Research Data Management in Turkey: Perceptions and Practices

Introduction

With the penetration and suffusion of information and communication technology (ICT) in our lives, scientific research has evolved as well. As such, scientific research is more data intensive and derives information from massive volumes of digitized data. As of 2013, 2.5 quintillion bytes of data are being produced every day (https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html), 90% of which was produced in the last two years (SINTEF, 2013). A correct assumption is that the amount of data being produced will continue to increase. For instance, Internet users numbered 2.8 billion in 2013, whereas today, they number more than 3.5 billion (http://www.internetlivestats.com/internet-users/). The use of social media has increased the amount of data being produced. The total amount of data in the world is expected to be 4.1 zetabytes in 2016 and is estimated to be 40 zetabytes in 2020. Therefore, data management has become an important issue.

Likewise, in the scientific arena, data has become so prominent that it has been given a new name in "The Fourth Paradigm: Data-Intensive Scientific Discovery" in which "all of the science literature is online, all of the science data is online, and they interoperate with each other" (Hey et al., 2009). In previous paradigms scientific activities were driven by experimentation, theory, and computation (Hey et al., 2009). The traditional hypothesis-based scientific approach has been gradually replaced by the analyses of electronic databases that can hold large amounts of information. As papers, lab books, tapes, and photographic films have moved to digital archives, cloud storages, and data warehouses, science has gone beyond the boundaries of hypotheses. Analyses are built on the collections themselves, and patterns, anomalies, and diversities on which questions will be posed later are sought. Hence, the term "data-intensive science" has emerged, and this practice derives information from the datasets collected by various computerized modeling and simulation systems, imaging devices, sensors and sensor networks, and other data gathering and storage techniques (Hey et al., 2009; Knyazkov et al., 2012). The vision is to have "all of the science literature online, all of the science data online, and interoperate with each other" (Hey et al., 2009).

These mega-scale databases consist of data captured by various novel scientific tools, sometimes on a real-time basis. With this continuous flow of electronic information, the need to collect, store, curate, integrate, and analyze data in a way that could help inter-institutional and interdisciplinary collaboration has gained importance for the advancement of science in the twenty-first century.

According to Birnholtz and Bietz's study (2003, p. 339), data is an evidence for validation of scientific contribution and it makes a social contribution to the establishment of practice. Therefore, understanding the importance of the data is vital to design, sustain and curate well-structured research data management systems. In the light of all these developments and rising importance of "research data management" subject, this paper aims to reveal the perceptions and practices of Turkish researchers on the subject of RDM. In a nutshell, the current study addresses the question of the perceptions toward and practices of RDM in Turkey. Main research questions are as follows;

- What are the common research data types and formats among Turkish scholars?
- To whom and what degree research data is shared in Turkey?
- What are the main reasons for not sharing research data with others?
- What are the most preferred places to store the data?
- What is the awareness level of scholars about the benefits of data sharing?
- What are the current conditions and facilities provided by universities or research institutions for RDM?

According to research questions, current condition for RDM in Turkey evaluated from two angles; skills/awareness levels of scholars and current policies on research data. As first five research questions aim to reveal skills and awareness levels of scholars about data management, the last question is designed to understand the approaches of decision makers and managers. Answers of the research questions are grouped in the discussion section to provide general framework on research data approaches in Turkey.

Literature Review

Various techniques and tools are required to analyze datasets. High-performance computers and advanced software help scientists to process large arrays of datasets to produce results that could be later reused, tested, and verified. High-quality datasets, if stored in a way which facilitate the instantaneous global access, could be used anywhere, anytime, thereby resulting in new scientific theories and studies.

The literature on research data management (RDM) is growing rapidly. Current studies focus on understanding the current situation, storing research data, the role of libraries and data warehouses in the process, opinions toward RDM, and so on (Faniel & Jacobsen, 2010; Tenopir et al, 2011; Corrall et al., 2013; Faniel et al., 2013; Calvert, 2015; Lee, 2015; Surkis and Read, 2015; Steiner, 2015; Cox et al., 2016; AL-Omar and Cox, 2016).

That the full potential of this new era is being utilized is difficult to argue. What we have now, both technologically and policy-wise, can provide only inefficient and unsatisfactory results compared with what we need, and as a result, the progression of science is slowed by the absence or insufficiency of regulatory measures for RDM (NSF, 2007; Chen and Zhang, 2014). Today, much like in the past, the majority of research data collected for a specific purpose are not archived digitally in a way that allows inter-institutional knowledge transfer, and the possibility of accessing such datasets after the relevant research paper is published declines by 17 percent per year (Wallis et al., 2013; Borgman et al., 2016; Vines, 2014). Considering the amount of lost data that could be used for developing new theories, training scientists to investigate diversified datasets collected by various instruments and techniques, and reproducing reported results to verify fabrication and falsification or to compare with past or future results, funding agencies have been establishing RDM and sharing mandates, which encourage research bodies to plan and implement data storage, curation, and analysis services (Hey et al., 2009; Douglass et al., 2014).

Despite the obvious shift toward the fourth paradigm (Hey et al., 2009), data-intensive science has its limitations because of data management issues. An important part of the topic is the behavioral aspect of RDM by scientists. Attitudes toward data sharing and preservation, data behaviors, and institutional support given to scientists are critical in establishing RDM systems (Tenopir et al., 2011; Piwowar & Vision, 2013; Tenopir et al., 2015a; Aydinoglu et al., 2014). Scientists collect, generate, and gather large amounts of data during the course of a study, and most of the time, they end up not knowing what to do with it after the results have been published. Personal digital archives lack the guarantee of permanency, and the storage quality may differ. Furthermore, when personally stored, dataset may also be sifted so that the information relevant to the hypotheses remains, and the rest of the information that may be significant to other studies is eliminated. In addition, personal data storage does not allow sharing most of the time; thus, the information that may be omnipresently required for verification or training issues remains inaccessible. Moreover, other stakeholders such as libraries and data managers play an important role in the data life cycle (Douglass et al., 2014; Tenopir et al., 2015b).

RDM schemes are developed to overcome such barriers and to guide scientists on how to handle their data. To provide reliability, quality, and availability, such schemes work together with ICT solutions and policy mandates to unify efficient scientific production. The rationale here is that imposing a common data management scheme is imposed by funding agencies and research institutions, the verification, reuse, and expansion of datasets will be ensured, thereby resulting in sustainability and efficiency in scientific production and advancement. It is too early to tell whether this rationale is going to work or not; however, funding agencies and research institutions have been quick to take action and have added RDM schemes to their grant agreements for the past few years. The schemes that have been implemented in the highest number of studies could potentially be listed as those planned for EU funds, those developed and/or adopted by major U.S. research agencies, and those developed by the Organisation for Economic Co-operation and Development (OECD) for access to research data from public funding. The European Commission has been piloting an open access program since 2008, during which the beneficiaries were encouraged to self-archive (green publishing) or to publish their work in open access mode (gold publishing) so that data are deposited in a repository to be accessed and reused by third parties later (Horizon 2020, 2013). In the U.S., each funding agency has its own separate policy. For instance, the National Science Foundation requires project administrators to prepare a data management plan with their proposals

(NSF, 2010); the National Institute of Health mandates data sharing with safeguards to ensure privacy and confidentiality of health data, and encourages an open access culture through PubMed (NIH, 2003); and the National Aeronautics and Space Administration has been investing in data management for years through different data repositories, such as those for earth science, planetary missions, and astronomical observations (NASA, 2016). In addition, recognizing the need for an international initiative, 30 OECD countries and Russia, China, South Africa, and Israel have signed the Declaration on Access to Research Data for Public Funding in 2004 and created guidelines (OECD, 2007).

In Turkey, few studies focus on RDM, and efforts are being made to increase awareness on the issue. Open access is a relatively important topic, and the same scholars are interested in both topics. The MedOANet project in Turkey conducted a nationwide survey and found that RDM is not mentioned in open access policy papers (Tonta, 2012; Tonta, 2013). The first paper was a conference proceeding on the challenges of research data practices for environmental scientists (Allard and Aydinoglu, 2012). Hacettepe University organized an international workshop in November 2014 on RDM, in which best practices on RDM were shared with the participants and discussions were held for future actions in Turkey (http://rdm.bilgiyonetimi.net/index.html). A detailed assessment of the workshop is published for Turkish audiences (Tonta and Al, 2012). The same year, the theme for the 5th International Symposium on Information Management in a Changing World was RDM; papers were presented and a half-day workshop was held during the symposium (IMCW2014, 2014). A limited number of scholars have published on the issue (Onder, 2013; Gurdal and Bitri, 2015; Malkoc, 2015). However, activities geared toward increasing awareness have not succeeded. Despite the OECD paper, not even a single agency has an RDM policy (Tonta, 2013). Our study sheds light on the attitudes of Turkish scientists toward RDM.

Methods

Survey instrument

The survey instrument is a derivation of the seminal study of Tenopir et al. (2011). This version is used to gain a better understanding of the perceptions toward and practices of scientific data management in the astrobiology community (Aydinoglu et al., 2014). The survey is a shorter version of the Tenopir et al. survey but has new questions on data storage and backup. That version is translated into Turkish by the co-authors of this study. In addition, some parts of the survey are adjusted to the Turkish academic context, such as academic roles. Finally, relevant questions to the astrobiology community are broadened, such as questions on data repositories and data formats, as this survey is distributed to academics from all domains instead of a single domain. Despite the edits, the goal is to keep questions similar to the original survey to facilitate potential comparisons between international and Turkish RDM behaviors.

The surveys asks about i) demographic information; ii) data management practices (types of data collected, data formats, metadata standards; and iii) data backup practices through a five-point Likert scale (*disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat,* and *agree strongly*) attitudes, perceptions, and practices with regard to research data sharing. The Appendix shows the full set of questions. The survey is uploaded to SurveyMonkey.com, and the link is distributed to the potential participants.

Participants

The survey instrument is distributed to academicians from the top 25 most scholarly productive universities in Turkey1. The universities are selected because they have the most business with research data as they publish frequently. To obtain the list of top 25 universities, the researchers employed the report entitled "*Türkiye Üniversiteleri'nin Bilimsel Yayın Performansı: 2004–2014/Scholarly Production Performance of Turkish Universities: 2004–2014*" (TUBITAK ULAKBIM, 2016), which was prepared based on data from Thomson Reuters InCites. The total number of publications is divided by the number of academic staff in

---

1 Turkey has 193 universities (http://www.yok.gov.tr/web/guest/universitelerimiz).

these universities to measure the publications per academic. Such data come from the Higher Education Council database. The top 25 most productive universities in Turkey are listed in Table I.

Table I. Top 25 universities in Turkey based on the number of publications per academics, number of e-mails sent, number of responses, and response rate of these 25 universities

University	# of publications per person	E-mails sent (N)	E-mails responded (n)	Response rate % (n/N)
Hacettepe University	3.46	2096	74	4
Ankara University	3.01	1955	31	2
Ege University	3.36	1513	67	4
Middle East Technical University	3.84	1078	50	5
Erciyes University	2.94	1247	29	2
Ataturk University	2.88	1742	33	2
Istanbul University	2.51	985	28	3
Cukurova University	3.26	944	23	2
Gaziosmanpasa University	2.68	764	13	2
Gazi University	2.64	627	12	2
Gaziantep University	3.28	503	12	2
Bilkent University	4.64	295	4	1
Istanbul Technical University	3.54	505	13	3
Ondokuz Mayis University	3.42	621	23	4
Firat University	3.23	543	18	3
Gebze Technical University	4.24	426	16	4
Kirikkale University	2.49	352	8	2
Dicle University	2.87	226	10	4
Bogazici University	3.87	583	12	2
Kahramanmaras Sutcu Imam University	2.46	478	7	1
Yuzuncu Yil University	2.70	484	10	2
Harran University	2.70	309	10	3
Koc University	7.13	315	3	1
Fatih University	3.47	410	13	3
Baskent University	3.24	630	13	2
Total	-	19631	532	3

The e-mail addresses of the academics are collected from the university websites. A total of 19,631 academicians are contacted via e-mail and invited to participate in the survey. A total of 1,082 e-mail addresses bounced back for various reasons. A total of 532 academics from 25 universities participated in the survey. Eleven responses came from academics that are from different universities, and their responses are not included in the analysis. Thus, the response rate is approximately 3%.

According to Cochran's (1963) formula for a sample to represent the population, 377 participants can be used to represent 19,631 people with a 95% confidence interval for e = 0.05, and 582 participants indicate a 99% confidence interval for e = 0.05. Therefore, we are satisfied with the number of participants to our survey.

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

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

In the formulas,

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

IBM SPSS Statistics software package (v. 21) is used to analyze data. Descriptive statistics such as frequencies, cross-tabulations, descriptive ratio statistics, and chi-square tests are employed.

Findings

Among the 532 participants, the universities with the most participants are Hacettepe University (13.9%), Ege University (12.6%), and METU (9.4%). The others are Ataturk University (6.2%) and Koc University (.6%). The largest participant group according to domain is from humanities and social science (36.8%), followed by engineering (18.8%), health sciences (14.8%), agricultural and fisheries (11.7%), and sciences (11.3%). As for the academic titles of the participants, the number of graduate research assistants (38.9%) who participated in the survey was double that of any other group (assistant professors, 17.9%; associate professors, 18.6%; professors, 17.1%).

In addition to research responsibilities, academicians in Turkey are expected to teach and conduct administrative tasks. Therefore, knowing how much of their time is dedicated to research is important when analyzing the results. The participants are asked how much of their weekly 40 hours is distributed among research, teaching, administrative duties, and others (Figure 1). The responses indicate that the amount of time allocated to research and teaching is similar, and the time spent on administrative tasks is lower. For half of the participants, five hours or less are allocated to administrative tasks; in other words, less than one-eighth of their labor is consumed by non-research and non-teaching activities. Twenty-five percent of the respondents can spend a minimum of 10 hours/week on research, and 10% spend 29-40 hours/week on research. Overall, the respondents conduct research and deal with data; thus, they are the correct sample to ask about RDM.
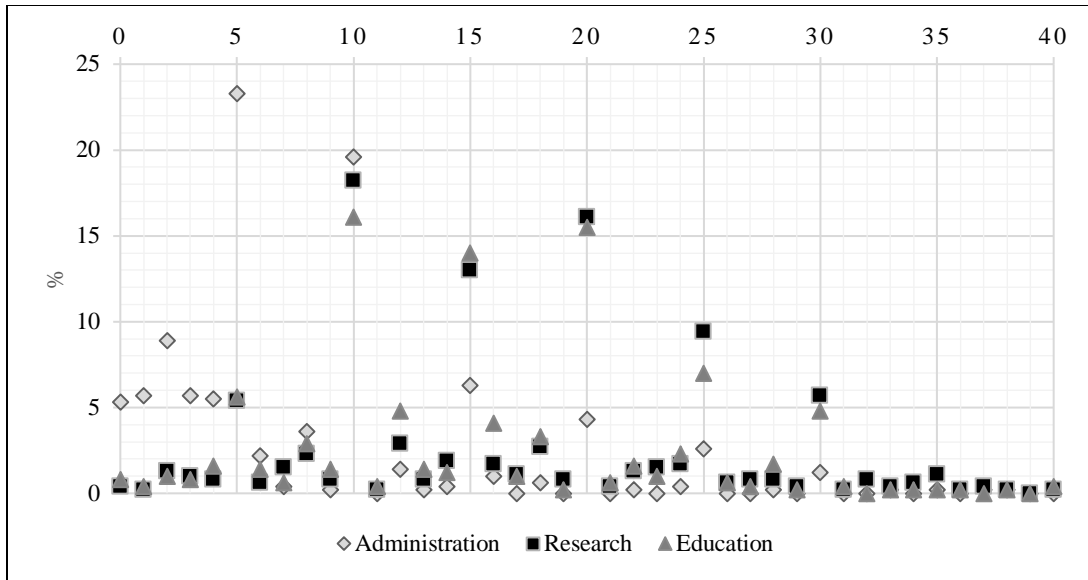
Figure 1. Distribution of 40 hr/week work on administrative duties, research, and education.

We also asked how much of the respondents' time is used for research, education, and administrative duties. On the basis of their academic titles, the width of the distribution for assistant professors and postdocs is significant. Professors and associate professors have a balanced distribution. Although the latter is not as great as the former, it is considerably better than the rest.
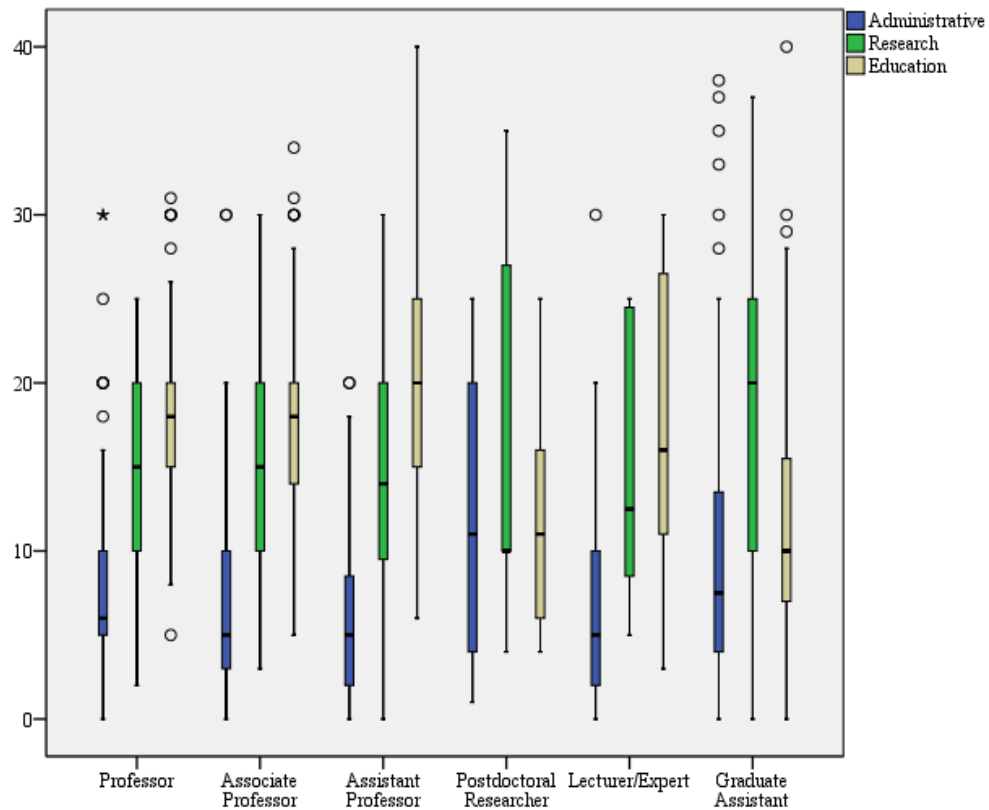

Figure 2. Distribution of 40h/week time spent on administrative duties, research, and education according to academic titles.

Table II provides the responses on data types. According to the responses, experimental data (52.8%) and text data (47.0%) are the two types of data that were used by half of the respondents. Survey data use is also significant (~41%). Approximately, a quarter of the respondents reported that they use other types of data: still image (pictures and photos) (26.1%) and model-algorithm-code data models (25.2%), lab notebook (22.7%), and audio recordings (22.2%). Only a small group (2.8%) mentioned that they do not use research data. The chi-square test provided statistically significant differences for the use of different data types based on academic ranking. The use of experimental data by professors (67.0%), associate professors (62.6%), and postdoctoral researchers (64.3%) is greater than any other academic rankings. The use of text data is greater among graduate assistants (55.1%) and assistant professors (49.5%). Postdoctoral researchers utilize survey data the most (64.3%). As a data type, data models are not employed as much as the rest; however, a statistically significant difference exists among their use according to academic titles: graduate assistants (31.9%) and lecturers/experts (34.6%). Audio data are popular among the graduate students as well (29.5%).

Table II. Research data types (frequencies and chi-square test results for academic title)

Data Type	n	%	χ^2	p
Experimental	278	52,3	22.749	0.000
Text	250	47,0	13.941	0.016
Survey	216	40,6	12.405	0.030
Still image	139	26,1	8.025	0.155
Data models	134	25,2	13.455	0.019
Lab notebook	121	22,7	3.425	0.635
Audio	118	22,2	23.583	0.000
Video	102	19,2	5.464	0.362
Remote sensing	28	5,3	-	-
Others	23	4,3	-	-
Not using research data	15	2,8	-	-

"-" indicates that because of a high number of "no" responses, chi-square test cannot be applied.

We also asked about the format they use to define their data. The most frequent response is spreadsheet, such as Excel and Google Spreadsheet (53.9%). One-third of the respondents indicated text, and 30.1% reported free text. A little over a quarter of the respondents (27.4%) uses SAV format. SAV and XML as data formats are favored more by postdoctoral researchers (57.1%, $\chi^2 = 18.923$, p = 0.002, and 28.6%, $\chi^2 = 14.683$; p = 0.012). DOC, which is not a data format, is the most reported format among the other data types. In addition, the most frequent formats are not "smart" or "networked."
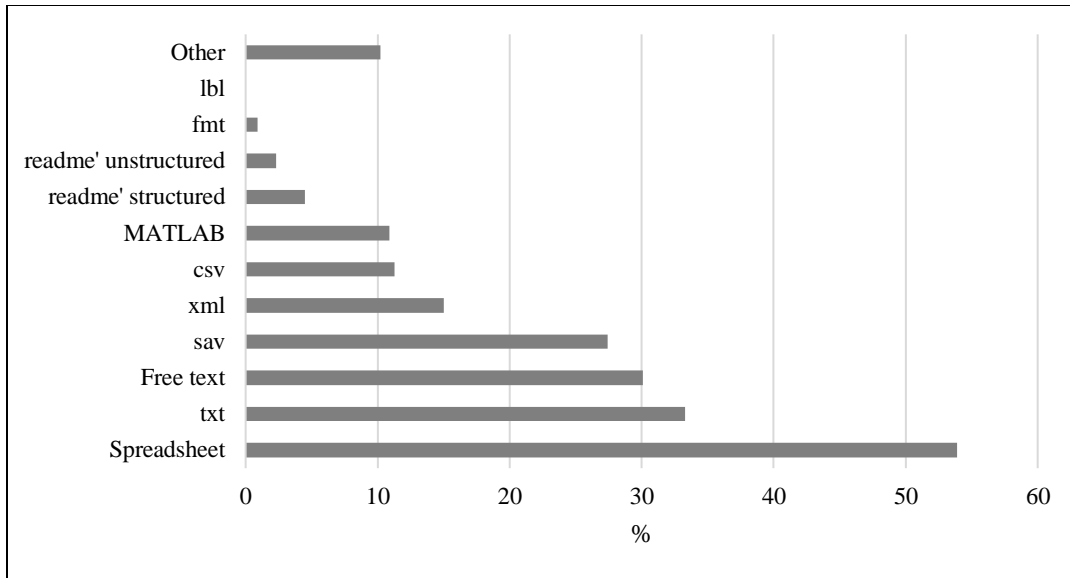
Figure 3. Research data formats.

A striking result is that 27.1% of the participants acknowledged that they do not know anything about metadata (who collected the data, when, where, why, etc.). Of the respondents (n = 484), only 176 (36.4%) reported that they record metadata, which is an extremely low figure. Academicians mostly use the metadata standard they developed in their lab (13.3%, n = 71). The second most frequent metadata standard is ISO (8.8%, n = 47). Each of the standards (AWM, DwC, DIF, EML, FGDC, CSDGM, NISO, MIX) account for less than 1%.

The participants are asked what they think of data sharing. One-eighth of them did not respond to this question. Of the responses, a little less than two-thirds (62.4%) reported that they do share, whereas 37.4% reported they do not. Among the 62.4%, when they are asked with whom they share their data and to what degree, almost all (98.9%) answered that they share their data with their research team, followed by scholars in their own discipline (76.6%), researchers in their organization (73.6%), and the scientific community (72.6%).
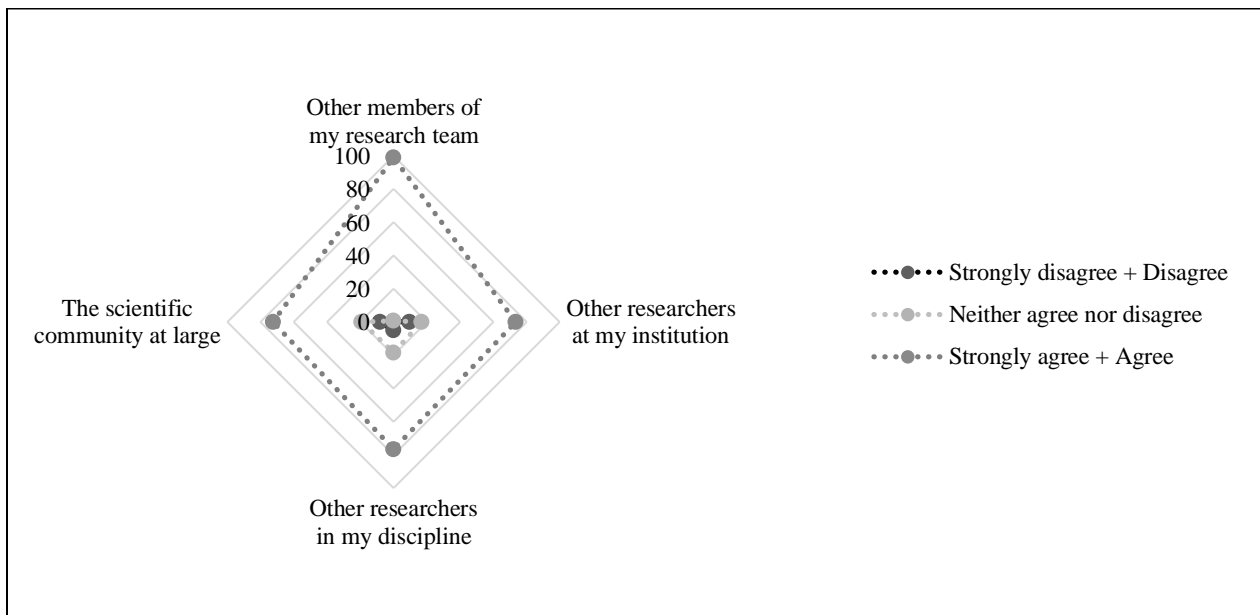

Figure 4. To whom and to what degree the research data is shared (%).

For the respondents who do not share their research data with others (37.6%), their reasons for not sharing data are provided below (Fig. 5). The most important reason is not wanting others to access their data (65.5%). Lack of technical skills and expertise to make them available, no place to store them, lack of funds, people do not need them, lack of metadata standards, lack of time, and lack of the funding agency's enforcement are other prominent reasons the participants do not share data with others.
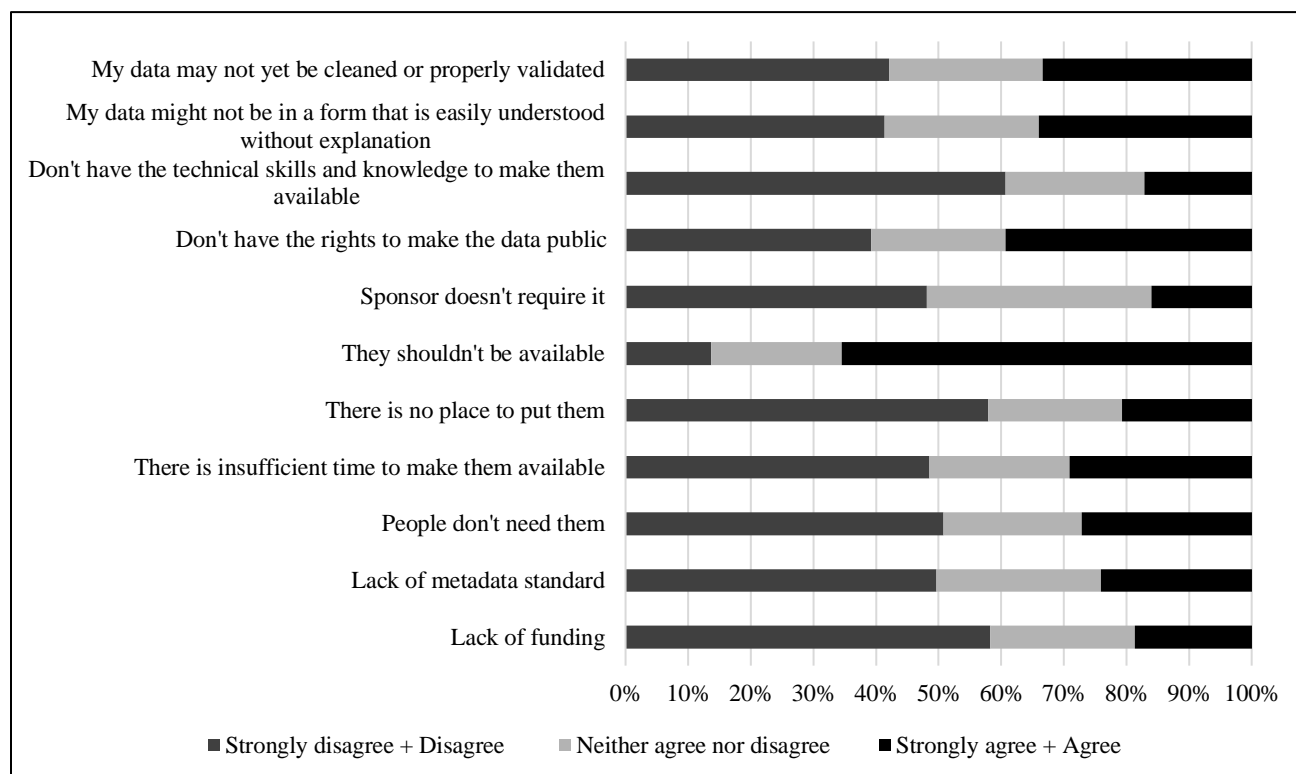
Figure 5. Reasons for not sharing data.

Different places are used to store data (see Table 3). Local computers (71.6%) are the most common storage place. Close to half of the participants also use the cloud (45.9%). The use of an open access data repository is quite low (8.3%). However, the data suggest that the increase in academic title results in the decrease in the use of the cloud for data storage ($\chi^2 = 32,978$; p = 0,000). In fact, graduate assistants use the cloud almost twice as much as the professors (58.9% and 30.8%, respectively).

Table III. Places to store data

Medium	n	%
Local computers	381	71.6
Cloud	244	45.9
Open access data repository	44	8.3
Institutional open repository	17	3.2
Commercial data repository	2	0.4

The participants are also asked what medium they prefer for data backup. Four out of 10 people use only discs (CD/DVD/external hard disk and thumb drive) (41.1%). Only one out of 10 people (10.4%) use the cloud. Close to half of the respondents utilize both discs and the cloud (47.0%), which shows that academics in Turkey do not fully trust in using only the cloud for storage. An important detail to acknowledge is that in addition to six participants who reported that they do not back up their data, 109 people did not answer this question. Thus, the percentages are calculated according to n = 423 (of the 532). Of the 423 academicians, 26.7% back up their data instantly, almost half of them (49.6%) back up once a week, and a quarter of them (25.8%) back up once a month.

The participants showed a positive attitude toward data sharing and acknowledge its benefits. A great majority of the participants (93.5%) think that "well-maintained data helps retain data integrity." Interestingly, fewer people (57.2%) agree with the statement that "data sharing reduces redundant data." Eighty-two percent of the participants think that data sharing encourages interdisciplinary collaborative science. Moreover, 84.2% agree that data management practices are beneficial "to the scientific process itself (re-analysis of data helps verify results data)," 78.4% think that data sharing helps "the training of the next generation of researchers," and 75.5% believe that data sharing "prevents data fabrication and falsification."
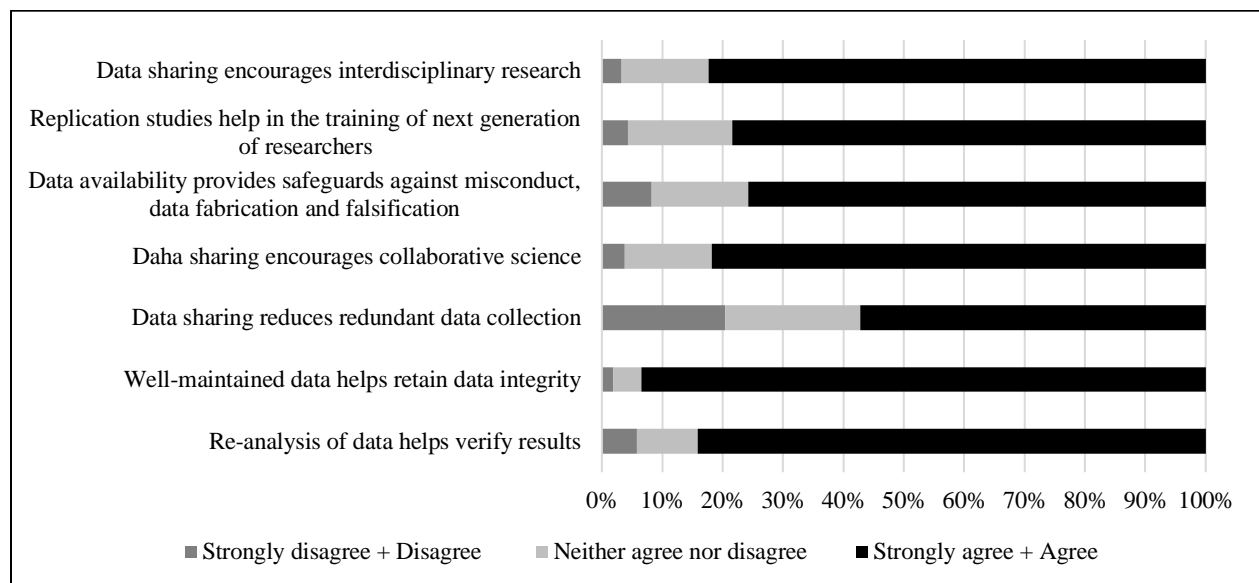

Figure 5. Benefits of data sharing.

Despite the individual positive attitudes toward data sharing, institutional support for RDM is nonexistent among the top 25 most productive universities in Turkey. Consequently, only 6.1% of the academicians reported that an RDM plan is mandatory in their institutions. Around 30% of the participants do not know whether an RDM policy is in effect in their organization. One-fifth of the institutions (22.1%) support RDM in technical issues only. Fifty-nine point nine percent of the participants reported that no RDM procedure exists, and 59.3% state that no policy with regard to RDM exists in their institutions. Only 13.3% of the institutions provide training on RDM, and 11.8% provide monetary support for RDM.
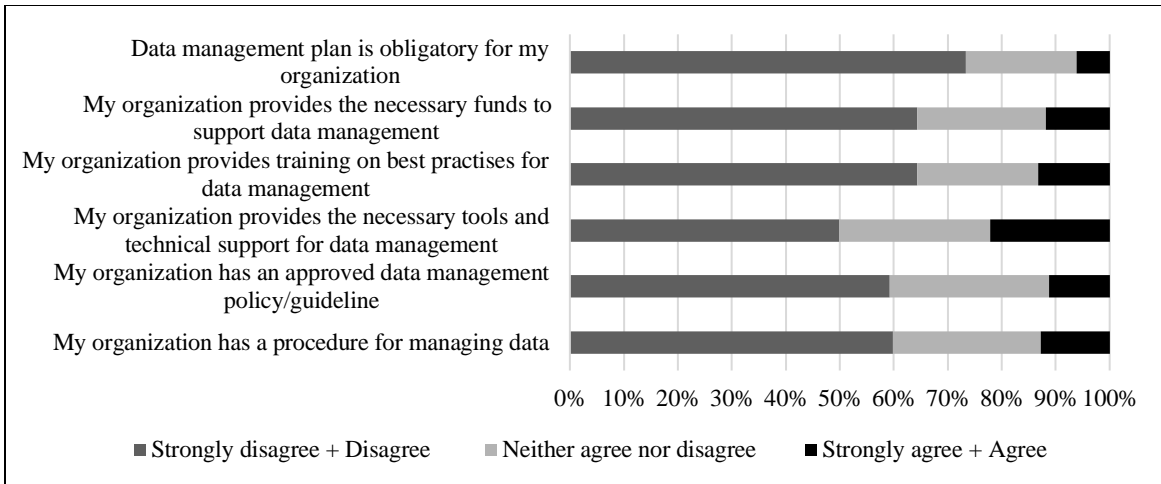
Figure 6. Institutional support for RDM.

Figure 6 shows that a great majority of the participants think that for them to share their data with others, having "formal citation of the data providers and/or funding agencies in all disseminated work making use of the data" (92.8%) is important. Other conditions that are important for sharing research data are as follows: "Formal acknowledgment of the data providers and/or funding agencies in all disseminated work making use of the data" (89%); "results based on the data could not be disseminated in any format without the data provider's approval" (84.3%); "mutual agreement on reciprocal sharing of data" (84.1%); and "the opportunity to collaborate on the project" (81.5%).
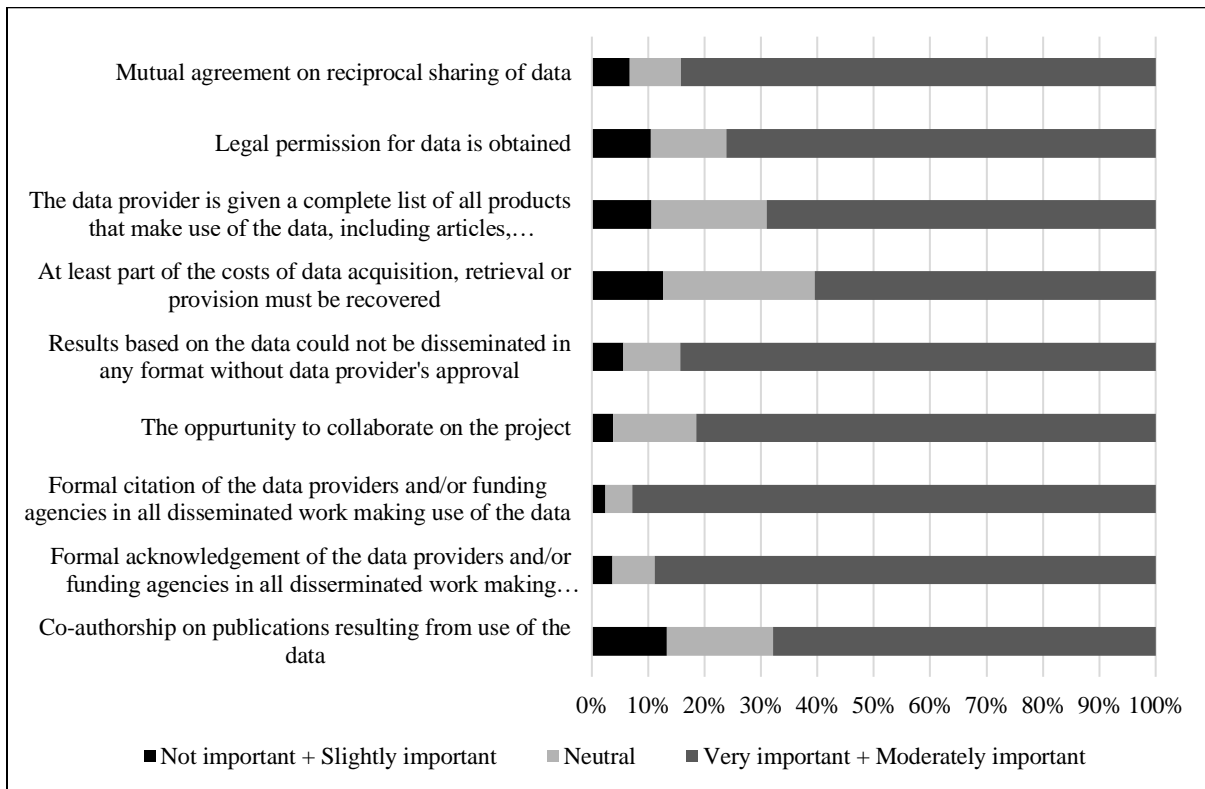

Figure 6. Conditions to sharing data with other researchers

Discussion

The amount of scientific data and information has increased so much that processing, analyzing, and storing have become arduous tasks. RDM seems to be the only way to perform such tasks because RDM ensures that data collection, processing, and curation can be performed effectively, as well as minimizes costs. However, the Turkish research community does not seem ready to adopt such a strategy as Allard & Aydinoglu (2012) found earlier for environmental scientist in Turkey. Now, we took a snapshot of the perceptions toward and practices of RDM by Turkish academics in the top 25 universities in Turkey. Our findings can be grouped into the following two areas:

Lack of research data policy or strategy. RDM does not exist from an institutional perspective. The main funding agency in Turkey (TUBITAK) neither has an RDM policy/strategy nor asks for an RDM plan from the scientists it funds. The universities do not have an established mechanism (policy, guidance, staff, software, hardware, training, etc.) to support their staff with regard to RDM activities. Incentives and sanctions do not exist. Even though research is becoming increasingly conducted through data, the benefits of RDM, the resources that RDM needs, and the vision for research data are not acknowledged by the people who govern science.

To address this problem, TUBITAK should prepare a research data strategy/policy document with input from all the stakeholders. Without a strategy, individual efforts would be unlikely to amount to something. Turkish research institutions and researchers have to adopt better RDM practices because international programs require RDMs. For instance, when the institutions receive funding from the H2020 Program, an RDM has to be submitted within six months. In addition, the academic activities with regard to RDM or open data can be added to the academic promotion system and/or other incentive systems run by TUBITAK or the Higher Education Council. As a result, not only will academicians take better care of their research data and share it with others, but also funding money can be used more effectively through reuse of research data.

Lack of skills and knowledge. Our results indicate that a great majority of academics in Turkey lack the technical skills and knowledge for effective RDM. Basic knowledge, such as collecting/curating data according to a metadata standard or formats to store data, is lacking. The .doc file name extension, a proprietary format by Windows for Word documents, is thought to be a data format, or one-third of the participants do not know what metadata is. The academics may lack technical knowledge and skills; nevertheless, they are aware of the benefits of data sharing, such as how data sharing facilitates interdisciplinary research and collaboration, as well as help verify results. They expressed that under certain conditions, they are willing to share, but for many reasons, they cannot. This finding is supported by a quick search on the Data Citation Index on September 30, 2016. Only 413 datasets were posted by 48 Turkish scholar groups. Compared with the number of publications per year (~30,000) (WoS, 2016), this number is abysmally small. Yet, investigating the motivations and practices of these 48 groups can be illuminating and help TUBITAK and the universities to craft RDM policies and practices, and spread best practices.

Trust is also an important factor for the RDM practices of Turkish academics. The closed network style of the Turkish academic system makes researchers more protective of their data. It also affects data preservation practices. Researchers use multiple mediums to ensure their data is safe.

Turkish researchers have similarities with researchers around the world in some areas and are not similar in other areas (via Tenopir et al., 2011 and Tenopir et al., 2015a). For instance, in both cases, institutional support is low, and the metadata standard that is developed in one's lab is the most common standard. However, Turkish academics seem to have less knowledge of metadata. Experimental data among the types of data used in research come first in both; however, other data types (observational, biotic surveys, etc.) are not used by Turkish researchers. The most contrasting finding is the reason for not sharing data. For Turkish scholars, "data shouldn't be available" is the first reason. By contrast, this reason is the last for the international community, whose primary reason is "lack of time."

To address the lack of skills and knowledge, early career scholars can be utilized. Our study reveals that graduate research assistants have the highest awareness of RDM. They are also the ones who use research data the most. In fact, a high academic ranking corresponds to low use of research data. This finding may not

be surprising because early career people are more tech savvy and open to learning, and they are often assigned tedious tasks such as data cleaning and curation (Powell, 2016; Tenopir et al., 2011). It is easier to adapt good data habits for them as they are still in training and through them a sustaining impact on the data culture can be achieved (Vogeli et al., 2006; Aydinoglu et al., 2014). Fostering collaboration among people of different academic ranks is important to benefit all parties particularly those in more data-intensive fields. Data science courses can be added to the curriculum in science departments. In addition, extracurricular seminars and workshops can be organized for graduate students and scientists who deal with research data.

In conclusion, although our study confirms some of the barriers to efficient RDM, more research is needed to uncover the specific barriers and how to bypass them. Identifying the training that researchers need at different levels is another crucial area. In our study, we looked at university researchers, but some government agencies generate data as well, such as the Ministry of Environment and General Directorate of Mineral Research and Exploration; these agencies need to be studied. Moreover, needs assessment for hardware, software, data repository, and technical knowledge is critical. Most importantly, a data strategy or policy for Turkey is needed. TUBITAK should lead a RDM strategy and policy in collaboration with other stakeholders (academia, government, NGOs). The open access community, which has been quite active in the last decade in Turkey, can support open data (and RDM) and TUBITAK in crafting the strategy/policy document.

References
Allard S., Aydınoğlu A.U. (2012), "Environmental researchers' data practices: An exploratory study in Turkey", In: Kurbanoğlu S., Al U., Erdoğan P.L., Tonta Y., Uçak N. (eds) *E-Science and Information Management*. IMCW 2012. *Communications in Computer and Information Science*, vol 317. Springer, Berlin, Heidelberg.

AL-Omar, M. and Cox, A.M. (2016), "Scholars' research-related personal information collections A study of education and health researchers in a Kuwaiti University", *Aslib Journal of Information Management*, vol 68, no 2, pp. 155-173.

Aydinoglu, A.U., Suomela, T. and Malone, J. (2014), "Data management in astrobiology: Challenges and opportunities for an interdisciplinary community", *Astrobiology*, vol 14, no 6, pp. 451-461.

Birnholtz, J.P. and Bietz, M.J. (2003), "Data at work: Supporting sharing in science and engineering", In GROUP'03 Proceedings of the 2003 International ACM SIGGROUP Conference, pp. 339-348. ACM, Florida,

Borgman, C.L., Golshan, M.S., Sands, A.E., Wallis, J.C., Cummings, R.L., Darch, P.T. and Randles B.M. (2016), "Data management in the long tail: Science, software, and service", *International Journal of Digital Curation*, vol 11, no 1, pp. 128-149.

Calvert, P. (2015), "Should all lab books be treated as vital records? An investigation into the use of lab books by research scientists", *Australian Academic and Research Libraries*, vol 46, no 4, pp. 289-303.

Chen, C.L.P. and Zhang, C.Y. (2014), "Data-intensive applications, challenges, techniques and Technologies: A survey on big data", *Information Sciences*, vol 275, pp. 314-347.

Cochran, W.G. (1963), *Sampling Techniques*, 2nd Ed.. New York: John Wiley and Sons, Inc.Corrall.

Corrall, S., Kennan, M.A. and Afzal, W. (2013), "Bibliometrics and research data management services: emerging trends in library support for research", _Library Trends,_ vol 61, no 3, pp. 636-674.

Cox, A.M., Pinfield, S. and Smith, J. (2016), "Moving a brick building: UK libraries coping with research data management as a 'wicked' problem", _Journal of Librarianship and Information Science_, vol 48, no 1, pp. 3-17.

Douglass, K., Allard, S., Tenopir, C., Wu, L. and Frame, M. (2014), "Managing scientific data as public assets: Data sharing practices and policies among full-time government employees", _Journal of the Association for Information Science & Technology_, vol 65, no 2,pp. 251-262.

Faniel, I.M., and Jacobsen, T.E. (2010), "Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data", _Computer Supported Cooperative Work_, vol 19 no 3-4, pp. 355-375.

Faniel, I., Kansa, E., Kansa, S.W., Barrera-Gomez, J. and Yakel, E. (2013). "The challenges of digging data: A study of context in archaeological data reuse." In _JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries_, pp. 295-304. New York, NY: ACM.

Gürdal, G. and Bitri, E. (2015), "Araştırma verisi yönetimi, açık veri ve Avrupa Birliği Bilimsel Veri Altyapısı: OpenAIRE2020 [Research data management, open data and the European Scholarly Communication Data Infrastructure: OpenAIRE2020]", paper presented at XVII. Akademik Bilişim Konferansı [XVII. Academic Computing Conference]. Eskisehir, Turkey, 4-6 February 2015, viewed on 5 May 2016, http://ab.org.tr/ab15/ozet/124.html

Hey, T., Tansley, S. and Tole, K. (2009), _The Fourth Paradigm: Data-intensive Scientific Discovery_, ebook, viewed from http://research.microsoft.com/enus/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf

Horizon 2020. (2013), Guidelines on data management in Horizon 2020: Version 1.0, viewed 2 February 2016, http://www.gsrt.gr/EOX/files/h2020-hi-oa-data-mgt_en.pdf

IBM. (2016), Bringing big data to the enterprise, viewed from https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

IMCW2014. (2014), viewed 4 May 2016, http://imcw2014.bilgiyonetimi.net/

Knyazkov, K.V., Kovalchuk, S.V., Tchurov, T.N., Maryin, S.V. and Boukhanovsky A.V. (2012), "CLAVIRE: e-Science infrastructure for data-driven computing", _Journal of Computational Science_, vol 3, no 6, pp. 504-510.

Lee, D.J. (2015), "Research data curation practices in institutional repositories and data identifiers". Unpublished PhD Dissertation, Florida State University, Tallahassee.

Malkoç, B. (2015), "Research data alliance ve DataCite", paper presented at 4. Ulusal Açık Erişim Çalıştayı [4th National Open Access Workshop], Ankara, Turkey, 19-21 October 2015, viewed on 11 May 2016, http://www.acikerisim.org/dokumanlar/ae2015_program.pdf

National Aeronautics and Space Agency (NASA). (2016), _Open Gov Plan 2016 Outline_, viewed 6 August 2016, https://open.nasa.gov/blog/open-gov-plan-2016-outline/

National Institute of Health (NIH). (2003), _Final NIH statement on sharing research data_, viewed 4 May 2016, http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

National Science Foundation (NSF). (2010), _Data management for NSF SBE Directorate proposals and awards_, viewed 4 May 2016, https://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf

National Science Foundation. (2007), _NSF 07-28, Cyberinfrastructure Vision for 21st Century Discovery_, viewed 6 July 2016, http://www.nsf.gov/pubs/2007/nsf0728/index.jsp

OECD. (2007), _OECD principles and guidelines for access to research data from public funding_, viewed 4 May 2016, http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm

Önder, A. (2013), "Büyük veri [Big data]", paper presented at 2. Ulusal Açık Erişim Çalıştayı [2nd National Open Access Workshop], Izmir, Turkey, 21-22 October 2013, viewed 6 May 2016, www.acikerisim.org

Piwowar, H.A. and Vision T.J. (2013), "Data reuse and the open data citation advantage", _PeerJ_, viewed 9 September 2016, https://peerj.com/articles/175/

Powell, K. (2016), "Young, talented and fed-up: scientists tell their stories", *Nature News*, viewed 16 February 2017, http://www.nature.com/news/young-talented-and-fed-up-scientists-tell-their-stories-1.20872

SINTEF. (2013), "Big data, for better or worse: 90% of world's data generated over last two years". *ScienceDaily*, viewed from www.sciencedaily.com/releases/2013/05/130522085217.htm

Steiner, K. (2015), "Research data management and information literacy - new developments at New Zealand University libraries", *Information-Wissenschaft und Praxis*, vol 66, no 4, pp. 230-236.

Surkis, A. and Read, K. (2015), "Research data management", *Journal of the Medical Library Association*, vol 103, no 3, pp. 154-156.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M. and Frame M. (2011), "Data sharing by scientists: practices and perceptions", *PLoS One*, viewed 5 June 2016, http://dx.doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D. and Dorsett, K. (2015a), "Changes in data sharing and data reuse practices and perceptions among scientists worldwide", *PLoS One*, viewed 8 May 2016, http://dx.doi.org/10.1371/journal/pone.0134826

Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, W.B., Baird, L., Sandusky, R., Langseth, M. and Lundeen A. (2015b), "Research Data Services in Academic Libraries: Data Intensive Roles for the Future?", *Journal of eScience Librarianship*, vol 4, no 2, 24 pages

Tonta, Y. and Al, U. (2012), "Araştırma verilerinin yönetimi [Research data management]", *Türk Kütüphaneciliği [Turkish Librarianship]*, vol 29, pp. 36-45.

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. Viewed 9 May 2016 http://www.acikerisim.org/sunumlar/yasar_tonta.pdf

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, 17 December 2013, viewed 8 May 2016, library.bilkent.edu.tr/activities/librarianship-seminars/presentations/yasar-tonta.pptx

TUBITAK ULAKBIM. (2016), "Türkiye üniversitelerinin bilimsel yayın performansı: 2004-2014 [Scholarly production performance of Turkish universities: 2004-2014]", viewed on 3 March 2016, http://ulakbim.tubitak.gov.tr/tr/hizmetlerimiz/turkiye-universitelerinin-bilimsel-yayin-performansi-2004-2014

Vines, T.H. (2014), "The availability of research data declines rapidly with article age", *Current Biology*, vol 24, no 1, pp. 94-97.

Vogeli, C., Yucel, R., Bendavid, E., Jones, L.M., Anderson, M.S., Louis, K.S., and Campbell, E.G. (2006). Data withholding and the next generation of scientists: Results of a national survey. *Academic Medicine, 81*: 128-136.

Wallis, J.C., Rolando, E. and 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*, viewed 4 May 2016, http://dx.doi.org/10.1371/journal.pone.0067332

Web of Science. (2016), viewed 8 September 2016, http://isiknowledge.com