541–549 (2014). https://doi.org/10.1016/j.acalib.2014.06.011
5. Koltay, T.: Data literacy for researchers and data librarians. J. Librariansh. Inf. Sci. **49**(1), 3–14 (2017)
6. Schneider, R.: Research data literacy. In: Kurbanoğlu, S., Grassian, E., Mizrachi, D., Catts, R., Špiranec, S. (eds.) ECIL 2013. CCIS, vol. 397, pp. 134–140. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-03919-0_16
7. UNESCO: Media and Information Literacy (2014). http://www.uis.unesco.org/Communication/Pages/information-literacy.aspx
8. Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., et al.: Data sharing by scientists: practices and perceptions. PLoS One **6**(6), e21101 (2011). https://doi.org/10.1371/journal.pone.0021101
9. PARSE.Insight: Insight into Digital Preservation of Research Output in Europe: Survey Report (2009). http://libereurope.eu/wp-content/uploads/2010/01/PARSE.Insight.-Deliverable-D3.4-Survey-Report.-of-research-output-Europe-Title-of-Deliverable-Survey-Report.pdf
10. Research Information Network (RIN): If you build it, will they come? How researchers perceive and use web 2.0. Research Information Network Report (2010). http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/use-and-relevance-web-20-researchers
11. Thessen, A.E., Patterson, D.J.: Data issues in the life sciences. ZooKeys **150**, 15–51 (2011). https://doi.org/10.3897/zookeys.150.1766
12. Hartter, J., Ryan, S.J., MacKenzie, C.A., Parker, J.N., Strasser, C.A.: Spatially explicit data: stewardship and ethical challenges in science. PLoS Biol. **11**(9), e1001634 (2013). https://doi.org/10.1371/journal.pbio.1001634