

Research data management in Turkey: perceptions and practices

Research data
management
in Turkey

Arsev Umur Aydinoglu
Middle East Technical University, Ankara, Turkey, and
Guleda Dogan and Zehra Taskin
Hacettepe University, Ankara, Turkey

271

Received 23 November 2016
Revised 22 February 2017
Accepted 30 March 2017

Abstract

Purpose – The massive increase in research data being produced nowadays has highlighted the importance of research data management (RDM) to science. Research data not only have to be cost effective but also reliable, discoverable, accessible, and reusable. In this regard, the purpose of this paper is to investigate the perceptions and practices of Turkish researchers on the subject of RDM.

Design/methodology/approach – An online survey was distributed to the academicians in 25 universities in Turkey, and 532 responses were gathered.

Findings – Results indicate that although Turkish researchers are aware of the benefits of data management, are willing to share their research data with certain groups, and have decent preservation habits, they express that they lack the technical skills and knowledge needed for RDM. In addition, no institutionalized support (staff, training, software, and hardware) is provided to researchers.

Research limitations/implications – A well-structured data strategy or policy that includes resource allocation (awareness, training, software/hardware) and is supported by Turkish research agencies is required for better data management practices among researchers in Turkey.

Originality/value – This is the first study that investigates the data practices of Turkish academics who produce around 30,000 scientific articles annually that are indexed by Web of Science. It contributes to the growing literature on RDM.

Keywords Data preservation, Research data management, Data storage, Data repository, Data sharing, Research data

Paper type Research paper

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 (www-01.ibm.com/software/data/bigdata/what-is-big-data.html), 90 percent 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 (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 zettabytes in 2016 and is estimated to be 40 zettabytes in 2020. Therefore, data management has become an important issue.

Likewise, in the scientific arena, data have 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

The study was funded through the TÜBİTAK-Marie Curie FP7 Co-funded Brain Scheme (Project No. 114C011). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the paper. The authors would like to acknowledge the DataONE Usability and Assessment Group for preparing the original survey and sharing their survey and data sets from the original study with public.



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 data sets 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 Birmholtz and Bietz’s (2003, p. 339) study, data are 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 RDM systems. In the light of all these developments and rising importance of “RDM” 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:

- RQ1.* What are the common research data types and formats among Turkish scholars?
- RQ2.* To whom and what degree research data are shared in Turkey?
- RQ3.* What are the main reasons for not sharing research data with others?
- RQ4.* What are the most preferred places to store the data?
- RQ5.* What is the awareness level of scholars about the benefits of data sharing?
- RQ6.* 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 data sets. High-performance computers and advanced software help scientists to process large arrays of data sets to produce results that could be later reused, tested, and verified. High-quality data sets, 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 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 and Jacobsen, 2010; Tenopir *et al.*, 2011; Corral *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 (National Science Foundation, 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 data sets 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 data sets 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 and Vision, 2013; Tenopir, Dalton, Allard, Frame, Pjesivac, Birch, Pollock and Dorsett, 2015; 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, data set 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, Hughes, Allard, Frame, Birch, Baird, Sandusky, Langseth and Lundeen, 2015).

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 data sets 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 US 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 USA, 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 (National Science Foundation, 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 (National Institute of Health, 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 (National Aeronautics and Space Agency, 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, 2013). The first paper was a conference proceeding on the challenges of research data practices for environmental scientists (Allard and Aydınoglu, 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, 2015). 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 (Önder, 2013; Gürdal and Bitri, 2015; Malkoç, 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 (Aydınoglu *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 survey asks about demographic information; data management practices (types of data collected, data formats, metadata standards; and 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 Turkey[1]. 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 ULAKBİM, 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 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.

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. In total, 11 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 percent.

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 percent confidence interval for $e=0.05$, and 582 participants indicate a 99 percent confidence interval for $e=0.05$. Therefore, we are satisfied with the number of participants to our survey:

$$n_0 = \frac{z^2 pq}{e^2} \tag{1}$$

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} \tag{2}$$

University	No. of publications per person	E-mails sent (<i>N</i>)	E-mails responded (<i>n</i>)	Response rate % (<i>n/N</i>)
Hacettepe University	3.46	2,096	74	4
Ankara University	3.01	1,955	31	2
Ege University	3.36	1,513	67	4
Middle East Technical University	3.84	1,078	50	5
Erciyes University	2.94	1,247	29	2
Ataturk University	2.88	1,742	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 Mayıs 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
Kahramanmaraş 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	–	19,631	532	3

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

In the formulas, N is the population size, n_0 the sample size, n corrected sample size, z the table score for the selected confidence interval, p the estimate of variance, q the $1 - p$, e the 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 χ^2 tests are employed.

Findings

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

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. In total, 25 percent of the respondents can spend a minimum of 10 hours/week on research, and 10 percent 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.

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).

Table II provides the responses on data types. According to the responses, experimental data (52.8 percent) and text data (47.0 percent) are the two types of data that were used by half of the respondents. Survey data use is also significant (~41 percent). Approximately,

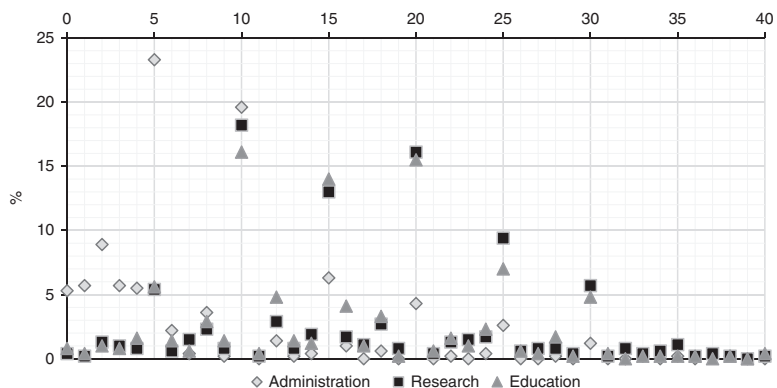


Figure 1.
Distribution of
40 hr/week work
on administrative
duties, research,
and education

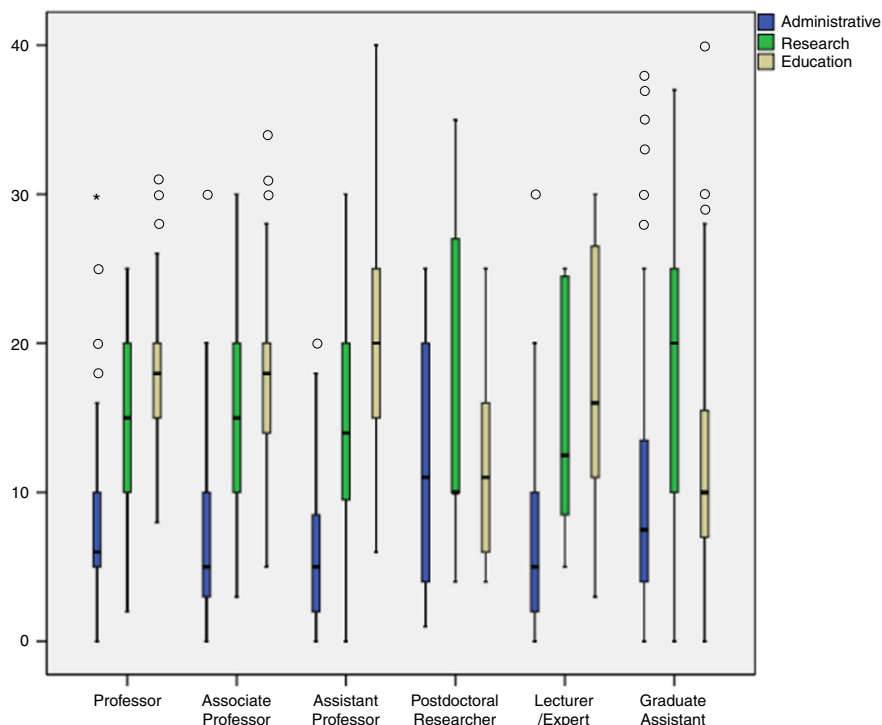


Figure 2. Distribution of 40 hr/week time spent on administrative duties, research, and education according to academic titles

Data type	<i>n</i>	%	χ^2	<i>p</i>
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	–	–

Note: “–” indicates that because of a high number of “no” responses, χ^2 test cannot be applied

Table II. Research data types (frequencies and χ^2 test results for academic title)

a quarter of the respondents reported that they use other types of data: still image (pictures and photos) (26.1 percent) and model-algorithm-code data models (25.2 percent), lab notebook (22.7 percent), and audio recordings (22.2 percent). Only a small group (2.8 percent) mentioned that they do not use research data. The χ^2 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 percent), associate professors (62.6 percent), and postdoctoral researchers (64.3 percent) is greater than any other academic rankings. The use of text data is greater among graduate assistants (55.1 percent)

and assistant professors (49.5 percent). Postdoctoral researchers utilize survey data the most (64.3 percent). 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 percent) and lecturers/experts (34.6 percent). Audio data are popular among the graduate students as well (29.5 percent).

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 percent). One-third of the respondents indicated text, and 30.1 percent reported free text. A little over a quarter of the respondents (27.4 percent) uses SAV format. SAV and XML as data formats are favored more by postdoctoral researchers (57.1 percent, $\chi^2 = 18.923$, $p = 0.002$, and 28.6 percent, $\chi^2 = 14.683$; $p = 0.012$). 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).

A striking result is that 27.1 percent 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 percent) reported that they record metadata, which is an extremely low figure. Academicians mostly use the metadata standard they developed in their lab (13.3 percent, $n = 71$). The second most frequent metadata standard is ISO (8.8 percent, $n = 47$). Each of the standards (AWM, DwC, DIF, EML, FGDC, CSDGM, NISO, MIX) account for less than 1 percent.

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 percent) reported that they do share, whereas 37.4 percent reported they do not. Among the 62.4 percent, when they are asked with whom they share their data and to what degree, almost all (98.9 percent) answered that they share their data with their research team, followed by scholars in their own discipline (76.6 percent), researchers in their organization (73.6 percent), and the scientific community (72.6 percent) (Figure 4).

For the respondents who do not share their research data with others (37.6 percent), their reasons for not sharing data are provided in Figure 5. The most important reason is not wanting others to access their data (65.5 percent). 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.

Different places are used to store data (see Table III). Local computers (71.6 percent) are the most common storage place. Close to half of the participants also use the cloud (45.9 percent). The use of an open access data repository is quite low (8.3 percent). However, the data suggest that the increase in academic title results in the decrease in the use of the

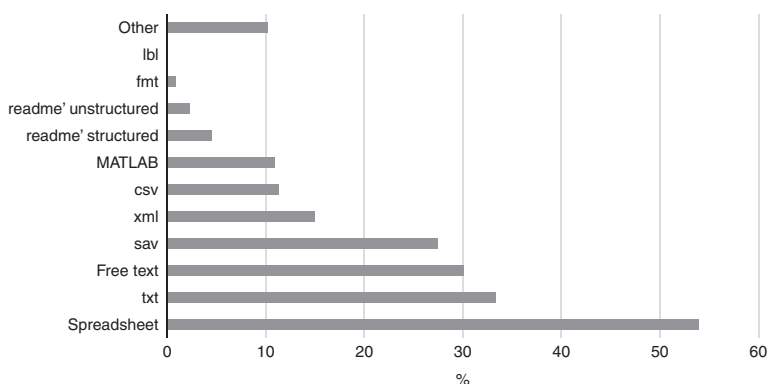


Figure 3.
Research data formats

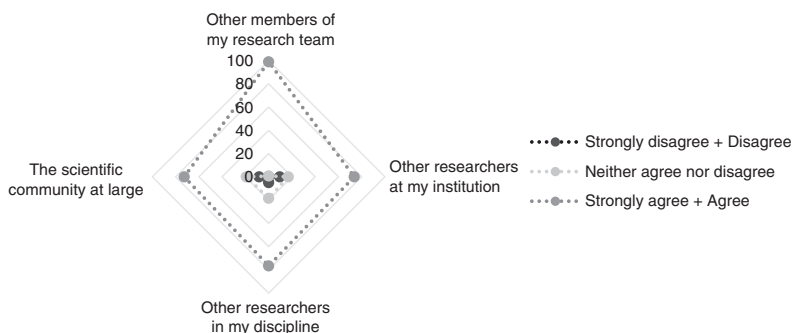


Figure 4. To whom and to what degree the research data is shared (%)

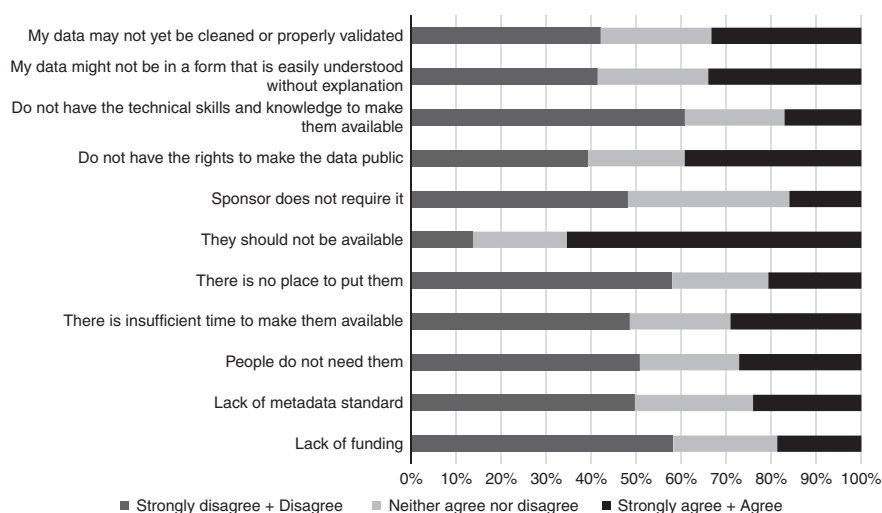


Figure 5. Reasons for not sharing data

Medium	<i>n</i>	%
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

Table III. Places to store data

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 percent, respectively).

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 percent). Only one out of 10 people (10.4 percent) use the cloud. Close to half of the respondents utilize both discs and the cloud (47.0 percent), 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 percent back up their data instantly, almost half of them (49.6 percent) back up once a week, and a quarter of them (25.8 percent) 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 percent) think that “well-maintained data helps retain data integrity.” Interestingly, fewer people (57.2 percent) agree with the statement that “data sharing reduces redundant data.” In total, 82 percent of the participants think that data sharing encourages interdisciplinary collaborative science. Moreover, 84.2 percent agree that data management practices are beneficial “to the scientific process itself (re-analysis of data helps verify results data),” 78.4 percent think that data sharing helps “the training of the next generation of researchers,” and 75.5 percent believe that data sharing “prevents data fabrication and falsification” (Figure 6).

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 percent of the academicians reported that an RDM plan is mandatory in their institutions. Around 30 percent of the participants do not know whether an RDM policy is in effect in their organization. One-fifth of the institutions (22.1 percent) support RDM in technical issues only. In total, 59.9 percent of the participants reported that no RDM procedure exists, and 59.3 percent state that no policy with regard to RDM exists in their institutions. Only 13.3 percent of the institutions provide training on RDM, and 11.8 percent provide monetary support for RDM (Figure 7).

Figure 7 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

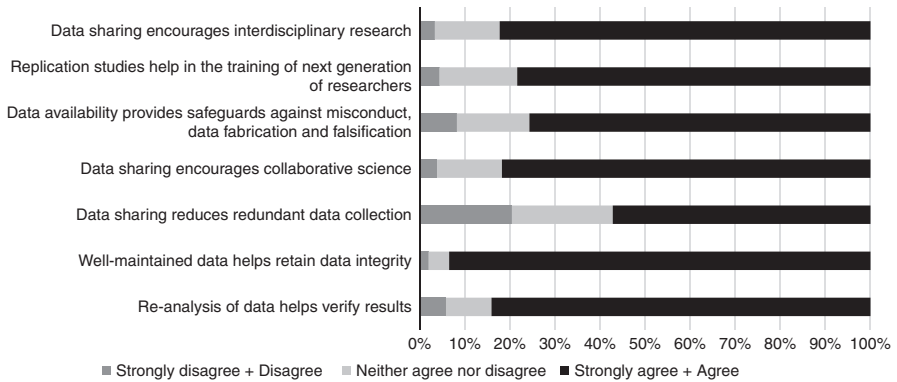


Figure 6.
Benefits of data sharing

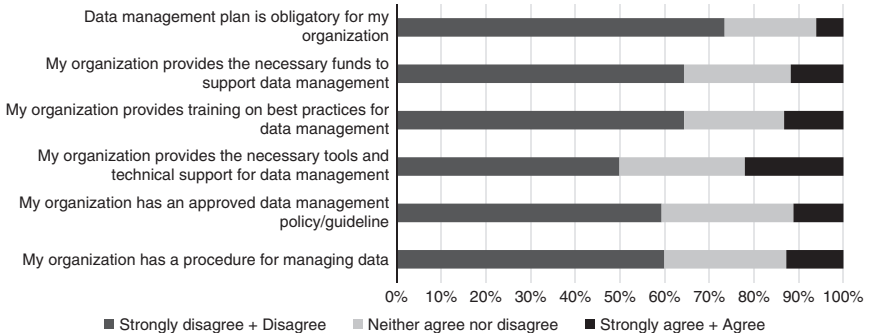


Figure 7.
Institutional support for RDM

disseminated work making use of the data” (92.8 percent) 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 percent); “results based on the data could not be disseminated in any format without the data provider’s approval” (84.3 percent); “mutual agreement on reciprocal sharing of data” (84.1 percent); and “the opportunity to collaborate on the project” (81.5 percent) (Figure 8).

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 and Aydınoglu (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,

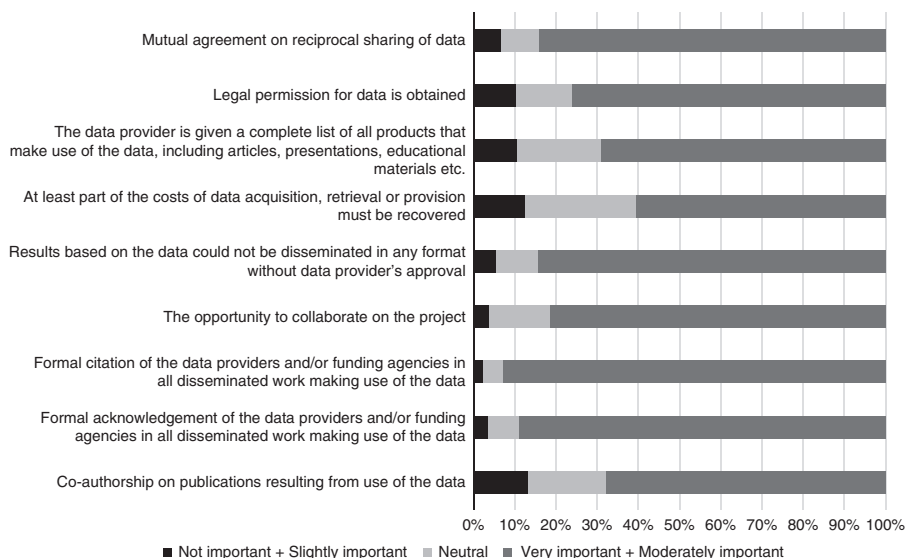


Figure 8. Conditions to sharing data with other researchers

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 data sets were posted by 48 Turkish scholar groups. Compared with the number of publications per year (~30,000) (Web of Science, 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 are 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; Tenopir, Dalton, Allard, Frame, Pjesivac, Birch, Pollock and Dorsett, 2015). 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.

Note

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

References

- Allard, S. and Aydinoglu, A.U. (2012), "Environmental researchers' data practices: an exploratory study in Turkey", in Kurbanoglu, S., Al, U., Erdogan, P.L., Tonta, Y. and Ucak, N. (Eds), *E-Science and Information Management, IMCW 2012, Communications in Computer and Information Science*, Vol. 317, Springer, Berlin and Heidelberg, pp. 13-24.
- 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", *Proceedings of the 2003 International ACM SIGGROUP Conference, FL*, pp. 339-348.
- 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., John Wiley and Sons, Inc., New York, NY.
- Corral, 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., 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", *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, ACM, New York, NY, pp. 295-304.
- 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 Nos 3-4, pp. 355-375.
- 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, February 4-6, available at: <http://ab.org.tr/ab15/ozet/124.html> (May 5, 2016).

- Hey, T., Tansley, S. and Tole, K. (2009), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, Redmond, WA, available at: 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", available at: www.gsrt.gr/EOX/files/h2020-hi-0a-data-mgt_en.pdf (accessed February 2, 2016).
- IMCW2014 (2014), "Research data management: getting your organization started", Workshop, available at: <http://imcw2014.bilgiyonetimi.net/> (accessed May 4, 2016).
- 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, FL.
- Malkoç, B. (2015), "Research data alliance ve DataCite", paper presented at 4. Ulusal Açık Erişim Çalıştayı (4th National Open Access Workshop), Ankara, October 19-21, available at: www.acikerisim.org/dokumanlar/ae2015_program.pdf (accessed May 11, 2016).
- National Aeronautics and Space Agency (2016), "Open gov plan 2016 outline", available at: <https://open.nasa.gov/blog/open-gov-plan-2016-outline/> (accessed August 6, 2016).
- National Institute of Health (2003), "Final NIH statement on sharing research data", available at: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html> (accessed May 4, 2016).
- National Science Foundation (2007), "NSF 07-28, cyberinfrastructure vision for 21st century discovery", available at: www.nsf.gov/pubs/2007/nsf0728/index.jsp (accessed July 6, 2016).
- National Science Foundation (2010), "Data management for NSF SBE directorate proposals and awards", available at: www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf (accessed May 4, 2016).
- OECD (2007), "OECD principles and guidelines for access to research data from public funding", available at: www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm (accessed May 4, 2016).
- Önder, A. (2013), "Büyük veri (Big data)", paper presented at 2. Ulusal Açık Erişim Çalıştayı (2nd National Open Access Workshop), Izmir, October 21-22, available at: www.acikerisim.org (accessed May 6, 2016).
- Piowar, H.A. and Vision, T.J. (2013), "Data reuse and the open data citation advantage", *PeerJ*, available at: <https://peerj.com/articles/175/> (accessed September 9, 2016).
- Powell, K. (2016), "Young, talented and fed-up: scientists tell their stories", *Nature News*, October 26, pp. 446-449, available at: www.nature.com/news/young-talented-and-fed-up-scientists-tell-their-stories-1.20872 (accessed February 16, 2017).
- SINTEF (2013), "Big data, for better or worse: 90% of world's data generated over last two years", *ScienceDaily*, May 23, available at: 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*, Vol. 6 No. 6, available at: <http://dx.doi.org/10.1371/journal.pone.0021101> (accessed June 5, 2016).
- Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D. and Dorsett, K. (2015), "Changes in data sharing and data reuse practices and perceptions among scientists worldwide", *PLoS One*, Vol. 10 No. 8, available at: <http://dx.doi.org/10.1371/journal.pone.0134826> (accessed May 8, 2016).
- Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, W.B., Baird, L., Sandusky, R., Langseth, M. and Lundeen, A. (2015), "Research data services in academic libraries: data intensive roles for the future?", *Journal of eScience Librarianship*, Vol. 4 No. 2.

- Tonta, Y. (2012), "Açık erişim, kurumsal arşivler ve MedOANet Projesi (Open access, institutional repositories and MedOANet Project)", paper presented at Ulusal Açık Erişim Çalıştay (National Open Access Workshop), Ankara, November 8-9, available at: www.acikerisim.org/sunumlar/yasar_tonta.pdf (accessed May 9, 2016).
- 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, December 17, available at: <http://library.bilkent.edu.tr/activities/librarianship-seminars/presentations/yasar-tonta.pptx> (accessed May 8, 2016).
- Tonta, Y. and Al, U. (2015), "Araştırma verilerinin yönetimi (research data management)", *Türk Kütüphaneciliği [Turkish Librarianship]*, Vol. 29 No. 1, pp. 36-45.
- TUBITAK ULAKBİM (2016), "Türkiye üniversitelerinin bilimsel yayın performansı: 2004-2014 (Scholarly production performance of Turkish universities: 2004-2014)", available at: <http://ulakbim.tubitak.gov.tr/tr/hizmetlerimiz/turkiye-universitelerinin-bilimsel-yayin-performansi-2004-2014> (accessed March 3, 2016).
- 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*, Vol. 81 No. 2, pp. 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*, Vol. 8 No. 7, available at: <http://dx.doi.org/10.1371/journal.pone.0067332> (accessed May 4, 2016).
- Web of Science (2016), available at: <http://isiknowledge.com> (accessed September 8, 2016).

Further reading

- IBM (2016), "Bringing big data to the enterprise", available at: www-01.ibm.com/software/data/bigdata/what-is-big-data.html (accessed April 19, 2017).

Appendix. Research Data Management in Turkey – Survey Instrument (translated from Turkish)

What percentage of your work time overall is allocated to the following activities? (total to equal 100%) % (0-100)

Administration
Outreach
Research
Teaching
Other

What is your subject discipline?

Which terms best describe the type of data you use (check all that apply)?

- Abiotic surveys (soils, microclimate, hydrology, etc.)
- Biotic surveys
- Data models
- Experimental (involving some degree of manipulation)
- Observational (no manipulation involved)
- Remote-sensed abiotic data (including meteorological data)
- Remote-sensed biotic data
- Social Science Survey data (ethnographic interviews, quantitative surveys, interviews, etc.)

LHT
35,2

- Other (please specify)
- No Answer

What format is used to describe your data? (check all that apply)

- Free Text
- “readme” (structured)
- “readme” (unstructured)
- Spreadsheet
- CSV
- FMT
- LBL
- TXT
- XML
- MATLAB
- Other (please specify)
- No Answer

286

Metadata is data about data that describes what data is about and how the data is formatted. This would include how, when, where and by whom a particular set of data was collected. Do you capture metadata?

- Yes
- No
- Don't know
- No Answer

Are you using any scientific metadata standards?

- ABCD (Access to Biological Collections Data Schema)
- AVM (Astronomy Visualization Metadata Standard)
- DC (Dublin Core)
- DwC (Darwin Core)
- DIF (Directory Interchange Format)
- EML (Ecological Metadata Language)
- FGDC Metadata Standard (Federal Geographic Data Committee)
- CSDGM (Content Standard for Digital Geospatial Metadata)
- ISO (International Standards Organization)
- NetCDF (Network Common Data Form)
- NISO MIX
- OGIS (Open GIS)
- Metadata standardized within my lab
- None
- Other (please specify)
- No Answer

In general, I currently share my data with: (check all that apply)

- no one other than myself
- other members of my research team
- other researchers at my institution
- other researchers in my discipline
- the scientific community at large
- Other (please specify)
- No Answer

My data is stored on: (check all that apply)

- on a personal computer connected to the network
- on a computer operated by my research team
- on a computer operated by my school, company, or organization
- on an open access repository
- on a commercial repository
- on a public website
- Other (please specify)
- No Answer

What percentage of your data is stored on: % (0-100)

- on a personal computer connected to the network
- on a computer operated by my research team
- on a computer operated by my school
- company, or organization
- on an open access repository
- on a commercial repository
- on a public website
- Other

I currently backup my research data using the following methods:

- cloud storage
- CD / DVDs / BluRay Disk
- hard drive (internal to PC)
- hard drive (external to PC)
- magnetic tape
- flash disk
- Other (please specify)
- No Answer

I backup my research data (frequency):

- immediately
- daily

- weekly
- bi-monthly
- monthly
- Other (please specify)
- I don't backup my research data
- No Answer

If your data are not available electronically to other researchers, why not (check all that apply)?

- Lack of funding
- Lack of metadata standards
- People don't need them
- There is insufficient time to make them available
- There is no place to put them
- They shouldn't be available (please give reason why not)
- Sponsor doesn't require it
- Don't have the rights to make the data public
- Don't have the technical skills and knowledge to make them available
- Other (please specify)
- No Answer

The following group of statements relates to your views on why your data should not be shared. Tell us how much you agree with each using the following scale: agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly.

My data might not be in a form that is easily understood without explanation.

My data may not yet be cleaned or properly validated.

The following group of statements relates to how you collect and use research data. Tell us how much you agree with each statement using the following scale: agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly.

I am satisfied with the process for collecting my research data.

I am satisfied with the process for searching for my own data.

I am satisfied with the process for cataloging/describing my data.

I am satisfied with the process for storing my data during the life of the project.

I am satisfied with the process for storing my data beyond the life of the project.

I am satisfied with the process for analyzing my data.

I share my data with others.

Others can access my data easily.

I am satisfied with the tools for preparing metadata.

I am satisfied with the tools for preparing my documentation.

On the benefits of data sharing and/or preservation...

Tell us how much you agree with each using the following scale: agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly.

Re-analysis of data helps verify results data.

Well-maintained data helps retain data integrity.

Data sharing reduces redundant data collection.

Data sharing encourages collaborative science.

Data availability provides safeguards against misconduct, data fabrication and falsification.

Replication studies help in the training of next generation of researchers.

Data sharing encourages interdisciplinary research.

The following group of statements relates to how your organization is involved with your data. Tell us how much you agree with each using the following scale: agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly.

My organization or project has a formal established process for managing data.

My organization or project provides the necessary tools and technical support for data management.

My organization or project provides training on best practices for data management.

My organization or project provides the necessary funds to support data management.

How important are the following conditions to sharing your data with other researchers? Not important, Slightly important, Neutral, Moderately important, Very important.

Co-authorship on publications resulting from use of the data.

Formal acknowledgement of the data providers and/or funding agencies in all disseminated work making use of the data.

Formal citation of the data providers and/or funding agencies in all disseminated work making use of the data.

The opportunity to collaborate on the project (including, for example, consultation on analytic methods, interpretation of results, dissemination of research results, etc.).

Results based (at least in part) on the data could not be disseminated in any format without the data provider's approval.

At least part of the costs of data acquisition, retrieval or provision must be recovered.

The data provider is given a complete list of all products that make use of the data, including articles, presentations, educational materials, etc.

Legal permission for data use is obtained.

Mutual agreement on reciprocal sharing of data.

Preliminary data should be labeled as such so that people know when data are not completely vetted.

Data would be available to team, but not outside the team, until publication.

Do you have the sole responsibility for approving access to your data?

- Yes – for all my datasets
- Yes – for some of my datasets
- No
- No Answer

My primary project funding agency requires me to provide a data management plan.

- Yes
- No
- Don't know
- No Answer

List the data repositories you use to access and/or submit research data, if any.

Your title

Your institution

Your department

Gender

Corresponding author

Arsev Umur Aydinoglu can be contacted at: arsevu@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com