Research Data Management in Turkey: A Survey to Build an Effective National Data Repository[1]

Güleda Doğan, Zehra Taşkın, Arsev Umur Aydinoglu

Abstract

Research data management (RDM) is an important topic for funding agencies, universities, and researchers. In this context, the main aim of this study is to collect preliminary information for Aperta, which is being developed by The Scientific and Technological Research Council of Turkey, to fulfil the following goals: determine the RDM awareness levels of researchers in Turkey; understand current RDM practices in their research environments; and find out their experiences about policy issues. For this, a questionnaire was distributed to 37,223 researchers, with 1577 researchers completing it. The results indicated that researchers who spend more time with data have more concerns about data management issues. The levels of experiences for creating a data management plan were quite low. The importance of this study lies in how it is able to show the current RDM practices of Turkish scholars during the new repository's foundational development stage.

Introduction

The data deluge in recent decades has created not also great opportunities for scientists but the challenge of managing enormous information. Discovering, accessing, storing, migrating, integrating, and re-using data have become easier and more common. Hence, research data management, which concerns the organization and dissemination of data (Whyte & Tedds, 2011), has become an important topic for all stakeholders involved in the research. It includes "the organization, storage, preservation, and sharing of data" (University of Pittsburgh, 2020). It has become both an opportunity (Royal Society, 2012) and a challenge (McAfee & Brynjolfsson, 2012; Pinfield, Cox & Smith, 2014) for research organizations.

Many researchers, funders and countries who understand the importance of research data management take important steps in this regard and carry out their studies in this direction. For researchers and funding agencies in Turkey, the biggest problem concerns the awareness of, and attitudes toward, RDM practices. Up until 2019, there has been a dearth of encouragement for researchers or a lack of drive to enforce guidelines, regarding research data management, from national funding agencies. However, researchers who are involved in the Horizon2020 grant scheme are required to submit research data management plans within six months after their project has been funded like any other grantee in Europe. Even though some researchers are aware of the benefits of RDM, they lack institutional support, have limited knowledge and technical skills, and are scared of falling victim to unethical practices such as being scooped (Allard & Aydinoglu, 2012; Aydinoglu et al., 2017; Ünal & Kurbanoğlu, 2018). Much needs to be done to address the problem but, at least those at the highest level are now offering commitment.

The head of the main funding agency in Turkey, TÜBİTAK, announced its commitment to open the data garnered during the 6th Turkey Open Science Summit held in 2018. In 2019,

[1] This paper is the author copy of the article accepted by *IFLA Journal*. Doi: 10.1177/0340035220917985

TÜBİTAK announced its Open Science Policy covering publications and research data that are produced through TÜBİTAK funding (TÜBİTAK, 2019). It recommends "establishing a research data management plan for open access to research data," "providing open access to publications along with research data," and also commits to preparing templates and guidelines for data management plans (TÜBİTAK, 2019, pp. 3-4). Moreover, an institutional repository, named Aperta, has been created (Aperta, 2019), and an RDM training portal (Araştırma Verileri Yönetimi, 2019) has been prepared. TÜBİTAK aims to provide open access to all kinds of funded projects' output, including research data and publications funded by TÜBİTAK's Support Program of International Scientific Publications (UBYT), TÜBİTAK-addressed publications, articles published in TÜBİTAK academic journals, and related research data (Aperta, 2019). On the other hand, the training portal provides templates, guidelines, and videos to researchers in need of training on RDM (Araştırma Verileri Yönetimi, 2019). The active initial steps that were taken by the country's main funding body may function as an accelerator regarding RDM and open research data in Turkish universities. In such an important process, this study aims to provide information regarding RDM behaviors of Turkish scholars which would help not only to the design of Aperta but also other data repositories.

Literature Review

RDM offers several opportunities not only to researchers but also to the scientific enterprise and society in general. As for researchers, because "proper data management is also a key prerequisite for effective data sharing" (Sesartic & Dieudé, 2017), one advantage is increased visibility for their research and, consequently, greater opportunities for their publications to receive more citations (Herold, 2015). This is because of evidence point to how studies that are made available receive more citations than similar studies that do not (Fecher et al., 2015; Ioannidis et al., 2009; Piwowar et al., 2007; Piwowar & Vision, 2013; Spires-Jones et al., 2016). As for the greater science, RDM helps in identifying questionable research ethics and in dealing with the reproducibility crisis, which has plagued science (John et al., 2012; Roettger et al., 2019). Integration and reuse of data spell the introduction of new methods and better science (Fecher et al., 2015; Tenopir et al., 2011; Xia et al., 2017).

However, challenges remain. Research suggests that, despite all the efforts by the scientific community, data sharing still tends to occur through personal exchanges (Fecher et al., 2015; Ferguson et al., 2014; MacMillan, 2014; Wallis et al., 2013). Policymakers have been encouraging researchers to share their research data and, in general, to have better data management practices. For instance, different funding agencies in the U.S. have adopted RDM policies into their grant schemes (NASA, 2010; NIH, 2008; NSF, 2008), and Europe followed (EC, 2016). Hingman and Pinfield (2015) looked at the adoption of data management practices at the Higher Education Institutes and found out that successful cases could be attributed to large research funders. Despite enviable open access data sharing practices, challenges persist. fMRI Data Center for the neuroimaging community is a good example (Mennes et al., 2103; Poldrack et al., 2017). Socio-cultural problems in ecology continue to hinder scientific advancement as well (Hardistry & Roberts, 2013). Dental research is plagued by similar concerns despite an environment of positive attitudes toward data sharing (Spallek et al., 2019). A recent survey conducted among earth and planetary geophysicists found that scientists are aware of the benefits of sharing their research data; however, they are concerned about data misuse and the risk of not receiving credit for their data sets (Tenopir et al., 2018). Data misuse as a concern for sharing data has appeared in different surveys (Aydinoglu et al., 2017; Bertzky & Stoll-Kleemann, 2009; Cragin et. Al, 2010; Elsayed & Saleh, 2018)

Management of research data creates an opportunity for library and information science practitioners and scholars. For academic institutions, university libraries play a critical role not only in managing research data but also in training the next generation of researchers to be well-versed in data management practices (Koltay, 2019; Tenopir et al., 2016). There exists a knowledge gap among scientists that can easily be addressed by these professionals (Steeleworthy, 2014; Strasser & Hamilton, 2012; Tenopir et al., 2017; Verbakel, 2013).

Several researchers think that incentives to researchers can be a solution to the socio-cultural barriers preventing data sharing in sciences (Ioannidis et al., 2014; Koole & Lakens, 2012; Michener, 2015; Nosek et al., 2012). Datasets can be cited and counted in the National Science Foundation grant applications as long as they are citable and accessible (Piwowar, 2013). Attitudes toward data sharing differ among scientists, depending on their academic disciplines (Tenopir et al., 2011; Tenopir et al., 2015a) or which sectors they belong to, such as if from academia or industry (Pollock, 2016). Even within a discipline, developments overtime could have changed the atmosphere and attitudes of scientists, such as the case of medical science that has been traditionally conservative regarding data sharing (Tenopir, 2011) but has now been moving toward the opposite direction (Yegros-Yegros & van Leeuwen, 2019).

Data citation is not the only thing researchers consider when it comes to RDM. Tenopir et al. (2011) conducted an international survey of researchers from different disciplines to explore current practices in, and perceptions toward, data sharing and found that scientists do not make their research data available nor do they receive institutional support, and only if certain conditions are met, they are willing to share their data. A follow-up study (Tenopir et al., 2015b) to observe the changes (if any) regarding data sharing was conducted amidst how many funding agencies made data management plans mandatory for their grantees and how awareness had increased. Apart from the increase in acceptance and willingness to engage in data sharing, actual data sharing behaviors also saw a spike. However, with greater awareness, risk perception among the participants had also spiked as barriers remained. Other surveys found similar results (Aydinoglu et al., 2014; Aydinoglu et al., 2017; Grootveld et al., 2018; Kratz & Strasser, 2014; Whitmire et al., 2015).

There have been a limited number of studies investigating RDM practices in Turkey. A nationwide survey found that the concept of RDM did not exist in open access policy papers (Tonta, 2012; Tonta, 2013). The studies so far have focused on the researcher's attitudes towards and practices of RDM and data sharing. Allard and Aydinoglu (2012) investigated data sharing practices among environmental scientists; Aydinoglu, Dogan, and Taskin (2017) examined RDM behaviors of academicians of research intensive universities in Turkey; Unal and Kurbanoglu (2018) researched attitudes towards RDM. A recent study by Unal et al. (2019) compared scholars from Turkey to France and the UK and found out that there are big differences in data behaviors such as "the use of data from outside sources", "expectations for funding for data storage and open access", and "concerns for sharing their data". It is important to have a better understanding of scholars data behaviors in comparison to each other as they collaborate frequently and what to do with the data consists a big chunk of the collaboration. Considering that Tubitak Open Science Policy is in effect and Aperta has been introduced to researchers; it has become imperative to get a baseline for Turkish scholards RDM behaviors in order to understand the mid- and long-term effects of the policy and archieve. Therefore, this study is designed to understand the RDM behaviors of the target audience of Tubitak funding: active researchers who received funding from TÜBİTAK. by addressing the following research questions:

- What is the level of the use, production, and citing of research data for researchers in Turkey? Do these practices differ by academic title and field of study?

- What are the most common data types and data formats used in research? Is there a difference between the fields according to the data types/formats used?
- What is the size of the data used in research? Is the field of study a determinant of the data size used?
- Where is the research data stored? Does the environment for data storage differ by academic title and study of the field?
- Do researchers support open access to research data? What are their experiences in preparing a Data Management Plan (DMP)? How are the approach of researchers for TÜBİTAK making data open and DMP mandatory for funded projects? Does the approach of researchers differ by field and title?
- What is the need for training on RDM? Does the need differ by title and the area?

Methodology

TÜBİTAK ULAKBİM (Turkish Academic Network and Information Center) sent a questionnaire to 37,223 researchers registered to ARBİS (Researcher Information System of TÜBİTAK) to gather the preliminary information for *Aperta,* mainly on RDM related issues. ARBİS, which is designed and developed by TÜBİTAK, is an updated database to hold information about researchers. The researchers who have registered to ARBİS are able to apply for TÜBİTAK scholarship and support programs or are able to serve as evaluation and monitoring phases for the submitted proposals (ARBİS, 2019). ARBİS is one of the biggest researcher repositories of Turkey. On the other hand, since the Aperta is aimed at researchers who conduct or evaluate projects for TÜBİTAK, only scholars registered to ARBİS are included in the research. The questionnaire was distributed online by TÜBİTAK ULAKBİM through LimeSurvey and remained open for three weeks between April 27 and May 17 in 2018.

The questionnaire consisting of 19 questions was answered by 1577 researchers during this period. Researchers were asked questions about research data usage and production; the types, formats, and the size of data they used/produced; the environment they store their research data in, and so on. The aim was to reveal trends and the behaviors and thoughts of researchers, as well as to determine the knowledge levels and educational needs of researchers.

The participants were asked three initial questions about whether they used, produced, and cited research data before. Since these three questions were not answered in 269 returned questionnaires, the analysis was based on 1308 returned questionnaires (1577 − 269), where the participants answered the initial questions.

According to Formula 1 and 2 (Cochran, 1963, p. 75), a population of 37,223 can be represented by a sample of 1736 for e = 0.04 and with by a sample of 996 for e = 0.03 both at 99% confidence level (z = 2.56; p = 0.5; q = 0.5). Based on this information, it is possible to say that our sample represents the population at 99% confidence level.

$$n_0 = \frac{z^2 pq}{e^2} \qquad \textit{(Formula 1)}$$

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} \qquad \textit{(Formula 2)}$$

In Formula 1 and 2:

N*: Population size*

n_0*: Sample size*

n*: Corrected sample size*

z: *Z table score for the selected confidence interval*
p: *Estimate of variance*
q: $1 - p$
e: *Desired level of precision*

SPSS (version 21.0) was used for the analysis of the results, and Excel was used to generate the graphs. In addition to descriptive statistics, cross-tables and chi-square tests were used to reveal the differences according to fields and titles. Interpretations were made based on percentages (row percentages) by field and title to prevent results from being affected by frequency by field or title. In addition to the comparisons made based on the fields and titles, several comparisons were made to determine whether there was a difference between those who used/produce and did not use/produce research data according to their approach to open data and research data management.

Findings

Respondents' General Information

The distribution of 1308 respondents covered in the analysis according to the field of the study showed that 27% were in engineering, 24% in science, 21% in medical and health sciences, 18% in social sciences and humanities, and 9% in agricultural sciences. Fifteen (15) respondents did not specify their field of study. Figure 1 shows the distribution by title. The majority, around 40%, were professors, followed by associate professors (22.5%), and assistant professors/lecturers with a Ph.D. degree group (24%). These three groups accounted for approximately 87% of respondents, with the remaining 13% consisting of lecturers without a Ph.D. degree (3.4%), research assistants with a Ph.D. degree (5.1%), and research assistants in the process of post-graduate education (2.8%).

Figure 1. Distributions of respondents by title (%)

RDM Awareness Levels of Turkish Scholars

The research data usage rate of the participants was about 83%. Approximately 73% of the participants said that they cited research data before, and the option "I did not know that data could be cited" was not selected by anyone. A group of approximately 71% of the participants said that they produced their research data (Figure 2). The tendencies of researchers to use, produce, and cite research data were also examined based on the field of study and title of the researchers. While the rate of researchers who stated that they used research data was higher than 77% for all five fields of study, the highest rate was observed in the medical and health sciences (91%) and agricultural sciences (90%). A similar trend in producing and citing research data was observed. Medical and health sciences (83%, 76%, respectively) and agricultural sciences (76%, 83%, respectively) were the two fields of study that exhibited the highest rate of producing and citing research data. Social sciences and humanities had the lowest rate of producing (63%) and citing research data (68%). The titles of researchers were

divided into four groups (1. Professor, 2. Associate Professor 3. Assistant Professor and Lecturer- with or without Ph.D. 4. Research Assistant- with or without PhD and other) and comparisons by title showed that there was no statistically significant difference in terms of the title in all three aspects of researchers involvement with research information (using research data: $\chi^2_{(3)} = 6.121$, $p = 0.106$; citing research data: $\chi^2_{(3)} = 7.706$, $p = 0.052$; producing research data: $\chi^2_{(3)} = 1.012$, $p = 0.798$). The rate of researchers who stated that they produced (70%–74%), used (79%–85%), and cited research data (68%–76%) were found close to each other for all four title groups.

Commonly Used Data Types and Formats

The most commonly produced/used data type was found to be experimental data (50%), indicated by one in every two participants. The next most commonly used or produced data type was text data (24%), followed closely by survey data (22.5%) and graphic data (22%). It would not be wrong to say that these three data types were produced or used by approximately one in four participants (Table 1). In exploring how the data formats used differed according to the field, it was found that the most important difference was in the use of experimental data ($\chi^2_{(4)} = 139.577$, $p = 0.000$). The most significant reason for this difference is that the use of experimental data was seen in 67% of researchers in the medical and health sciences and 54%–58% in engineering, science, and agricultural sciences, whereas it was used by only 17.4% of researchers in the social sciences and humanities. Contrary to that found for experimental data use, the highest rate of raw data use, although not of the same significance, was found in the social sciences and humanities (21%). The usage of survey data by field showed that the highest usage rate was in social sciences and humanities (46.4%) and then in medical and health sciences (37%), with the lowest level of use being in engineering (8.2%) and sciences (10.2%) ($\chi^2_{(4)} = 181.747$, $p = 0.000$). The two most prominent fields in terms of the use of graphical data were sciences (25%) and engineering (30%). Worth noting is that data models were used more in engineering compared to other fields (13%). The use of lab books, which had no use in social sciences and humanities, was found to be higher in medical and health sciences (21%) and sciences (16.6%) than in the other two fields ($\chi^2_{(4)} = 70.901$, $p = 0.000$). Audio record and video data, both of which had a low usage in general, were used more in the social sciences and humanities field (15% and 17.7 respectively) than in other fields.

Table 1. Distribution of used/produced data types

Data type	%
Experimental data	50.1
Text data	23.9
Survey data	22.5
Graph data	22.2
Raw data	13.1
Lab book	10.9
Data model	8.9
Audio record	5.1
Video	4.3
Remote sensing data	3.3

Note: Multiple options could be selected for this question.

According to 4% of participants, the type of data they used was different from the options seen in Table 1, but only 10 of them specified the data type as abiotic data, data from

non-living physical and chemical elements in the ecosystem, (5 participants) and observational data (5 participants). Another striking finding about the data types used/produced was that 36% of the participants responded that they did not use any of the 10 data types listed in Table 1 and did not specify any other data type in the “other” option, which is a figure higher than for those who stated that they did not use (17%) and did not produce (29%) research data.

Table 2. Distribution of used/produced data formats

Data format	%
xls	49.5
txt	34.8
free text	16.4
sav	14.3
csv	13.5
spreadsheet	9.6
readme structured	2.1
readme unstructured	1.0

Note: Multiple options could be selected for this question.

Based on Table 2, which shows the percentages of data formats produced/used, the two data formats that stood out were xls (49.5%) and txt (34.8%), used by approximately one in two people. Social sciences and humanities had the lowest rate of use of xls (39.6%) and txt (25.5%) data formats, with these two data formats being used by approximately one in two researchers in the other fields. The use of csv was found to be higher in engineering (19.4%) and agricultural sciences (18.6%). The free text data format was used at a higher rate in medical and health sciences (23.4%) compared to the use in other fields. Researchers’ use of the sav data format, which was about 8% in engineering, sciences, and agricultural sciences, was 29% in medical and health sciences, and 25% in social sciences and humanities. This finding may be related to the fact that survey data are generally highly used in these two fields. The most commonly used data formats for survey data were found to be xls (75%) and sav (50%). Upon examination of how the data formats differ according to the other data types used, xls and txt data formats stood out for being the most used data types, as shown in Table 1 (Figure 2).

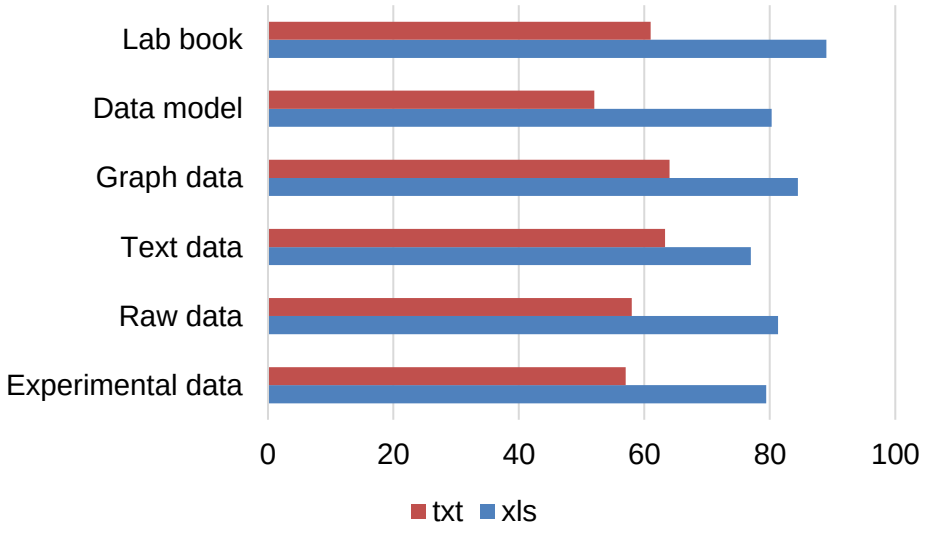

Figure 2. Use of xls and txt data formats by most commonly used data types (%)

When the data written to the “other” option part that was marked by 6% of the participants were examined, it was found that 16 researchers (approximately 1%) seemingly considered word/doc as a type of data format. Besides, although the data format of the statistical software SPSS is sav, two researchers specified SPSS as a data format in the “other” option part. Additionally, pdf was also specified as a data format. Another interesting point for the “other” option part was the picture/image and video formats (jpeg, tiff, png, mp4, etc.).

Average Data Size

It is thought that 36% of the participants who did not mark any of the options related to the average size of the data they used/produced in their most recent studies did not have sufficient working idea about the subject when they began. The percentage of researchers who did not indicate the size of the data used/produced was highest in social sciences and humanities (45%) and lowest in medical and health sciences (24%). This information should be evaluated together with the percentage of those who stated that they did not produce nor use research data in the related fields. The percentages of participants who did not produce and use research data for social and humanities were 37% and 19%, respectively, which were higher compared to the rate in medical and health sciences (17% and 9%, respectively). The percentage of participants who produced/used data larger than 10 GB was around 7%. In general, it can be said that the data used/produced were not “big data.” The size of the data used/produced by 39% of the participants was less than 1 GB. Concerning data for social sciences and humanities, 70% of the researchers replied that the amount of data they used or produced was less than 1 GB. The rate ranges from 58% to 62% for the other four areas.

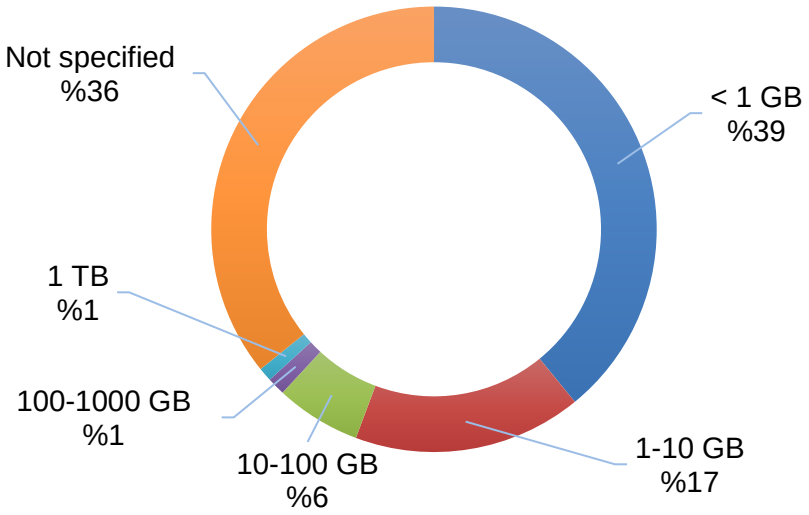

Figure 3. Distribution of the data used/produced by the participants in their latest study according to average data size

Figure 4 presents the findings on where the participants store the most recent data they produced/used. The most preferred medium for storing data was local computers (61.5%), with the highest rate (73.4%) seen in medical and health sciences and the lowest rate in social sciences and humanities (52.3%). Participants’ use of cloud storage was 7% and 9.6% for agricultural sciences and sciences, respectively. Although there was a “cloud” option among the options, it is interesting that “drive” and “dropbox” were indicated (as other mediums that the data is stored) under the “other” option by some of the participants who

selected the said option (5%). The storage medium specified by the vast majority of the participants who selected the "other" option was the "external disk." It can also be seen from Figure 4 that the use of institutional repositories, open access archives, etc. for data storage has not yet become widespread in Turkey. There were only 11 participants who stated that they used commercial databases such as the *Data Citation Index* for data storage. Although 71% of the respondents stated that they cited research data, they seemed hesitant to store their data in institutional repositories, open access archives, data archives, and other storage options that they believe could facilitate or make it possible for other researchers to cite their data. The use of open access archives, institutional repositories, and commercial databases for data storage was quite low in all areas.

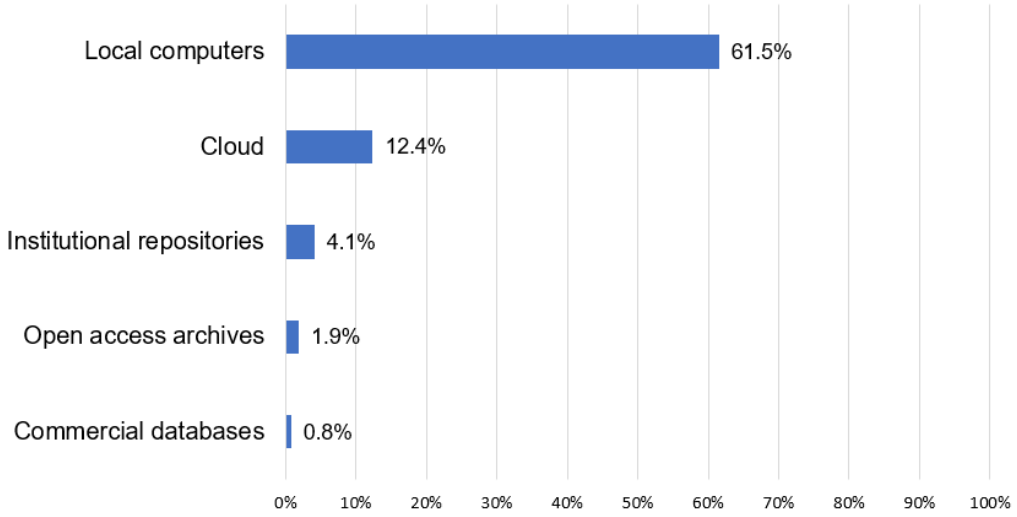

Figure 4. Storage environments of the most recently used/produced data (*Note: Multiple options could be selected for this question)*

On whether the data storage environment differed according to the title, there was a statistically significant difference in terms of cloud usage at a 95% confidence level ($\chi^2_{(3)}$ = 32.114, p = 0.000). It was found that professors (7.4% cloud usage) and research assistants' group (25% cloud usage) caused this difference.

Data Management Plans (DMPs)

In order to improve the answers to the questions about the DMP, the participants were asked whether they had previously conducted a project, with about 76% of the 1205 respondents replying yes. Approximately half (454 participants) of the 913 participants who answered the question of whether they had previously conducted a TUBITAK project answered in the affirmative. On the question of whether DMP was mandatory (at least once) for the funding of their previous projects (at least one of their previous projects), 121 of the participants (approximately 9% of the participants and approximately 14% of the respondents of the related question) answered affirmatively. This finding is supported by the study of Ünal ve Kurbanoğlu (2018, p. 299), who found that 13% of participants had previously prepared DMPs, and 16% of participants had DMPs in their current projects. Researchers in the medical and health sciences exhibited the highest rate (14.2%), while those in the social sciences and humanities showed the lowest rate (6.4%), in terms of prior preparation of DMPs. It is thought that 427 participants (approximately 33% of the respondents) who did not answer the related question had no idea about DMPs. Based on these findings, it would not be wrong to say that approximately 91% of the respondents did not know what a DMP is and/or had never prepared

a DMP before. Approximately half of the participants (49%) thought that open data and DMP that are mandatory in Horizon 2020 projects should also be mandatory for TUBITAK projects. Only 7% of the participants expressed negative opinions on this issue. It is thought that the 31% who did not express their opinions did not have any knowledge about the subject. Whether or not to use research data did not affect the opinion of the researchers ($\chi^2_{(5)} = 6.603$, $p = 0.252$) on said issue. The question of whether to produce research data cannot be said to have cast an important effect ($\chi^2_{(5)} = 11.196$, $p = 0.048$), even as those who did not produce research data had a higher positive opinion about DMPs (77.3%) in comparison with those who produced research data (68.4%). The distribution of 903 participants who gave their opinion about the necessity of mandatory DMPs for TUBITAK projects by field showed that the positive opinion ratio for each field ranged between 69.2% and 73.9% and that there was no statistically significant difference between the fields ($\chi^2_{(20)} = 7.687$, $p = 0.994$). Similarly, there was no statistically significant difference between titles ($\chi^2_{(15)} = 11.340$, $p = 0.728$), although the research assistants' group had the highest rate of positive opinions on making DMPs mandatory for TUBITAK (79%).

Training Needs

When asked for their opinions on funding institutions' (TUBITAK) checking of research data produced during projects, 40% of them gave their opinions, which were significantly positive. The majority believed that the person responsible for the control and supervision of the data was none other than the researcher who has produced or produces the data. It was stated that TUBITAK should wield control of research data under the supervision of experts, and it was understood that researchers had many reservations about data security and ethical violations.

This study also found that 67% of the participants and 90.5% of the respondents for the related question stated that they were willing to participate in data management training to hold by TUBITAK and to use open access resources and portals that will be created by TUBITAK. The presence of a group of 371 researchers who did not give either a positive and negative opinion on the subject is remarkable (approximately 28% of the participants). Although it was found that there was no significant difference in terms of educational needs according to the titles ($\chi^2_{(3)} = 8.138$, $p = 0.043$), it can be said that demand for training is lower for the associate professors (90%) and professors (88%) in comparison to other groups. Willingness for education in different fields ranged from 87% to 93%. ($\chi^2_{(4)} = 3.311$, $p = 0.507$). Besides, those who used and/or produced research data and those who stated that they did not were equally eager for training. One explanation for this was the presence of participants who declared that they used/produced research data but stated that they did not know about it.

The majority (59%) of the participants (76% of the respondents of the related question) thought that data produced by public sources is a public good and should be made available to the public as open data. The number of participants who gave negative opinions to this question was very low (8.2% of the participants, 10.6% of the respondents of the related question). The opinions of the research data users and non-users about open data were found to be similar ($\chi^2_{(5)} = 5.658$, $p = 0.341$), but the question of whether to produce research data affected the idea ($\chi^2_{(5)} = 19.801$; $p = 0.001$). An overwhelming majority of the researchers believed that data generated by public resources should be open to the public: this applied both to those who produced research data and those who did not, at 72.4% and 85%, respectively. The opinion on this subject was found to differ according to the fields ($\chi^2_{(20)} = 38.287$, $p = 0.008$). The reason for this difference is that the rate of supporting open data approach was approximately 86% in the social sciences and humanities but between 71%

and 76.5% in the other four fields of study. Although the participants' titles did not form an important factor in the opinion about open data ($\chi^2_{(15)} = 11.187$, $p = 0.739$), the rate of those who gave positive opinions was relatively higher in the research assistants group (80.4%). Examining the opinions expressed by 121 participants showed that the most prominent opinion could be considered the basis of the philosophy of open access, which is that "all scientific outputs produced by public resources belong to the public." It was stated that data sharing would prevent wastage of resources, would expand comprehensive research, and could wield widespread impact. The participants also expressed concerns about several issues, which were the use of data to advance personal instead of public interest, data with a high degree of confidentiality, anonymization of the data, guaranteeing the control and governance of data, how to determine the conditions relating to the use of the data, and protection of the rights of the researchers.

Discussion

This study utilized survey data aimed to collect preliminary information for Aperta, the Institutional Repository of TUBITAK, which is being developed by TUBITAK ULAKBIM, and to obtain information about ARBIS registered researchers' level of their knowledge and about the current situation in research data, research data management, open data, and related subjects.

1. **RDM awareness** was found to be high among the survey participants compared to the earlier studies conducted in Turkey (Allard & Aydinoglu, 2012 & Aydinoglu, Dogan & Taskin, 2017). It seems that two domains, the medical & health sciences and agricultural sciences, led the efforts in utilizing, producing, and citing research. Although social sciences exhibited the least efforts from among the five domains, researchers in the social sciences expressed the highest support for the argument that publicly funded research data should be made into open data. Furthermore, the scholars who did not generate research data were more in favor of open data policies. These arguments demonstrate that researchers who spend more time with research data have more reservations and concerns about open data.

The most surprising finding of the study concerns the measure of data citation. Even though the previous studies (Allard & Aydinoglu, 2012; Aydinoglu et al., 2017; Ünal & Kurbanoğlu, 2018) found that the notion and practice of data citation were not common among researchers, three out of every four respondents stated that they had previously engaged in data citation and knew about it. This can be interpreted as a positive sign that the efforts undertaken by TUBITAK, the main funding agency in Turkey, to increase awareness in RDM and data sharing have finally yielded some results. However, we also suspect that the respondents refer to something else when they think "data," such as a table or a graph in an article they see.

2. As for **data types, formats, and size**, experimental data that one of two participants used was extensively in medical and health sciences, sciences, agricultural sciences, and engineering. Not surprisingly, its usage in social sciences and humanities was quite low compared to other fields. Besides, data models were used more in engineering, and it may be a result of using big data. On the contrary, it was found that the survey data mostly used in social sciences and humanities and the medical and health sciences had the second-highest usage rate. Aydinoglu, Dogan, and Taskin (2017, pp. 276-277) found that the most commonly used data types were the same for this study, there were significant differences in the usage rate for data types other than for experimental data (53%) text data 47% and survey data (41%).

The highest use of the .sav data format, which is the data format of SPSS used widely for analyzing survey results, is again in these two fields of study. As the main subject of studies is human beings in the social sciences and humanities, audio record and video data were used more

in this field. In general, .xls and .txt. data formats are the most commonly used data formats, but they have the lowest usage rate in social sciences and humanities. These two data formats are the most commonly used data formats for each of the most used data types. In Aydinoglu, Dogan ve Taskin (2017, p. 278), the two most commonly used data formats had a proximate rate of use. The third (free text 30%) and fourth (sav 27.4%) commonly used data types were also the same in both studies, but the findings of Aydinoglu, Dogan, and Taskin (2017, p. 278) showed a higher usage rate. The usage of .sav files may indicate a problem as they are not recommended for archiving and publishing as it hinders the interoperability of research data.

Ünal and Kurbanoğlu (2018, p. 296) and Aydinoglu, Dogan, and Taskin (2017, p. 279) obtained generally similar results in terms of the data storage environments and TB size data usage and production of the researchers. Both studies found high use of participants' own devices for data storage (96% and 71.6% respectively) and of in-cloud storage (39% and 46% respectively). The cloud storage preference among early career researchers was twice more than professors which correspond to Aydinoglu, Dogan, and Taskin study (2017). Similar to the results of this study, Ünal and Kurbanoğlu (2018, p. 296) found that the use of university institutional archives (9%) and external institutional archives (6%) for data storage purposes was also very low. That being said, the Council of Higher Education initiated an "Open Academic Archieve System" which would increase the use of university institutional archives and indirectly help researchers adopt better data behaviors (the Council of Higher Education, 2019).

3. Although a direct question was asked about the **need for training** in the scope of the study, the answers given to some questions also show the level of knowledge about the subject and hence indicate the need for training. In fact, previous studies also identified the need for training in RDM (Aydinoglu, Dogan & Taskin, 2017; Ünal & Kurbanoğlu, 2018) and documented the demand from their participants for such training (Allard & Aydinoglu, 2012). Approximately four out of 10 participants did not specify the data type and data format they used/produced, and the size of the data they used in their most recent study. It is natural for participants that did not use and produce research data to not respond to this question, but it is understood that a considerable number of participants who produced and/or used research data also do not respond to these questions. For example, in the field of social sciences and humanities, although 19% indicated that they did not use research data and 37% indicated that they did not produce research data, almost one in two participants (45%) did not (could not) state the size of the data they used in their most recent study. Some of the participants—who answered the question about the data type they used, erroneously answered the data format question with Word, Doc, SPSS, PDF, and data storage question with Drive, Dropbox, etc. instead of choosing "cloud" option—may also be added to the group that needs training. On the other hand, although some participants engaged in using/producing research data, they indicated their need or willingness to undergo training.

Not only technical solutions such as Aperta but also socio-cultural issues have to be taken into consideration. Despite an attitude of willingness, there is a serious training need for the scientific community in Turkey. A great majority of the participants expressed their desire to attend training on RDM. Based on some of the responses, we suspect that the knowledge of the concept of data, metadata, interoperable data formats, etc. are not very clear in the minds of the participants. A small percentage of the researchers in Turkey have been managing research data, but incentives and mechanisms can be designed to spread their best practices. The training needs are twofold: First, the information professionals in Turkey need training. There is no study on how much they know about RDM (and that needs to investigate in a future study); however, the information professionals in other countries needed training to support the research community better (Tenopir et al., 2015; Tenopir et al., 2017; Wittenberg,

Sackman & Jaffe, 2018) and probably that would be the case for Turkey as well. Second, the researchers themselves need training to take better care of their research data as the participants of this study expressed and there are plenty of RDM training experiences that can be inspired upon (LEARN, 2017; Bishop et al., 2020; Sesartic & Dieudé, 2017).

4. The responses to **DMP** related questions revelated that researchers in Turkey have limited experience in preparing DMP even less than their international partners (Tenopir et al., 2015). Results of this study and other similar studies (Aydinoglu, Dogan & Taskin, 2017, p. 280; Ünal ve Kurbanoğlu, 2018, p. 299) demonstrated that researchers in Turkey do not have sufficient knowledge about DMPs despite TUBITAK's initiative to make DMPs mandatory for TUBITAK-funded projects. Ünal ve Kurbanoğlu (2018, p. 299) found that the rate of those who did not know whether their institution has a DMP or not was 73%, with more than 50% of the participants unaware what a DMP is; yet 84% of them stated that universities should have DMPs. Despite these results, almost half of the respondents were positive about DMPs as being mandatory for TUBITAK-funded projects.

There was also a contradiction between what had been expressed and done. When statements (such as support to make research data that have been publicly funded to be open data, made by six out of 10 respondents) were compared to the amount of actual research data that have been deposited or shared in organizational archives, open access archives, and commercial repositories, huge discrepancies were found. These findings revealed the confusion in the scientific community on how to deal with RDM. There exist not only awareness and goodwill but also trust issues and poor data habits.

Conclusion

Policies and tools designed to promote RDM and data sharing should consider the issues discussed above. Incentives (or their lack of) impact on how the scientific community adopts data sharing but unfortunately, there has been no incentive at all; however, TÜBİTAK Open Science Policy promises to change things at the funding agency level. DergiPark, which is the online journal hosting system for academic journals in Turkey with approximately 2000 journals, can enforce data sharing practices for the articles they publish. TÜBİTAK can nudge DergiPark in this direction. Furthermore, TÜBİTAK can ask for DMPs in grant applications and give additional points for the desired RDM behaviors. Along the same line, the Council of Higher Education, which is the central body that governs the recruitment and promotion of academicians in Turkey, can include data citation into its promotion scheme to encourage data sharing practices. We expect that such and similar incentives and encouragements would have a huge positive impact on research data handling and management.

As for the awareness and the training needs of Turkish researchers, library and information science practitioners and scholars seem to be the critical group. They have the basic concepts of RDM but lack experience with real-life RDM and the domain expertise the research data is coming from. A train-the-trainer for them would be the first step to address their lack of knowledge and experience. The second step would be embedding library and information science professional to research projects where they can provide support to active researchers so that their RDM needs are addressed real time. Simultaneously, RDM or DMP trainings can be provided as a general service course within the curriculum such as statistics courses to fourth year students or first year graduate students, students who are more likely to deal with research data. Another option would be seminars or workshops can be conducted that are targeted for active researchers; they can be discipline-specific as different disciplines have different data habits. Qualitative research and participatory workshops (information

professionals and domain scientists) are needed to design such seminars and workshops. TÜBİTAK can support such activities as they did with Aperta.

There is potential in Turkey: (i) single funding agency (TÜBİTAK has influence over scientists); (ii) decent research (Turkish affiliated researchers produce more than 30000 Web of Science indexed articles annually); (iii) awareness; and goodwill (from the current study). A well-thought organization of these resources would not only benefit Turkish researchers by supporting RDM activities but also can be an example for other countries.

As for the future, two studies are planned. Firstly, a survey of DergiPark journals' awareness of and attitudes towards RDM is being designed. The DergiPark journals can support healthy RDM habits. Secondly, a detailed analysis of the training needs for scholars from different disciplines is needed as different disciplines have different approaches to data.

Acknowledgement

We would like to thank people working for The Scientific and Technological Research Council of Turkey (TUBITAK); the Turkish Academic Network and Research Center (ULAKBIM) for their support; Filiz Mengüç and Ebru Aydın for the processes of designing the survey; Murat Köreke for the distribution of survey via LimeSurvey; and M. Mirat Satoğlu, director of ULAKBIM, for his support on matters relating to administrative processes.

References

Allard, S., Aydinoglu, A. U. (2012) Environmental researchers' data practices: An exploratory study in Turkey. In *E-Science and Information Management* pp.13-24. Berlin Heidelberg: Springer.

Aperta TÜBİTAK (2019). Aperta - TÜBİTAK Açık Arşivi [Aperta – Open archive of TUBITAK]. Available at: https://aperta.ulakbim.gov.tr/about (accessed 2 February 2020).

Araştırma Verileri Yönetimi Eğitim Portalı [Training Portal for Research Data Management] (2019) Available at: https://acikveri.ulakbim.gov.tr/acik-veri-acik-bilim/ (accessed 2 February 2020).

ARBİS. (2019). Available at: https://arbis.tubitak.gov.tr/home.jsf (accessed 2 February 2020).

Aydinoglu, A.U., Dogan, G., Taskin, Z. (2017) Research data management in Turkey: Practices and attitudes. Library Hi Tech 35(2): pp. 271-289.

Aydinoglu, A, Suomela T, Malone J (2014) Data management in astrobiology: challenges and opportunities for an interdisciplinary community. *Astrobiology*,14(6), 451-61. Doi: 10.1089/ast.2013.1127

Bertzky, B., Stoll-Kleemann, S. (2009). Multi-level discrepancies with sharing data on protected areas: What we have and what we need for the global village. Journal of Environmental Management, 90: pp. 8-24

Bishop, B., Gunderman, H., Davis, R., Lee, T., Howard, R., Samors, R., Murphy, F., Ungvari, J. (2020) Data curation profiling to assess data management training needs and practices to inform a toolkit. Data Science Journal 19(1): p. 4. DOI: 10.5334/dsj-2020-004

Cochran, W.G. (1963). *Sampling techniques*. New York: John Wiley and Sons, Inc.

Cragin, M.H., Palmer, C.L., Carlson, J.R., Witt, M. (2010) Data sharing, small science and institutional repositories. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. DOI: 10.1098/rsta.2010.0165

Elsayed, A. M., Saleh, E. I. (2018) Research data management and sharing among researchers in Arab universities: An exploratory study. IFLA Journal 44(4): pp. 281–299. DOI: 10.1177/0340035218785196

European Commission (2016) Open access & data management. Available at: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm (accessed 2 February 2020).

Fecher, B., Friesike, S.,& Hebing, M. (2015) What drives academic data sharing?. PLOS ONE 10(2): e0118053. DOI: 10.1371/journal.pone.0118053

Ferguson, A.R., Nielson, J.L., Cragin, M.H., Bandrowski, A.E., Martone, M.E. (2014) Big data from small data: data-sharing in the 'long tail' of neuroscience. Nature Neuroscience 17: pp. 1442–1447. DOI: 10.1038/nn.3838

Grootveld, M., Leenarts, E., Jones, S., Hermans, E., Fankhauser, E. (2018) OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans (Version 1.0.0). Available at: https://zenodo.org/record/1120245#.XJJuEqRRWM8 (accessed 2 February 2020).

Hardisty, A., Roberts, D., Addink, W., Aelterman, B., Agosti, D., et al. (2013) A decadal view of biodiversity informatics: Challenges and priorities. BMC Ecology 13: p. 16. DOI: 10.1186/1472-6785-13-16

Herold, P. (2015). Data sharing among ecology, evolution, and natural resources scientists: An analysis of selected publications. Journal of Librarianship & Scholarly Communication 3(2): pp. 1-23. DOI: 10.7710/2162-3309.1244

Higman, R., Pinfield, S. (2015) Research data management and openness: The role of data sharing in developing institutional policies and practices. Program: Electronic library and information systems 49(4): pp. 364-381. DOI: 10.1108/PROG-01-2015-0005

Ioannidis, J.P.A. et al. (2009) Repeatability of published microarray gene expression analyses. Nature Genetics 41: pp. 149-155. DOI: 10.1038/ng.295

Ioannidis, J.P.A., Munafo, M.R., Fusar-Poli, P., Nosek, B.A., David, S.P. (2014) Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends in Cognitive Sciences 18: pp. 235-241.

John, L. K., Loewenstein, G., Prelec, D. (2012) Measuring the prevalence of questionable research practices with incentives for truth telling. Psychological Science 23(5): pp. 524-532. DOI:10.1177/0956797611430953

Koltay, T. (2019) Accepted and emerging roles of academic libraries in supporting research 2.0. The Journal of Academic Librarianship 45(2): pp. 75-80. DOI: 10.1016/j.acalib.2019.01.001

Koole, S.L., Lakens, D. (2012) Rewarding replications: a sure and simple way to improve psychological science. Perspectives in Psychological Sciences 7: pp. 608-614. DOI: 10.1177/1745691612462586

Kratz, J.E., Strasser, C. (2015). Researcher perspectives on publication and peer review of data. PLOS ONE 10(2). DOI: 10.1371/journal.pone.0117619

Leaders Activating Research Networks (LEARN) (2017). LEARN Toolkit of Best Practice for Research Data Management. Available at: https://discovery.ucl.ac.uk/id/eprint/1546518/ (accessed 2 February 2020).

MacMillan, D. (2014) Data sharing and discovery: What librarians need to know. Journal of Academic Librarianship 40(5): pp. 541-549.

McAfee A, Brynjolfsson, E. (2012) Big data: The management revolution. Harvard Business Review 90: 60-68. Available at: https://hbr.org/2012/10/big-data-the-management-revolution (accessed 2 February 2020).

Mennes, M., Biswal, B.B., Castellanos, F.X., Milham, M.P. (2013) Making data sharing work: the FCP/INDI experience. NeuroImage 82: pp. 683–691. DOI: 10.1016/j.neuroimage.2012.10.064

Michener, W.K. (2015) Ecological data sharing. Ecological Informatics 29: pp. 33-44. DOI:10.1016/j.ecoinf.2015.06.010

NASA (2011) Data & information policy. Washington, D.C.: NASA.

NIH (2008) Revised policy on enhancing public access to archived publications resulting from NIH-funded research. National Institutes of Health (NIH), Bethesda, MD.

NSF (2010) Scientists seeking NSF funding will soon be required to submit data management plans. Press Release 10-777, National Science Foundation (NSF), Arlington, VA.

Nosek, B.A., Spies, J.R., Motyl, M. (2012) Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. Perspectives of Psychological Science 7: pp. 615-631. DOI: 10.1177/1745691612459058

Pinfield, S., Cox, A.M., Smith, J. (2014) Research data management and libraries: Relationships, activities, drivers and influences. PLOS ONE 9(12): e114734. DOI: 10.1371/journal.pone.0114734

Piwowar, H.A., Vision, T.J.(2013) Data reuse and the open data citation advantage. PeerJ 1: e175,7872013. DOI:10.7717/peerj.175

Piwowar, H.A., Day, R.S., Fridsma, D.B. (2007) Sharing detailed research data is associated with increased citation rate. PLOS ONE. DOI: 10.1371/journal.pone.0000308

Piwowar, H. (2013) Value all research products. Nature 493(159). DOI: 10.1038/493159a

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J.-B., Vul, E., Yarkoni, T. (2017) Scanning the horizon: Towards transparent and reproducible neuroimaging research. Nature Reviews Neuroscience 18: pp. 115-126. DOI: 10.1038/nrn.2016.167

Pollock, D. (2016) Understanding scientific data sharing outside of the academy. Poster presented at the 79th Association for Information Science & Technology (ASIS&T) Annual Meeting, Copenhagen, Denmark.

Roettger, T.B., Winter, B., Baayen, H. (2019) Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility. Journal of Phonetics 73, pp. 1-7. DOI: 10.1016/j.wocn.2018.12.001

Royal Society (2012) Science as an open enterprise: Final report. Available at: https://royalsociety.org/policy/projects/science-public-enterprise/Report/ (accessed 2 February 2020)

Sesartic, A., Dieudé, A. (2017) Research data management training and support services at both ETH Zurich and EPF Lausanne. DOI: 10.3929/ethz-b-000224456

Spallek, H., Weinberg, S. M., Manz, M., Nanayakkara, S., Zhou, X.,Johnson, L. (2019) Perceptions and attitudes toward data sharing among dental researchers. JDR Clinical & Translational Research 4(1): pp. 68-75. DOI:10.1177/2380084418790451

Spires-Jones, T.L., Poirazi, P., Grubb, M.S. (2016) Opening up: Open access publishing, data sharing, and how they can influence your neuroscience career. European Journal of Neuroscience 43(11): pp. 1413-1419. DOI: 10.1111/ejn.13234

Steeleworthy, M. (2014) Research data management and the Canadian academic library: an organizational consideration of data management and data stewardship. Partnership: The Canadian Journal of Library and Information Practice and Research 9(1). DOI: 10.21083/partnership.v9i1.2990

Strasser, C.A., Hampton, S.E. (2012) The fractured lab notebook: undergraduates and ecological data management training in the United States. Ecosphere 3(12): art116. DOI: 10.1890/ES12-00139.1

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M. (2011) Data sharing by scientists: practices and perceptions. PLOS ONE 6(6). DOI: 10.1371/journal.pone.0021101

Tenopir, C., Christian, L., Allard, S., Borycz, J. (2018) Research data sharing: Practices and attitudes of geophysicists. Earth and Space Science 5: pp. 891–902. DOI: 10.1029/2018EA000461

Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., Dorsett, K. (2015a) Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PLOS ONE 10(8). DOI: 10.1371/journal.pone.0134826

Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., Sandusky, R., Langseth, M., Lundeen, A. (2015b) Research data services in academic libraries: Data intensive roles for the future? Journal of eScience Librarianship 4(2): e1085. DOI: 10.7191/jeslib.2015.1085.

Tenopir, C., Pollock, D., Allard, S., Hughes, D. (2016) Research data services in European and North American libraries: Current offerings and plans for the future. Proceedings of Association for Information Science and Technology 53: pp. 1-6. DOI: 10.1002/pra2.2016.14505301129

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Allard, S. (2017) Research data services in European academic research libraries. Liber Quarterly 27: pp. 23-44. DOI: 10.18352/lq.10180/

Tonta, Y. (2012) Açık erişim, kurumsal arşivler ve MedOANet Projesi [Open access, institutional repositories and MedOANet Project], paper presented at Ulusal Açık Erişim Çalıştayı [National Open Access Workshop], Ankara, Turkey, 8-9 November 2012. Available at: http://ae2012.acikerisim.org/wp-content/uploads/sites/6/2016/11/yasar_tonta.pdf (accessed 2 February 2020).

Tonta, Y. (2013) Açık erişimin geleceği ve araştırma verilerine açık erişim [The future of open access and open access for research data]", paper presented at Bilkent'te Kütüphanecilik Seminerleri [Librarianship Seminars at Bilkent], Ankara, Turkey. Available at: http://yunus.hacettepe.edu.tr/~tonta/yayinlar/tonta-acik-erisim-bilkent2013.pdf (accessed 2 February 2020)

TÜBİTAK. (2019) Tubitak Open Science Policy. Available at: https://acikveri.ulakbim.gov.tr/wp-content/uploads/2019/03/TUBITAK-Open-Sciency-Policy.pdf (accessed 2 February 2020)

Unal, Y., Kurbanoglu, S. (2018) Araştırma verilerinin yönetimi: Türk araştırmacılar verilerini açmaya hazır mı? [Research data management: Are Turkish researchers prepared to open their data?]. Türk Kütüphaneciliği [Turkish Librarianship 32(4): pp. 287-311. DOI: 10.24146/tkd.2018.42

Unal, Y., Chowdhury, G., Kurbanoğlu, S., Boustany, J., Walton, G. (2019). Research data management and data sharing behaviour of university researchers'. In Proceedings of ISIC, The Information Behaviour Conference, Krakow, Poland, 9-11 October: Part 2. Information Research, 24(1), paper isic1818. Available at: http://InformationR.net/ir/24-1/isic2018/isic1818.html (accessed 2 February 2020)

University of Pittsburgh (2020) Research data management @ Pitt. Available at: https://pitt.libguides.com/managedata (accessed 2 February 2020).

Verbakel, E, Noordegraaf, M, de Smaele, M. (2013) Data-intelligence training for library staff. In: Framing the digital curation curriculum, International conference, Florence, Italy, 6–7 May 2013, CEUR Workshop Proceedings, paper 13.

Wallis, J.C., Rolando, E., Borgman, C.L. (2013) If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLOS ONE 8(7): e67332. DOI: 10.1371/journal.pone.0067332

Whitmire, A.L., Boock, M., Sutton, S.C. (2015) Variability in academic research data management practices: Implications for data services development from a faculty survey. Program 49(4): pp. 382-407. DOI: 10.1108/PROG-02-2015-0017

Whyte, A., Tedds, J. (2011) Making the Case for Research Data Management. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Available at: http://www.dcc.ac.uk/resources/briefing-papers (acceded 2 February 2020).

Wittenber, J., Sackmann, A., Jaffe, R. (2018) Situating expertise in practice: Domain-based data management training for liaison librarians. The Journal of Academic Librarianship 44(3): pp. 323-329. DOI: 10.1016/j.acalib.2018.04.004

Xia, F., Wang, W., Bekele, T. M., Liu, H. (2017) Big scholarly data: A survey. IEEE Transactions on Big Data 3(1): pp. 18-35. DOI: 10.1109/TBDATA.2016.2641460

Yegros-Yegros, A., van Leeuwen, T.N. (2019) Production and uptake of open access publications involving the private sector: The case of big pharma. In *Proceedings of the 23rd International Conference on Science and Technology Indicators.* DOI: 10.31235/osf.io/zt6kc.

Yükseköğretim Kurulu [Council of Higher Education, Turkey] (2019). YÖK açık erişim - açık bilim çalışmaları [Open access and open science works of The Council for Higher Education]. Available at: https://acikerisim.yok.gov.tr/acik-erisim (accessed 2 February 2020).

Appendix: Questionnaire

1. Department: ……..
2. Title:
   a. Professor
   b. Associate professor
   c. Assistant professor
   d. Lecturer, Ph.D.
   e. Research assistant, Ph.D.
   f. Research assistant (as a graduate student)
   g. Lecturer (without Ph.D.)
   h. Other
3. Do you use research data?
   Yes / No
4. Have you cited any research data before?
   Yes / No / I didn't know that I could cite research data
5. Do you produce research data?
   Yes / No
6. Choose the 3 data types that you mostly used/produced.
   a. Experimental data
   b. Raw data
   c. Text data
   d. Survey
   e. Graphs
   f. Data models
   g. Lab notebooks
   h. Audio
   i. Video
   j. Remote sensing
   k. Other (Please specify)
7. Choose the 3 data formats that you mostly used/produced.
   csv / xls / txt / spreadsheet / sav / free text / readme structured / readme unstructured / other (please specify)
8. What is the average size of the data you used/produced in your latest study? (Please specify as MB, GB, etc.)
   ……
9. Where do you store the data for your latest study?
   Local computer / Cloud / Open access archives / Institutional repository / Commercial databsases (For ex, Data Citation Index) / Other (please specify):
10. Did you conduct a project before?
    Yes / No
11. Did you conduct a TÜBİTAK funded project before?
    Yes / No
12. What is the number of your TÜBİTAK funded projects?
    ……..
13. Type of your TÜBİTAK funded projects?
    ……

14. What is the number of publications produced from your TÜBİTAK funded projects?
    Article: ……
    Proceedings: ……
    Patent: ……
    Other: ……
15. Was it required to prepare a data management plan in any of your previous projects?
    Yes / No
16. What do you think about open data and data management plan practices being mandatory for TÜBİTAK projects that are already required for Horizon 2020 projects?
    Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree
17. What do you think about making the data produced by public resources/funds available to the public as open data??
    Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree
18. Would you like to participate/use if TÜBİTAK makes data management awareness studies (training, portal, open access resources etc.)?
    Yes / No
19. What are your suggestions for controlling the research data produced during the projects by the funding institution TÜBİTAK?
    ……..