

**ATTITUDES AND PRACTICES OF DATA SHARING AND DATA
PRESERVATION AMONG SOCIAL SCIENCE RESEARCHERS IN
THE REPUBLIC OF NORTH MACEDONIA**

Vesna ZABIJAKIN- CHATLESKA

Associate professor, Ss. Cyril and Methodius University in Skopje,
Institute for sociological, political and juridical research
E-mail: vesna.catleska@isppi.ukim.edu.mk

Aneta CEKIKJ

Associate professor, Ss. Cyril and Methodius University in Skopje,
Institute for sociological, political and juridical research
E-mail: anetac@isppi.ukim.edu.mk

Abstract

The paper aims to analyse current practices and attitudes towards research data sharing among social science researchers in the Republic of North Macedonia. Considering the growing importance of the principle of open access to research data, particularly data collected with public funding, and the unification of data sharing practices on a global level, the goal is to examine the sharing and archiving potential for research data in the Republic of North Macedonia. According to the findings, a significant amount of research data is created in the country. Despite being predominantly financed from international funds, respondents' positive attitudes towards data sharing and the satisfactory degree of willingness for depositing data into an accredited data archive/repository, demonstrate a good potential in the country for saving existing data towards further use of existing data for new scientific purposes.

Key words: *research data sharing, open access, science policy, research data archiving, R. North Macedonia*

INTRODUCTION

Research data sharing and the open access initiative are among the more important scientific developments in the last three decades, when information and communications technology peaked its development. New technologies considerably influence scientific research by transforming the processes of conducting, managing and documenting research data and research findings. Digital materials and outputs had a huge impact on research practices. All this is known as e-science. Actually, a new research paradigm was established, called ‘the fourth paradigm: data-intensive scientific discovery’ where ‘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 Tenopir *et al*, 2011:1).

The question of data sharing and preservation is far from recent, dating before the 1970s of the 20th century. Already in the 1990s, developed countries began designing policies for national preservation and access to research data, mostly data from publically funded projects. The definition of research data sharing policies is exceeded by the Organization for Economic Cooperation and Development’s (OECD) member countries, particularly the United States of America (USA). The European Union (EU) reached a higher degree of development in respect to defining the rules in this field in the last 5-10 years. After a pilot phase, from 2017, for projects financed by the biggest research program – Horizon 2020, in addition to publications from conducted research, there is a requirement also for preparation of a data management plan by researchers, and proper documentation and deposition of the collected primary research data. At present, the European Open Science Cloud (EOSC), a large platform uniting in one place digital outputs and resources from all scientific disciplines is active, with the Social Sciences and Humanities Open Cloud (SSHOC) functioning within EOSC specifically for the social sciences and humanities.

Research data sharing is significant for numerous reasons, the most highlighted in the literature being: shared data are used in new analyses towards addressing new research questions; accumulation of scientific knowledge; encouraging interdisciplinarity; transparency and accountability in science; increased quality through diverse insight in data by numerous parties etc. Reusing/different use of existing data is cost-efficient in research. Also, data can be used for education purposes by academic staff/faculty and students.

While the open access initiative has been introduced to and is part of official government policies in Balkan countries, with state institutions publishing documents and publications available for open access, data sharing among scientists from the social sciences remains limited. Until recently, Slovenia was the only country in the Balkans with its own research data archive. The number of institutional repositories also remains minimal. However, in light

of the future joining of the European research area, in cooperation with the respective ministries, progress was made, and now North Macedonia, Croatia and Serbia are in the beginning phase of establishing national archives, becoming members of the Consortium of European social science data archive (CESSDA ERIC) in 2019. Although these data archives are not yet fully operational, as a result of the several international projects, as well as the cooperation with CESSDA ERIC, institutions have at their disposal trained staff for archiving and disseminating research data.

The paper aims to research the attitudes social sciences researchers in North Macedonia have towards sharing research data, i.e. their current practices and willingness to share data, as well as their need for a specialized institution dealing with archiving and dissemination of such data. The paper is based on data collected from a research conducted within South East European Data Services (SEEDS) project.¹

IMPORTANCE AND COMPLEXITY OF RESEARCH DATA SHARING

Research data sharing is a significant part of contemporary living, despite being considered as solely the interest of scientists. The present context, having to deal with the COVID-19 pandemics, is an excellent example illustrating the importance of research data sharing to society and the world. Namely, the latest knowledge on the virus is being published with open access in the most renowned journals, under quick review processes and with priority. In March 2020, over 700 genome sequences of the virus were already shared among scientists from different research centres towards finding a treatment and vaccine for the illness, which would accelerate this process significantly (Le Guillou, 2020). Authors often list the prevention of infectious diseases as an example to illustrate the importance of data sharing (Arzberger *et al*, 2004, Tenopir *et al*, 2011).

Before going into specification of the various aspects of data sharing principles and practices, a short elaboration of the key terms ‘data’ and ‘sharing of data’ is needed. According to the widely cited definition of OECD: “data are defined as the factual records (numerical scores, textual records, images and sounds) used as sources and base material for scientific research.” (Arzberger *et al*, 2004, p. 139). According to Borgman, sharing of

¹An international project financed by the Swiss National Science Foundation, conducted from May 2015 to April 2017, coordinated by the Swiss centre for the expertise in the social sciences (FORS), at the University of Laussane. Project web site:<https://seedsproject.ch/>

data means making data available (release) to others to use (Borgman, 2012). Kim and Adler define the term 'data sharing' "...as an individual scientist's behaviour in providing their raw (or preprocessed) data of his/her published work to other scientist by making it accessible through central/local data repositories or by sending data via personal communication methods upon request." (Kim and Adler, 2015, p. 409). Arzberger and his colleagues understand the term 'access to data' as "the act of making the data available for use by others..." (Arzberger *et al*, 2004, p. 138), while the term 'sharing' means "a researcher allowing one or more other individuals to use data, typically with the implicit, if not explicit assumption that it is on a reciprocal basis" (Arzberger *et al*, 2004, p. 139).

Borgman suggests that the release of data can have multiple forms: from a private exchange at a request, to depositing in a public data collection. Sharing also includes posting on a website or sending data to journals, as an additional material to a paper (Borgman, 2012). Similar to Borgman, (Mauthner, 2018) makes a distinction between two basic manners of data sharing: informal data exchange between researchers and formal data exchange with the mediation of specialized institutions/agencies, such as data archives and repositories. Informal data sharing involves researchers and/or institutions directly asking and receiving copies of data produced by other researchers and/or institutions. Inadequate data structuring and documentation, as well as unstandardized electronic formats of data files and folders is considered a serious weakness in the informal manner of data sharing.

Questions on sharing and open access to research data are becoming an important agenda within national and international frameworks, mostly due to the fact that a significant part of the research is financed from public funds. Such is the opinion of the Organization for Economic Cooperation and Development (OECD), whose basic recommendation is that 'publicly funded research data should be openly available to the maximum extent possible' (Arzberger *et al.*, 2004, p. 136). Funds invested in science are perceived as an investment meaning to encourage value chain (Arzberger *et al.*, 2004, p. 135). Consequently, in 2004, ministers responsible for national science and technology policies of OECD countries adopted a Declaration on Access to Research Data from Public Funding, while in 2007 the OECD Principles and Guidelines for Access to Research Data from Public Funding (OECD, 2007) were developed.

Regulation of access to publically funded data in the United States of America has been functioning since 1997 with certain restrictions, for instance, security questions (national security). Starting from January 2011, NSF funded research must include a data management plan so that "digital data are routinely deposited in well-documented form, are regularly and easily

consulted and analysed by specialist and nonspecialist alike, are openly accessible while suitably protected, and are reliably preserved” (Tenopir *et al.*, 2011, p. 3). This is the case with most European countries, while on EU level, the issue has been prevailing in the past six years and it refers to research outputs – publications and raw data – from the Horizon 2020 program.

Analysing policy documents and studies of data sharing, Borgman distinguishes four rationales for data sharing: (a) reproduction and verification of research, (b) availability of publicly funded research to the public, (c) possibility to ask new questions of existing data, and (d) advancement of science and innovation (Borgman, 2012, p. 1067). To this we can also add data use for teaching goals, interdisciplinarity development, and historical reasons for preservation, transparency and accountability in science, cost-efficiency and numerous other benefits. However, organization of data sharing, and particularly international data sharing, is not a simple task having in mind the variety of data, variety of existing practices for data documentation, differences across disciplines, and differences in the opinions and attitudes of researchers, policy makers and potential users. Arzberger *et al.* point to five broad group of issues related to data sharing and access: 1) issues related to technological infrastructure and its operation; 2) institutional and managerial issues of the specialized institutions, 3) financial and budgetary issues, 4) legal and policy issue and 5) cultural and behavioural issues of data producers and data managers (Arzberger *et al.*, 2004, p. 136).

There are significant differences among practices and regimes of data sharing in scientific fields and disciplines. Some disciplines, such as meteorology, genomes (genome sequencing), astronomy and other, have a long tradition of free data sharing. In other scientific disciplines the situation is quite different. Data from survey research tell us that respondents from the medical fields and social sciences are less likely to make their data electronically available to others. While 90% of respondents of atmospheric science and 85% of biologists report sharing their data with others, this is only so with 65% of researchers in medicine, 64% in computer science/engineering and 58% in social sciences (Tenopir *et al.*, 2011, p. 11-12). The authors point to various limitations to data sharing in medicine and social sciences, because they have human subjects as units of analysis. At the same time, social scientists (80%) agree at a higher rate than respondents from other disciplines that lack of access to research data generated by others is a major impediment to progress in science (Tenopir *et al.* 2011, p. 13).

Although the benefits of data sharing are obvious, in the context of social sciences, policies and standards are inconsistent among different subdisciplines of social sciences, and despite the significant progress in the last ten years, data sharing is smaller than expected. Pienta, Alter, and Lyle, who compiled a database with administrative data from 40 years of

social science research and found that “very few social science data collections are preserved and disseminated by an archive or institutional repository” (Pienta, Alter, and Lyle (2010) in Kim and Adler, 2015, p. 409). Although the situation has probably changed in recent years as a result of public policies in this area increasingly creating requests, rules and procedures for data depositing, still data sharing in social sciences remains limited. Below we present the most important conditions and factors affecting the behaviour of researchers regarding research data sharing.

ATTITUDES, INCENTIVES AND BARRIERS IN DATA SHARING

A basic overview of data sharing literature, from the perspective of researchers as creators but simultaneously potential data users, reveals that data sharing attitudes and practices are influenced by factors from institutional/organizational nature and factors from individual nature. Institutional factors are considered to be existing data sharing policies and practices on a national and disciplinary level, requirements of research funders for availability of the data gathered, requirements from scientific journals, as well as the rules of the organization where researchers work at regarding the data produced (whether data depositing is compulsory, and whether unified internal rules and procedures exist). Individual factors are also a topic of analysis, focusing mostly on researchers’ motivation and perceived benefits and barriers/risks related to data sharing. Fecher *et al* (2015, p. 9) quote studies examining the impact of personality traits (big 5 personality traits) on attitudes of data sharing. Kim and Adler list normative expectations among the disciplines as a factor in the data sharing attitudes and practices (Kim and Adler, 2015, p. 410), and also speak of “resource factors”, which is basically data repository existence.

Their comprehensive research analyses the role in most abovementioned individual, institutional and resource factors for data sharing.² They found that “social scientists’ data sharing behaviours are significantly driven by personal motivations (i.e., perceived career benefit and risk, perceived effort, and attitude toward data sharing) and perceived normative pressure”. Pressures from funding agencies, journal requirements and availability of data repositories do not affect the behaviour of researchers to a larger extent” (Kim and Adler, 2015, p. 414). This is a very important finding which shows the direction of policies and measures, if authorities aim to achieve better data

² The study was conducted on a random sample of all researchers in the social sciences in the USA (N=77 784), with 361 responses out of 2285 (sample size); response rate 15.80%.

sharing. Namely, they should focus more on securing benefits and eliminating the risks that impede data sharing.

A recent survey conducted by the Swiss Centre for the Expertise in the Social Sciences (FORS) reveals that the users of this data base share data usually at a colleague's request (64%), but also have other motives, such as creating new collaborations (53%). In addition, it is important to note that almost half of the respondents share data 'by principle' (48%), indicating to a data sharing culture where this practice is voluntary. Funders' (22%) and scientific journals' requirements (13%) are less common reasons for sharing data (Heers *et al*, 2017).

The most important **benefits** of sharing data for researchers is the citation of their work and possibility of co-authorships (In some disciplines the researchers who provide the data receive authorship of the publication). According to Tenopir *et al*, (2011) the most important condition for sharing their data with other researchers is to receive proper citation credit when others use their data. For 92% of the respondents, it is important that their data are cited when used by other researchers, while 61% find it fair to use other people's data if they give them co-authorship on publications resulting from use of the data. A vast majority (93%) find it a fair condition 'to use other people's data if there is formal acknowledgement of the data providers and/or funding agencies in all disseminated work making use of the data' (Tenopir *et al*, 2011, p. 10).

At present, acknowledgements for authors of data differ from acknowledgements for authors of publications. Fecher stresses that, for instance, there isn't an impact factor measuring the influence of datasets following the example of publications (Fecher *et al*, 2015, p.11).

According to Tenopir and al, also the possibility of assigning Digital Object Identifiers (DOI's) to data in order to cite material was particularly important for researchers (Tenopir *et al*, 2011, p. 4).

Some of the most important **risks** connected to data sharing include possible misuse and misinterpretation of data, the criticism that researchers who produced the data might face, and losing publication opportunities in case other researchers manage to publish results before the actual data producer. Consequently, an embargo on data release in a certain period is among the main instruments researchers use when sharing their data. This means that researchers who create data will receive a guaranteed period to publish their analyses/publications based on this data, before sharing with others or depositing the data to a repository. This argument is often used to encourage data sharing since it is perceived as an important risk by researchers. (Eschenfelder and Johnson, 2011, in Fecher *et al*, 2015, p. 10)

According to Fecher, the degree of control on once deposited data is also an important factor in the willingness to share data. According to his study, only 18% of researchers refuse to share data, while the degree of control on the distribution process of their data is central for the rest. Therefore, 21% of the latter would share their data provided they maintain control over access (usually through a user registration system), while 35% would share their data only if allowed to personally approve the request (Fecher *et al*, 2015, p. 10). This manner of data sharing nowadays is available through the mediation of data archives/repositories and is a standard possibility, provided the funder's legal stipulations allow it. Similar are the findings of the research of Tenopir and his colleagues. When asked whether they were willing to place all of their data into a central data repository with no restrictions, 41% of the respondents were not willing to place all of their data. Nearly two thirds of the respondents (65%) reported that they would be more likely to make their data available if they could place conditions on access (Tenopir *et al*, 2011, p. 9).

The necessary time and resources for data publications is also a factor discouraging researchers from sharing data (Fecher *et al* 2015, p. 11). Tenopir's research shows that the main reasons not to share data for 56% of researchers is insufficient time, while for 40% it is lack of financing (Tenopir *et al* 2011, p. 9).

RESEARCH DESIGN

Research data sharing is still a relatively new topic in North Macedonia from the aspect of an institutionalized approach to this field. Our analysis of relevant legislation and national science policies indicate complete lack of recognition and regulation of this issue (Cekik *et al*, 2015). The situation of higher education and research institutions is similar regardless of whether the institution is public or private: institutional practices for documenting and preserving research data are rare. In the NGO sector it seems that data are better preserved on an organizational level in comparison to higher educational institutions, although these organizations mostly use internal, not international documentation standards (Cekik *et al*, 2015). At present, according to our knowledge, social science journals do not require depositing the datasets on which the published papers are based. We lack previous knowledge on researchers' attitudes. Consequently, in our research and this paper we analyse the basic characteristics of researchers' practices and attitudes on documenting and sharing research data, as well as the attitudes and opinions regarding the usefulness of a national research data archive for the social sciences. In fact, this research is exploratory by nature and has the goal to answer the following research questions:

- What are the existing practices for research data preservation and sharing?

- What are researchers' attitudes to data sharing generally, as well as their willingness to share data?
- What are researchers' attitudes regarding the need for a national institution specialized in archiving and dissemination of research data?

In order to receive answers to these research questions, we used a survey questionnaire with 38 open-ended and closed-ended questions divided in four sections/topics.

Sample and time frame of the research

The research was realized through an online survey conducted between July and September 2015. The sample included the whole target population-researchers in the social sciences in the R. North Macedonia who work in the public and private higher education and research institutions, as well as researchers from the NGO sector, which produces considerable amount of research relevant for social science research. We compiled the sample using information from higher education and research institutions' web sites and/or officially contacting them with a request for information. We also contacted 105 NGOs and think tank organizations and asked them to provide us with list of researchers engaged in collecting and analysing data within their organisations. The resulting dataset consists of 1,159 researchers with e-mail addresses that were contacted, and 156 for whom we could not provide e-mails. We are confident that we managed to map more than 90% of active researchers in social sciences in North Macedonia.

One invitation e-mail and two reminders were sent between July 2 and September 3, 2015. The survey was accessed by 278 researchers, with 181 completing the survey. Thus, our response rate stands at around 15% of the mapped population, and is comparable with similar research like Fecher, 2015.

DESCRIPTION OF RESPONDENTS

The majority of our respondents are (senior) researchers or university professors (60.2%). The second largest category of respondents is doctoral students or teaching/research assistants (13.8%). 12.7% of researchers were currently project leaders, 6.6% were heads of institutions. 3.9% of researchers were BA or MA students (most probably working in the NGO sector), while 2.8% chose the option "other".

Table 1. Frequency and percentage distribution of respondents by principal activity

Principal activity	Frequency	Percent
Researcher/professor	109	60.2
Doctoral students or teaching/research assistants	25	13.8
Project leader	23	12.7
Head of institution	12	6.6
BA or MA student	7	3.9
Other	5	2.8
Total	181	100

With regard to institutional affiliation, a majority of respondents work at a higher education institution (56.9%). The second largest category of respondents is affiliated with NGO/think tanks (21.5%), followed by the employees in university research institutes (10.5%). 2.8 percent of researchers are currently not employed, and 7.7% of researchers are affiliated with other types of organizations.

Table 2. Frequency and percentage distribution of respondents by institutional affiliation

Institutional affiliation	Frequency	Percent
Higher education institution	103	56.9
University research institute	19	10.5
Public research institute	1	0.6
NGO/Think tank	39	21.5
Currently not employed	5	2.8
Other	14	7.7
Total	181	100

48.9% of the respondents work in the public sector, 25.6% in the private sector, and 25.6% in the nongovernmental sector (Table 3).

Table 3. Frequency and percentage distribution of respondents by sector of employment

Sector of employment	Frequency	Percent
Public sector	88	48.9
Private sector	46	25.6
NGO sector	46	25.6
Total	180	100

With regard to research disciplines, economists and political scientists dominate among respondents in the survey, followed by researchers in education science, law, organizational sciences, and psychology (Table 4).

Table 4. Frequency and percentage distribution of respondents by principal research discipline

Principal research discipline	Frequency	Percent
Economics	38	21
Journalism	8	3.6
Political science	37	6.3
Education science and teacher training	15	8.3
Psychology	10	20.4
Sociology	7	4.4
Law	15	1.9
Business and administration	3	8.3
Organizational sciences/Management	13	7.2
Public administration	6	3.3
History	4	2.2
Anthropology	3	1.7
Other	22	12.2
Total	181	100

DATA PRODUCTION

In order to get a basic idea of the potential for data sharing in the country, the amount of data production and the founding sources of research produced by 2015 are briefly examined. According to our results, a significant number of social science staff in Macedonia is involved in research activities. In the 5 years prior to the research, 74.6% of the respondents produced or helped in producing research data. The mean number of produced datasets stands at 9.23, while the median, which we use because of the several outliers, is 5. On average, every year, one dataset per researcher is produced. This is not a small amount of research produced by social science researchers in the country, and the potential for data preservation is quite decent.

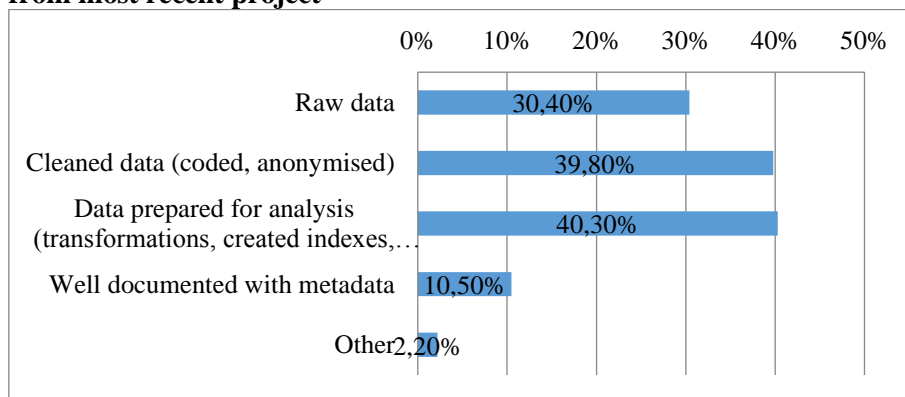
Funding of research is important issue for researchers and also important information when it comes to the potential for data sharing. In this survey, only 2% of researchers used public funds for their research, which is the obvious result from the very low investment in science in the last decade. The

level of private funding is also very low- it stands at 5%. The largest share of respondents (39%) indicated that their last research was financed by international funders, followed by 22% who said that their research was funded by the institution where they work. Other funding (90% of what is actually researchers' own private funding) is used for 8% of all research.

CURRENT DATA PRESERVATION, DATA DOCUMENTATION AND DATA SHARING PRACTICES

When asked if they have kept the data from their last project, 98% of researchers have indicated that they have kept the research data after the project was completed. However, only 10.5% of researchers keep their research data well documented with metadata. 39.8% of researchers keep data prepared for analysis (with transformations, created indexes, recoded), another 40.3% keep cleaned data, and 30.4% also keep raw data (multiple responses were possible). (Figure 1).

Figure 1. Percentage distribution of the answers about type of data kept from most recent project



When it comes to data preservation, aside from keeping their data on their own computers (39.8%), several kept copies on different computers/media (26.5%) or on colleagues' computers (14.4%), 16.6% of researchers indicated that their institution keeps the data- which is a somewhat more systematic approach to data preservation. However, only 12.2% of researchers have deposited their research to a data archive/repository.

When asked about documentation/metadata standards for description of research data, only one researcher indicated that he/she had used the ISO 11179 standard (DDI, Dublin core and ISO 11179 were listed as options). Also, 4% of researchers used some kind of an internal documentation standard of their institutions.

Who currently has access to these data is an important indicator of the situation with data access in the country (Table 5). The largest number of respondents (43.1%) answered that only the members of their research team have access to these data. Less number of respondents (18.8%) answered ‘members of my institution’ and even less (13.8%) ‘the broader scientific community’. Only 6.6% of the respondents stated that data are “publicly available – open access”, and slightly more (8.8%) reported that ‘the access is granted just to the project leader’.

For the largest number of respondents, the ideal level of access stands somewhere between ‘the broader scientific community’ (29.5%) and ‘open access’(31.1%). Still, 25.2% of the respondents have a somewhat restrictive understanding of data sharing- they think that only members of the research team should use these data. For 3.1% this should be only the project leader.

Table 5. Percentage distribution of answers for Current and Ideal level of access to researchers’ last project

	Current level of access (number and %)		Ideal level of access (number and %)	
	Number	%	Number	%
Just the project leader	16	8.8	4	3
Members of the research team	78	43.1	34	25.2
Members of my institution	34	18.8	9	6.7
Broader scientific community	25	13.8	40	29.5
Data is publicly available- open access	12	6.6	42	31.1
Other	11	6.1	4	3
Do not know	2	1.1	2	1.5
Total	178	100	181	100

The awareness of the importance of sharing research data within the researchers’ own discipline is quite high- four out of five respondents (80.1%) answered that sharing of data is “very important”, and 18.2 % estimated such practice as “somewhat important”. Additionally, very large percentage of

respondents (91.7%) have indicated that access to national and international research data will be beneficial for their scientific work.

If a data archive, which will enable proper preservation and access to data to accredited researchers, existed in the country, 38.9% of the respondents will 'certainly' deposit data from their research projects in it, while 44.2% of respondents would 'probably' do so. 14.2% of the respondents are not sure about that, while 0.9% probably won't share their data and 1.8% will certainly not share their data.

The final question asked the respondents 'How useful could be an institution that specializes in data archiving in the country'. For large majority of respondents (85.6%) a data archiving institution will be very useful, and 11.6% of somewhat useful.

FINDINGS FROM THE ANALYSIS OF THE RELATIONS BETWEEN THE VARIABLES

In order to examine more thoroughly the connection between researchers' attitudes and practices, and obtain additional information, we conducted a bivariate analysis with the application of certain statistical tests³.

The examination of the *relation between the number of produced datasets in the past five years*⁴ and the *manner of financing* revealed a significant difference in mean ranks ($U = 1577.000$, $p = 0.027$) between respondents according to the use of international funding⁵. Respondents who used international funding produced more data (mean rank = 71.14) than those who did not use international funding/project (mean rank = 56.73), (Table 6).

Furthermore, there was a significant difference in mean ranks ($U = 1233.500$, $p = 0.017$) between respondents regarding the use of institutional funding (the institution you work for paid from its own funds),⁶ regarding the number of produced datasets in the past five years, in that respondents who used institutional funding produced less data (mean rank = 52.34) in comparison to those who did not use such a manner of financing (mean rank = 69.45), (Table 6).

³ Spearman's rank-order correlation, Kruskal-Wallis H test and Mann-Whitney U test.

⁴ Ratio variable: "Enter the number of datasets that you have produced or helped to produce during the past 5 years."

⁵ Variable with several modalities, manner "Mark all that apply". Dichotomous variable produced, with "yes/no" answers.

⁶ Ibid

Table 6. Results from the Mann-Whitney U statistical test

Mann-Whitney U test	Enter the number of datasets that you have produced or helped to produce during the past 5 years.	If you knew that your data would be preserved for the long-term in a secure environment, and shared only with accredited researchers, would you be willing to provide your data to a social science data archive?	In your view, how useful could be an institution that specializes in data archiving in your country?
How was this research financed? [International funding/project]	U = 1577.000 Wilcoxon W = 3347.000 Z = -2.205 Sig. (2-tailed) = .027 N = 128	/	/
How was this research financed? [Own funding (institution you are working in paid from its own funds)]	U = 1233.500 Wilcoxon W = 1936.500 Z = -2.380 Sig. (2-tailed) = .017 N = 128	/	/
What kind of data was kept? [Well documented with metadata]	/	U = 418.500 Wilcoxon W = 5368.500 Z = -2.592 Sig. (2-tailed) = .010 N = 113	/
How was this research financed? [Public funding from other sources (other ministries, state agencies, cities and municipalities...)]	/	U = 223.500 Wilcoxon W = 251.500 Z = -1.904 Sig. (2-tailed) = .057 N = 113	/
Who may be granted access to the data from your last project for research use? [Broader scientific community]	/	/	U = 1659.000 Wilcoxon W = 13905.000 Z = -1.966 Sig. (2-tailed) = .049 N = 181
What kind of data was kept? [Cleaned data (coded, anonymised, ...)]	/		U = 3311.500, Wilcoxon W = 9306.500 Z = -2.917 Sig. (2-tailed) = .004 N = 181

While researching *the importance of research data sharing*, the results revealed several important findings. There is a statistically significant relationship between researchers' attitudes regarding the importance of data sharing and the quantity of the data produced. Researchers who believe that data sharing is important also produced more data in the past five years ($r_s = .182$; $p = .040$). Researchers who are on the opinion that data sharing is important are considerably more willing to share ($r_s = .287$; $p = .002$); they think that their scientific work would benefit from a better access to research data produced in the country ($r_s = .307$; $p = .000$) and access to international research data ($r_s = .212$; $p = .004$); they also grade higher the benefit of a national institution for archiving research data ($r_s = .234$; $p = .002$), (Table 7).

Table 7. Results from the bivariate Spearman's rank-order correlation statistical test

Spearman's rho		With respect to your own discipline, how important is the sharing of research data?	In your view, how useful could be an institution that specializes in data archiving in your country?
Enter the number of datasets that you have produced or helped to produce during the past 5 years.	Corr. Coefficient	.182*	.112
	Sig. (2-tailed)	.040	.209
	N	128	128
If you knew that your data would be preserved for the long-term in a secure environment, and shared only with accredited researchers, would you be willing to provide your data to a social science data archive?	Corr. Coefficient	.287**	.161
	Sig. (2-tailed)	.002	.088
	N	113	113
Would your scientific work benefit if you had better access to research data produced in your country?	Corr. Coefficient	.307**	.161*
	Sig. (2-tailed)	.000	.031
	N	181	181
Would your scientific work benefit if you had better access to international research data?	Corr. Coefficient	.212**	.107
	Sig. (2-tailed)	.004	.152
	N	181	181
In your view, how useful could be an institution that specializes in data archiving in your country?	Corr. Coefficient	.234**	1.000
	Sig. (2-tailed)	.002	.
	N	181	181

*Correlation is significant at 0.05 level (2 - tailed)

**Correlation is significant at the 0.0level (2- tailed)

Regarding the *willingness to deposit data to a social science data archive*⁷, statistically significant differences were revealed in the scores of groups of researchers who belong to different types of institutions, $\chi^2(3) = 8.196$, $p = 0.042$, with a mean rank of willingness to deposit data of 36.00 for ‘University research institute’, 55.28 for ‘Higher education institution’, 66.00 for ‘NGO/Think tank’, and 61.19 for ‘Public research institute’. The post hoc test for examining the differences between the pairs showed marginal statistical significance (Bonferroni-adjusted) between the two groups, or more precisely, those who belong to university research institutes are more willing to deposit their data in a specialized archive in comparison to the NGO/Think tank ($p = 0.055$), (Table 8).

Table 8. Results from the Kruskal-Wallis H statistical test

Kruskal-Wallis H test	If you knew that your data would be preserved for the long-term in a secure environment, and shared only with accredited researchers, would you be willing to provide your data to a social science data archive?	In your view, how useful could be an institution that specializes in data archiving in your country?
With what type of institution are you currently principally affiliated?	Test Statistic = 8.196 Degree Of Freedom = 3 Asymptotic Sig.(2-sided test) = .042 N = 108	Test Statistic = 16.618 Degree Of Freedom = 4 Asymptotic Sig.(2-sided test) = .002 N = 167
Do you work in public, private or nongovernmental sector?	/	Test Statistic = 9.704 Degree Of Freedom = 2 Asymptotic Sig.(2-sided test) = .008 N = 180

Furthermore, a significant difference was found in the mean ranks ($U = 418.500$, $p = 0.010$) between respondents according to the format in which the data was kept, and regarding the willingness to deposit data to the social science data archive, in that respondents who keep data which were well

⁷ The question “If you knew that your data would be preserved for the long-term in a secure environment, and shared only with accredited researchers, would you be willing to provide your data to a social science data archive?”, was measured on a Likert scale from 1= “Yes, certainly”, to 5 = “No, certainly not”.

documented with metadata are considerably less willing to deposit data (mean rank =76.61) in comparison to those who did not save their data in this format (mean rank = 54.23), (Table 6).

There were no statistically significant differences in scores for the willingness to deposit data to the social science data archive between groups of researchers who used or did not use (No = 0, Yes = 1) different sources for financing their research. From this data, it can be concluded that researchers whose research was publically funded through national science funding bodies (science ministry, science foundation...) and public funding from other sources (other ministries, state agencies, cities and municipalities...), do not indicate a significantly higher willingness to deposit and share research data in comparison to researchers who used private sector funding. Only marginal statistical significance was confirmed among researchers who used public funding from other sources (other ministries, state agencies, cities and municipalities...), (mean rank = 35.93) who were more willing to deposit and share data, and those financed in other manners (mean rank = 58.39), expressed through a larger willingness to deposit and share research data ($U = 223.500$, $p = .057$), (Table 6).

*Researchers' perception on the usefulness of a national institution specialized in data archiving*⁸ is significantly positively related to the perceived benefits of free access to research data produced in the country ($r_s = .161$; $p = .031$), (Table 7). We found a statistically significant difference in respondents' scores on the usefulness of a national data archive among researchers affiliated to different types of institutions, $\chi^2(4) = 16.618$, $p = 0.002$, with a mean rank of the perceived importance of 80.70 for researchers who belong to a 'higher education institution', 71.00 for 'university research institute', 71.00 for 'public research institute', 94.81 for 'NGO/Think tank', and 119.60 for 'currently unemployed', (Table 8). The post hoc test for examining the differences among the pairs showed a statistically significant difference (Bonferroni-adjusted) among respondents who belong to a 'university research institute', who assessed the usefulness of such an archive higher than those in the 'currently unemployed' group ($p = 0.015$). In addition, researchers who belong to a 'university research institute' highly assessed the usefulness of an archive than those belonging to 'NGO/Think tank', but with a marginal statistical significance ($p = 0.052$). Almost the same marginal statistical significance ($p = 0.053$) was found among researchers belonging to a 'higher education institution', who have a more positive opinion on the

⁸ The question "In your view, how useful could be an institution that specializes in data archiving in your country?" was measured on a Likert scale from 1= "Very important", to 4 = "Not at all important".

benefits of such an archive than those who were unemployed. Furthermore, according to researchers' perception on the usefulness of a data archive, statistically significant differences were found in the scores among the groups of researchers employed in different sectors, $\chi^2(2) = 9.704$, $p = 0.008$, with a mean rank of the usefulness of 85.63 for the 'public sector', 87.28 for the 'private sector', and 103.04 for the 'nongovernmental sector'. The post hoc test for examining the differences among the pairs showed a statistical significance (Bonferroni-adjusted) among employees in the public sector and those in the nongovernmental sector, more precisely, employees in the public sector assess highly the usefulness of a specialized archive than those employed in the nongovernmental sector, $p = 0.008$. There is a significant difference among respondents according to who has access to their last project, in relation to the usefulness of such an archive ($U = 1659.000$, $p = 0.049$), in that researchers whose data is already accessed by the broader research community assess lower the usefulness of a data archive (mean rank = 102.64) in comparison to those who don't share this opinion (mean rank = 89.13). Finally, results showed that there was a significant difference in the mean ranks ($U = 3311.500$, $p = 0.004$) between respondents according to the format in which the data was saved, in relation to the usefulness of a national institution specialized for archiving, in that respondents who saved cleaned data (coded, anonymised, ...) assess lower the benefits of a data archive (mean rank = 99.51) in comparison to respondents who did not save their data in such a format (mean rank = 85.38), (Table 6).

DISCUSSION

The descriptive analysis of our data indicates that at present, a significant amount of data is produced by social science researchers in the Republic of North Macedonia. However, considering that few of them receive public funds from a centralized body for research and science, there is a possibility that these data will not be publicly shared or that they might even get lost. Considering that most of the research is financed by international sources, public access will depend on donors' rules and behaviour. Although at present we lack more specific information in this respect, the responses to several other questions indicate that it might be possible to expect a significant part of these data, as well as data created in future, to become public. This was revealed from researchers' attitudes towards data sharing. Similar to the research of Kim and Adler (2015), Tenopir *et al* (2011), Fecher *et al* (2015), 80.1% of researchers believe that data sharing in their scientific discipline is very important, while for 91.7% access to domestic and international data is a big benefit to their scientific work. If an accredited data archive did exist, 38.9% of the respondents would deposit their data, while additional 40%

would probably deposit. Information regarding the current and ideal level of access is also illustrative (Table 5). There is an evident possibility for a change in the existing modes of dominant data sharing among researchers (within their research teams and institutions), towards sharing of these data with the broader research community and a completely open access.

According to the analysed responses to the open-ended question regarding the barriers that might prevent researchers from sharing data (N=33), most researchers, similarly to most conclusions in the literature, are worried from a possible abuse of their data, inadequate interpretation, improper referencing, or even plagiarism. Some stated that the great effort invested in the data collection, and sometimes even the investment of personal finances, avert them from simply sharing data with others. Researchers also stress that some data is confidential and sensitive. One respondent mentioned the lack of legal regulations. Contrary to the findings of Fecher *et al*, 2015 and Tenopir *et al*, 2011, none of the researchers in this sample mentioned lack of time as a reason not to share data.

At present, only 12% of the research data is kept in a data repository, while documenting the data by researchers according to the usual and internationally accepted documentation standards (important for international data sharing as well) is almost inexistent. There is obviously a need to educate researchers on research data documentation, as well as a more systematic approach to data preservation at an institutional level.

The conducted bivariate data analysis revealed several interesting findings. Productivity of researchers' scientific work, measured by the number of produced datasets in the past five years, differs according to the financing manner. Findings suggest that researchers relying on international funding are more productive in producing research data than researchers who provide their budget from other sources. On the other hand, researchers who use funds from the institutions that employ them have lower data production in comparison to other means of financing. This might be the case because of the limited amount of funds on researchers' disposal. Institutional financing of research in a developing economy cannot substitute public financing of research, which should be the primary manner of financing of research in each country.

Regarding the importance of research data sharing, the results revealed significant findings. Increased production of research data increase the importance of data sharing among researchers. In addition, it was confirmed that the willingness to share is positively related to the intention of sharing, and, as an attitude, it can be a motivator for social scientists' data sharing behaviours. This finding is in the same line as the findings from Kim and Adler's study (2015). Researchers connect the importance of sharing with the perceived benefit to their scientific work, i.e. the more data is shared, the

bigger the personal, professional and scientific benefit from greater availability of international and national research data.

The analysis of the willingness for depositing and sharing research data with accredited researchers, provided there is a national social science data archive which, in the long-term and in a safe environment, would preserve data, indicate that researchers' willingness to deposit and share data is related to the manner in which researchers manage (structure and document) data during the research process. Bigger willingness was observed among researchers who have saved data which is well documented with metadata.

Finally, the perceived benefit from a national institution specialized in data archiving is related with the researchers' benefits from free access to research data. The perception for the benefits differs among researchers employed in different sectors, and is significantly more evident among researchers employed in the public sector contrary to those employed in the nongovernmental sector. In addition, findings suggest that researchers who, through different methods, have already provided access to the data from their last research, assess as lower the importance of a specialized data service that would act as a mediator between data producers and data users.

CONCLUSION

The study aimed to examine attitudes and practices of social science researchers regarding research data sharing, as well as their attitudes on the benefits of a specialized archive for data preservation and dissemination to the broader scientific community. Findings suggest that, generally, researchers support data sharing practices, taking into account their individual professional development, as well as the scientific development as a whole. What is lacking and is challenging is a more systematic and holistic approach in managing research data deposition and dissemination processes.

Firstly, on a national and institutional level, there is a need to regulate the processes of research data sharing. The ministry responsible for science policy should address these issues, engaging in creating proper public policies and regulations for sharing scientific data in the public domain, and support the operability of the national data archive which would generate a broader social and economic value. In addition, it is necessary to develop preservation and dissemination standards for digital research data, considering personal data protection and researchers' integrity.

Secondly, considering that researchers need to invest an additional effort, time and finances in data processing and preparation of data for deposition, broader engagement of researchers in data sharing practices would be achieved

through: a) creating policies and practices for stimulating data sharing through different incentives, based on individual recognition and rewards for data sharing and data publishing, including benefits to researchers' professional promotion and advancement, and b) comprehensive skill development training for data management (including preparation of data management plan), creation of standardized metadata documents, and instructions for preservation of raw data in standardized digital formats.

REFERENCES

- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhler, P., and Wouters, P. (2004). Promoting Access to Public Data for Scientific, Economic, and Social Development. *Data Science Journal*. Vol. 3, pp. 135-152. Retrieved from:
- Borgman, C., L. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 63(6), pp. 1059-1078. Retrieved from: <https://doi.org/10.1002/asi.22634>
- Cekik, A., Babunski, K., and Zabijakin-Chatleska, V. (2015). Report on the evaluation of research and legal conditions for the establishment of a social science data archive in Macedonia, *SEEDS project*, Retrieved from:
- Fecher B., Friesike S., Hebing, M. (2015) What Drives Academic Data Sharing? *PLoS ONE* 10 (2):e0118053.
<https://doi:10.1371/journal.pone.0118053>
- Heers M., Ferez, E., & De Paula, E.M. (2017) Data sharing and re-use: Researcher's practices, attitudes and needs, FORS survey of social science researches in Switzerland.
<https://api.semanticscholar.org/CorpusID:210590836>
<https://doi.org/10.1016/j.ijinfomgt.2015.04.007>
<https://doi.org/10.1371/journal.pone.0021101>
https://seedsproject.ch/wp-content/uploads/2015/06/Report-Macedonia_merged1.pdf
https://www.jstage.jst.go.jp/article/dsj/3/0/3_0_135/_pdf
<https://www.oecd.org/sti/inno/38500813.pdf>
- Kim, Y. and Adler, M. (2015). Social Scientists' Data Sharing Behaviors: Investigating the Roles of individual motivations, Institutional Pressures, and Data Repositories. *International Journal of Information Management*. Vol. 35, pp. 408-418. Retrieved from:

- Le Guillou, Ian (2020) *Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research*, Horizon 2020 magazine, 20 March 2020, available at:https://horizon-magazine.eu/article/covid-19-how-unprecedented-data-sharing-has-led-faster-ever-outbreak-research.html?fbclid=IwAR0oZj6Y_DCAjwrjLrOB5aNkUZm8n5fur_O7tc5sXc3ARVVF-cngiUtssXI#utm_source=Facebook&utm_medium=share&utm_campaign=covid-19
- Mauthner, N. S. (2018). Data Sharing Methods. *The Blackwell Encyclopedia of Sociology*. Edited by G. Ritzer and C. Rojek. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405165518.wbeos1103>
- OECD, (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD Publishing. Retrieved from:
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., et al. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6(6): e21101.
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B. et al. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE* 10(8): e0134826. <https://dx.doi.org/10.1371/journal.pone.0134826>

