RESEARCH DATA IN THE CONTEXT OF OPEN SCIENCE: CASE STUDY OF THE REPUBLIC OF MOLDOVA

Andrei Rusu[1], Nelly Turcan[2],
Rodica Cujba[3] and Mihai Grecu[4]

DOI: 10.24989/ocg.v335.26

Abstract
At present, there are important changes in scientific research, changes that concern both the organization, realization and evaluation of research, as well as the use and dissemination of scientific results. These changes that are characteristic to the Open Science phenomenon are determined by the development of new technologies, by increasing the social role of scientific research in an institutional and political context. The Open Science concept represents a new approach to the way in which scientific research based on cooperation and new ways of disseminating knowledge is carried out and organized, using new digital technologies, new tools for collaboration, and RDI infrastructure to ensure open access to research data.

The paper presents the results of the study on mapping research data in the Republic of Moldova. The study is based on results of two surveys conducted in January-February 2016 and May-July 2018. The research has highlighted the general concern of RDI actors about the retention and use of scientific data. A new challenge is needed to solve scientific data issues by creating new type of infrastructure to ensure data archiving, broad access for the purpose of their dissemination and reuse, and thus creating new research opportunities based on research data.

1. Introduction

Open Science is a challenging phenomenon that is emerging around the world. Open Science brings social, economic, cultural, political and technological change, based on openness and connectivity, on how research is designed, performed, used, assessed, and preserved. Open access platforms, open infrastructures, open data tools, open educational resources, open evaluation methods, open collaboration, or open citizen science activities are irreversible trends that are impacting all

[1] Institute of Mathematics and Computer Science „Vladimir Andrunachievici", 5 Academiei street, MD-2028, Chisinau, Republic of Moldova, andrei.rusu@math.md, http://www.math.md/, Ovidius University of Constanta, bd. Mamaia 124, Constanta, Romania, agrusu@univ-ovidius.ro, http://math.univ-ovidius.ro/, https://orcid.org/0000-0002-0259-3060
[2] Moldovan State University, 60 Alexe Mateevici street, MD-2009, Chisinau, Republic of Moldova,
Information Society Development Institute, 5A Academiei street, MD-2028, Chisinau, Republic of Moldova, tsurcannelly@gmail.com, http://usm.md/, https://orcid.org/0000-0001-8475-0770
[3] Information Society Development Institute, 5A Academiei street, MD-2028, Chisinau, Republic of Moldova, rodica.cujba@idsi.md, https://idsi.md/, https://orcid.org/0000-0002-7982-6184
Technical University of Moldova, 168, Stefan cel Mare Blvd., MD-2004, Chisinau, Republic of Moldova, rodica.cujba@adm.utm.md, https://utm.md/
[4] Information Society Development Institute, 5A Academiei street, MD-2028, Chisinau, Republic of Moldova, mihai.grecu@idsi.md, https://idsi.md/, https://orcid.org/0000-0003-0283-6648

scientific actors and have the potential to accelerate the research cycle [14]. By increasing access to publications and data, universities, research institutions, firms and individuals may use and re-use scientific outputs to produce new products and services.

One of the key elements of open science is open access to research data. In a research context, there is a growing opinion that most research data should be open, particularly data from publicly funded projects. This point of view is driven especially by research funder requirements for sharing and re-use data, upon principles regarding research data such as to be findable, accessible, interoperable and re-usable (FAIR principles). The research data are made open for two purposes: to provide evidence that the research was conducted properly and to provide data for reuse and the generation of further findings and outputs [3].

Open research data is research data that combines the characteristics of open data and the types of research data [8].

As World Bank mentioned in his report [16, ch.1] data is considered the new gold, or the new oil, and like oil, unprocessed data has relatively little value and needs to be mined, refined, stored, and sold on to create value.

According to the Engineering and Physical Sciences Research Council [7] "**Research data** is defined as **recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings**; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created".

Research data can be generated for different purposes and through different processes [10]:

- Observational: data captured in real-time, usually irreplaceable. For example, sensor data, survey data, sample data, neuroimages.

- Experimental: data from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data.

- Simulation: data generated from test models where model and metadata are more important than output data. For example, climate models, economic models.

- Derived or compiled: data is reproducible but expensive. For example, text and data mining, compiled database, 3D models.

- Reference or canonical: a (static or organic) conglomeration or collection of smaller (peer reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals.

Scientists are particularly interested in data collection, and the success of each experiment is determined by the new data generated, which can contribute to advancing scientific knowledge. Any scientific research involves performing an observation, generating a hypothesis, running an experiment, and collecting data. Traditionally, for any research, the amount of data collected by scientists was not very extensive, and its analysis did not require the use of technology. Previously, for scientists, technology was used in a very limited way, and data evaluation was not done using algorithms or software. However, significant changes have taken place over the past two decades,

and changes in software and tools have made data acquisition and analysis a very important part of research.

At present, scientists and research are subject to a paradigm shift [1, 2, 15]. Various developments in data-based statistical software, tools and science, such as computational biology and computational chemistry, have led to a new generation of scientists focusing on the analysis and interpretation of the data obtained. Research projects, such as the Large Hadron Collider, the Hubble Telescope, and the Human Genome Project, are evidence of how science has become dependent on data computation and extraction. Thus, the outcome of various advances in technology has transformed the field of science. Scientists now have the ability to perform high performance experiments, which are basically intensive data projects that allow researchers to accumulate and store huge amounts of data.

In late years some studies have been conducted that were focused on researchers' attitude toward openness, access, sharing and re-use of research data among others [5, 9, 11, 12].

Thus, the Open Data survey [9] found that the attitude toward data sharing is generally positive, but open data is not yet a reality for most researchers. A global online survey of 1,200 researchers found that many perceive data as personally owned. Public data sharing primarily occurs through the current publishing system; less than 15% of researchers share data in a data repository. At the same time, 69% of survey respondents said sharing research data is important for doing research in their field. The survey also revealed that when researchers share their data directly, most of them (>80%) share with direct collaborators. Although data sharing seems to have a global benefit, cultural and national factors pose a significant challenge to a one-size-fits-all approach. Regardless of the benefits, deciding what kind of data can be shared, how it should be shared, and making it usable by others requires additional effort, training, and resources. Furthermore, freeing up data for reuse and sharing depends on accommodation or coordination of disciplinary, cultural, and local differences with respect to data privacy and licensing.

In another large surveys on research data [11], have found widespread data sharing associated with published works and a desire from researchers that their data are discoverable. The survey confirms and extends recent findings on general data sharing attitude and behaviour. Thus, 76% of researchers rated the importance of making their data discoverable highly – with an average rating of 7.3 out of 10 and the most popular rating being 10 out of 10 (25%). The main challenge to data sharing was identified by respondents as „Organising data in a presentable and useful way" (46%), with other challenges generally rated: „Unsure about copyright and licensing" - 37%; „Not knowing which repository to use" - 33%; „Lack of time to deposit data" - 26%; „Costs of sharing data" - 19%.

Figshare's annual report, *The State of Open Data Report 2018* [6], looks at global attitude toward open data. The report is the third in the series and the survey results continue to show encouraging progress that open data is becoming more embedded in the research community. The key finding is that open data has become more embedded in the research community – 64% of survey respondents reveal they made their data openly available in 2018. The percentage of respondents in support of national mandates for open data is higher at 63% than in 2017 (55%). 80% of respondents stated that they were aware of open data to reuse. However, a surprising number of respondents (60%) had never heard of the FAIR principles, a guideline to enhance the reusability of academic data.

In order to map the situation regarding generation, gathering, use, sharing and preservation of research data obtained within research projects carried out in the Republic of Moldova, Information Society

Development Institute conducted two surveys: in January-February 2016 and May-July 2018. Some of the surveys' results are presented in the book *Open Science in the Republic of Moldova* [4].

Thus, the *goal* of the present paper is to investigate how data is used and managed by the research community of the Republic of Moldova and how the use and management of research data evolve in time in the Moldovan academia.

2. Methodology

The study is based on responses of two surveys.

The first survey was conducted under the context of the Research and Development Strategy of the Republic of Moldova 2020 (Gov. Dec. no 920 of 07.11.2014) and National Strategy for information society development "Digital Moldova 2020" (Gov. Dec. no 857 of 31.10.2013), within the national project *The pilot platform for quality assurance and visualization of digital scientific content in the Republic of Moldova* (SCIFORM) (https://idsi.md/en/sciform) [13].

The survey's *main objectives* were:

- to map the existing digital scientific content;
- to assess the needs for the transposition of national scientific content in digital format;
- to find out the opinion of Moldovan academia vis-à-vis openness of publications and research data;
- to catch the attitude of leadership of research institutions, project managers and journals' editorial boards regarding open access policies.

Responses were collected from 39 representatives of R&D institutions (75% success rate); 48 editions of scientific journals (success rate: 63%); 83 managers of national research projects (success rate: 34%); 23 libraries (success rate: 71%).

The second survey on the mapping of the research data ecosystem in the Republic of Moldova was carried out within the framework of the project *Elaboration of conceptual and methodological framework for e-Infrastructure of data in the RDI field of the Republic of Moldova* (e-IDSM) (https://idsi.md/md/e-idsm). Unlike the previous survey this one *was focused* exclusively on research data.

The *main goal* of this study was to identify the needs of the RDI community in the Republic of Moldova on the management of scientific data over their lifecycle (creation, processing / analysis, storage / preservation, sharing / access and use). The specific objectives of the survey were:

- to identify the types / formats and sources of research data;
- to find out the modes of storing and preservation of the research data;
- to discover the ways the research data are processed and analyzed;

- to learn the procedures of the research data management;
- to determine methods of sharing, access and use of the research data.

Responses were collected from 48 RDI institutions (92% success rate), including 13 higher education institutions. Respondents with various positions within these institutions participated in the survey, including: 25 heads of RDI institutions (12.3%); 42 project managers (20.7%), 65 laboratory / research group managers (32%), 34 scientists (16.7%), 23 university teachers (11.3%), 4 PhD students (2%), other positions (10 - 5%).

The authors were members of the research projects' teams and have participated in the design of the surveys, collection and aggregation of the results.

3. Results and Discussions

Both surveys highlighted the need to improve the circulation of knowledge and access to research data. Thus, the respondents of the first survey totally or partially agreed that data obtained from research funded exclusively from public funds as well as those that are partly funded from public and partly from private funds should be available for reuse and free on the Internet (figure 1).

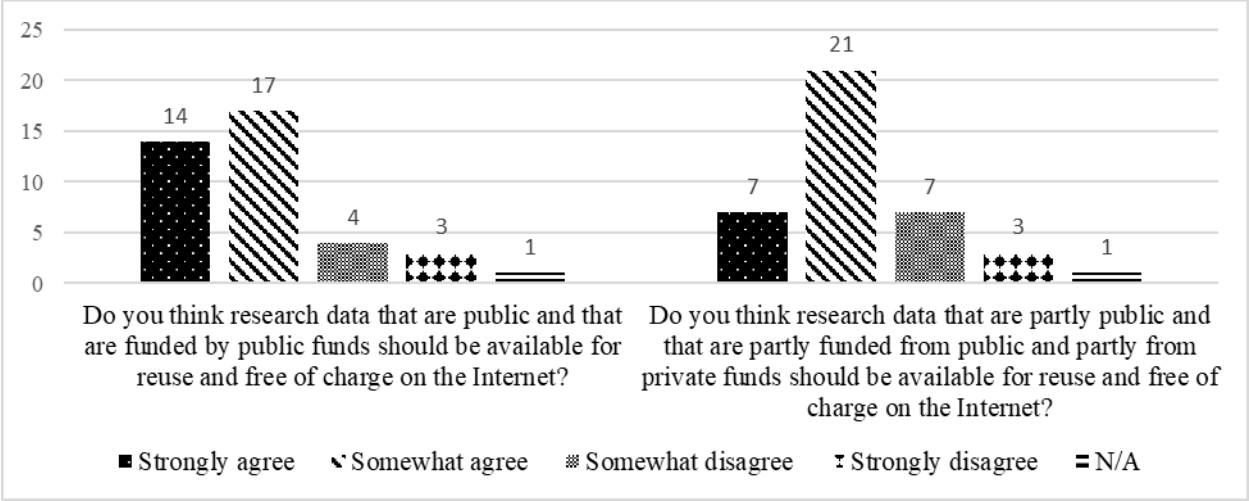

Figure 1. RDI institutions' opinion on open access to research data funded exclusively or partly from public funds

The comparative analysis of both surveys' answers shows that the rate of respondents who unconditionally accept open access to research data obtained from public funds decreased from 50% in 2016 to 28% in 2018. At the same time, those who accept conditional open access increased twice, from 23% to 45%. However, the rate of those who do not support open access to research data did not change significantly (10% in the first survey and 12.4% in the second survey) (figure 2).

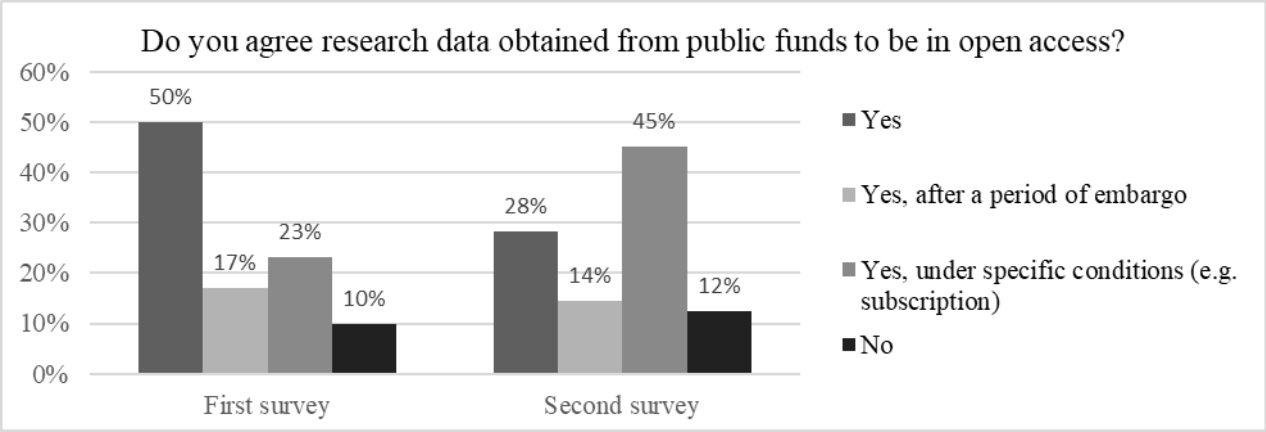

Figure 2. Moldovan academia's opinion on open access to research data obtained from public funds

At the same time, 2018' survey participants have specified that research data must be accessible to colleagues, scientific community, PhD students, decision-makers, educational institutions and other users, and one of the primary conditions for using research data is to cite the source. Also, it was mentioned that some data banks are not accessible to the public, and researchers do not have sufficient skills in managing research data.

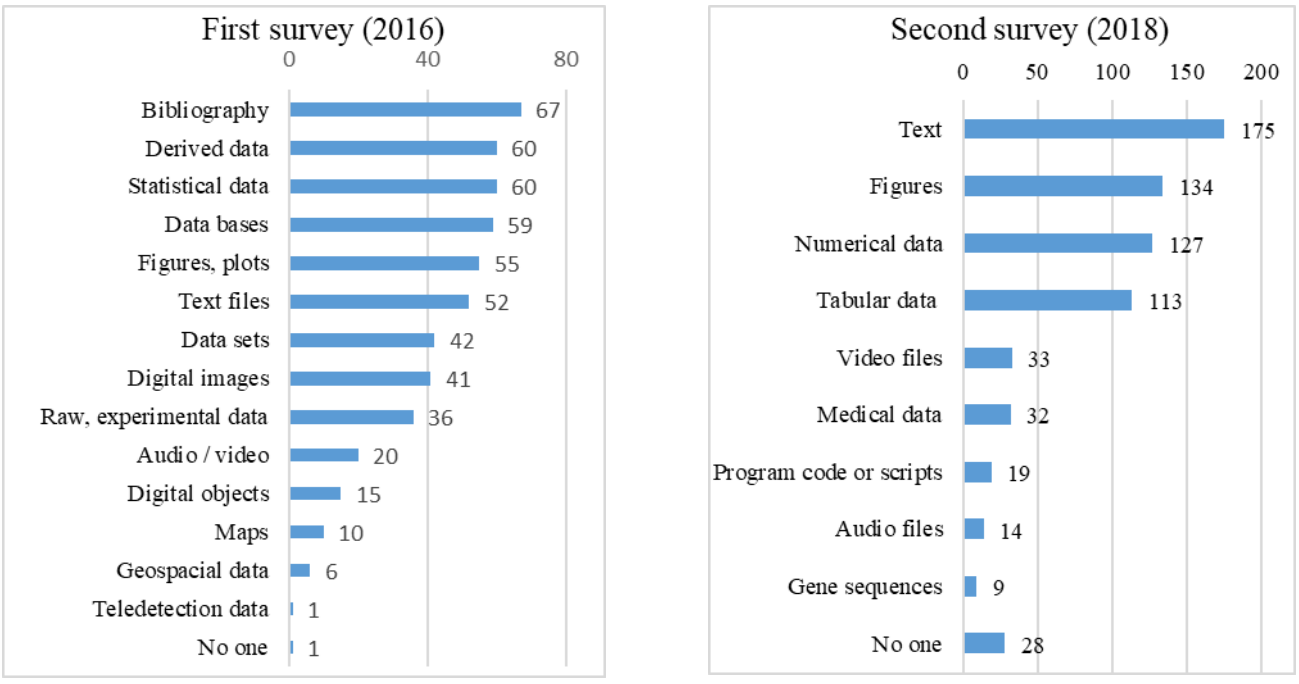

Figure 3. Types of digital data produced / generated for research

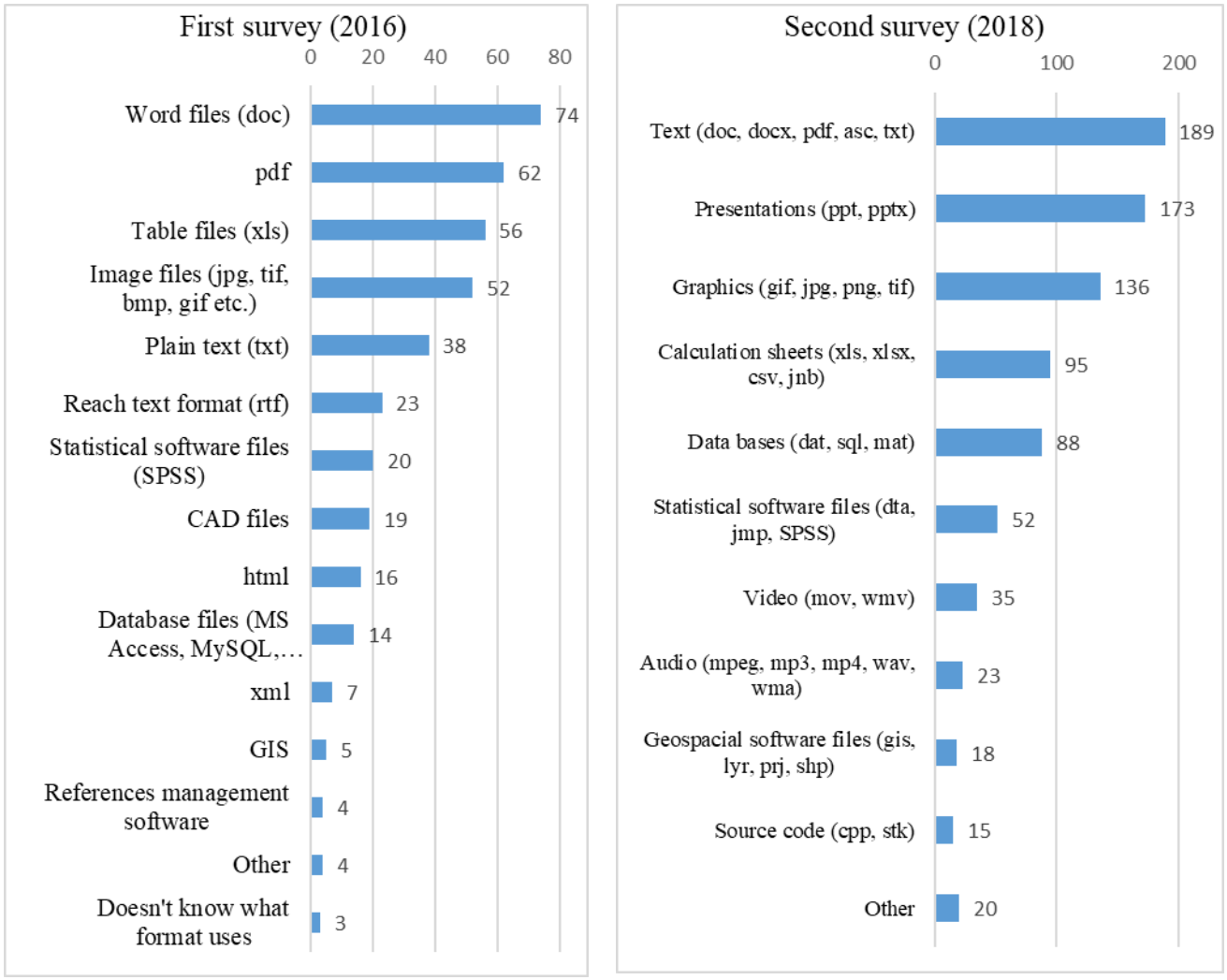

Figure 4. Types of generated / collected data formats

Regarding types of research data produced by Moldovan academia, answers obtained within two surveys show that in 2016 more bibliographic and derived data were produced in comparison with 2018 when more figures and numerical data prevailed (figure 3).

The data obtained from surveys regarding formats of data produced for research, revealed that the most popular formats are texts, tables, graphics, presentations (figure 4).

Taking into account data formats used by researchers, as well as the needs described by them in the second survey, Moldovan researchers can be divided into two categories:

- Researchers using relatively widespread software tools in the academic and research environment, such as: Microsoft Office, SPSS, Adobe FineReader, and others.

- Researchers using research-specific software solutions such as: ArGIS, Geoportal, Mathematica, FoxPro, Endnote, 1C, EViews, GAMESS, Gaussian09 and others.

For data management, it is important to have Data Management Plans for institutional data management policies or procedures as well as for research projects. The results of the second survey

revealed that 107 respondents, which make up more than 50% of the survey participants, do not know or believe that there are no institutional policies and procedures regarding the management of research data (Figure 5). With regard to the development of the data management plan for research projects, only 21 respondents (10.3%) stated that funding agencies had requested such a plan (Figure 6).

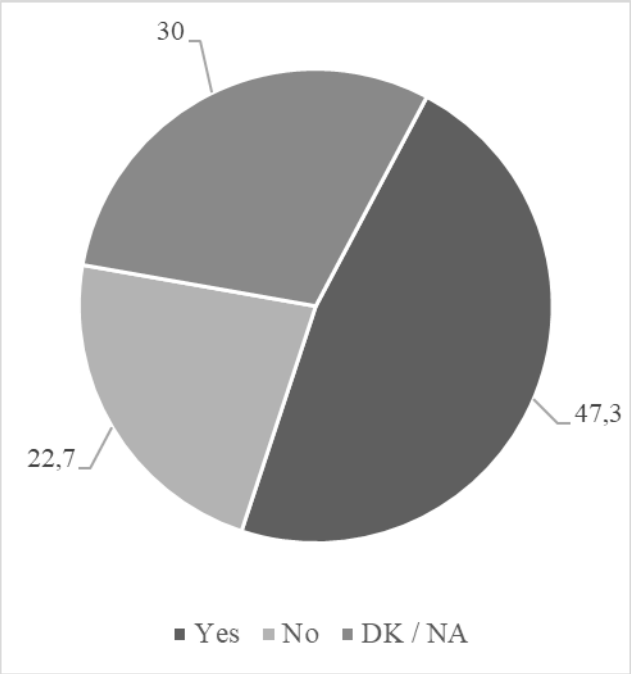

Figure 5. Existence of institutional management data policies and procedures

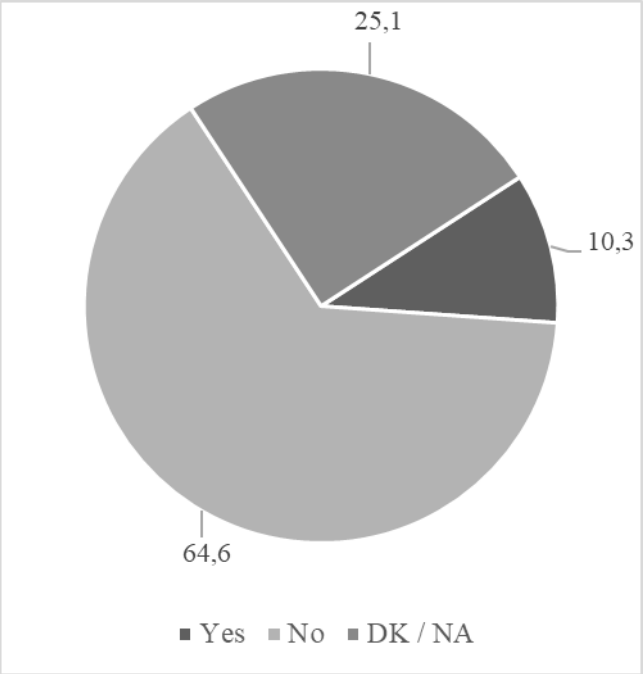

Figure 6. Request from financial agencies of Data Management Plan

Although only 96 respondents (47.3%) know about the existence of institutional data management policies and procedures, they noted that there are various policies and procedures in place to protect, store, archive, share research data, among which: privacy policy, data storage policy, institutional policy on open access, institutional policy on intellectual property and technology transfer, primary data verification policy, old data removal policy, strict journaling of records, experiments' and tests' registries, contracts with organizations, non-disclosure agreements, internal networks specifying data access rights, etc.

However, it should be noted that the majority of survey participants (170 respondents – 83.7%) believes that training on research data management is needed. Respondents emphasized the necessity of training researchers from different fields on research data management technologies.

4. Conclusions and Recommendations

In conclusion, we can mention that Moldovan academia is ready to provide access to research data. Most of researchers use digital media to access research data, but prefer to use traditional data storage formats (Word, Excel, PDF, etc.). Only some researchers use modern and innovative tools to process, access, store and archive research data. One of the main issues that discourages research data sharing is the issue of copyright protection. There are some concerns about the loss of property rights and copyright infringement in case of sharing and open access to research data. At the same time, in RDI institutions of the Republic of Moldova, the management of research data management

is not implemented. There are problems related to long-term preservation, storage, sharing and access to research data.

Concluding results of these studies, the following recommendations can be made:

- Continuous analysis of the tendencies in research data management.
- Analysis of the international experience in the field of research data management.
- Establishment and approval of policies on research data management in research projects and / or research institutions.
- Training and familiarization of Moldovan academia in different fields of research data management.
- Training researchers in the Republic of Moldova on open source solutions that could be used in the research process as well as in the special case of research data management.
- Establishing rules / procedures / customs for research data management to be known to researchers, to be adopted by researchers and addressing all stages of research data management.
- Organization of round tables or other activities for presentation of solutions used in the field, as well as for exchange of views and experience in the field of research data management.

Acknowlegments. The study was performed due to the support of two research projects: 15.817.06.13A "The pilot platform for quality assurance and visualization of digital scientific content in the Republic of Moldova (SCIFORM)" (2015-2019) funded by the Academy of Sciences of Moldova (ASM) and the National Agency for Research and Development of the Republic of Moldova (NARDRM); 18.50.07.10A/PS "Elaboration of conceptual and methodological framework for e-Infrastructure of data in the RDI field of the Republic of Moldova (E-IDSM)" (2018-2019) funded by the NARDRM.

5. References

[1] ADAMS, J., Impact of Open Science methods and practices on the economics of research and science. Case Studies from Life, Mathematical and Social Sciences, Luxembourg, Publications Office of the European Union, 2015, available: doi: 10.2777/14 [accessed 30.12.2018].

[2] CHANGHUI, P. et al. Towards a paradigm for open and free sharing of scientific data on global change science in China, in: *Ecosystem Health and Sustainability*, 2016, **2**(5), available: doi: 10.1002/ehs2.1225 [accessed 30.12.2018].

[3] CHILDS, S., MCLEOD, J., LOMAS, E., COOK, G., Opening research data: issues and opportunities, in: *Records Management Journal*, 2014, **24**(2), 142-162, available: doi: 10.1108/RMJ-01-2014-0005 [accessed 02.01.2019].

[4] CUCIUREANU, G. et al. Open Science in the Republic of Moldova, Chisinau, IDSI, 2018, available: doi: 10.5281/zenodo.146841 [accessed 30.12.2018].

[5] Digital Science. The State of Open Data Report 2017 [online]. Digital Science, 2017, available: doi: 10.6084/m9.figshare.5481187 [accessed 02.01.2019].

[6] Digital Science: The State of Open Data Report 2018, Digital Science, 2018, available: doi: 10.6084/m9.gshare.7195058 [accessed 02.01.2019].

[7] Engineering and Physical Sciences Research Council, EPSRC policy framework on research data - Scope and benefits, available: https://epsrc.ukri.org/about/standards/researchdata/scope/ [accessed 30.12.2018].

[8] ONYANCHA, O. B., Open Research Data in Sub-Saharan Africa: A Bibliometric Study Using the Data Citation Index, in: *Publishing Research Quarterly*, 2016, **32**(3), 227-246, available: doi: 10.1007/s12109-016-9463-6 [accessed 30.12.2018].

[9] Open Data. The Research Perspective, CWTS and Elsevier, April 2017, available: https://www.elsevier.com/about/open-science/research-data/open-data-report [accessed 02.01. 2019].

[10] Research Information Network classification. University of Leicester, Research Data – Definitions, available: https://www2.le.ac.uk/services/research-data/documents/UoL_Reserch DataDefinitions_20120904.pdf [accessed 30.12.2018].

[11] STUART, D. et al. Practical challenges for researchers in data sharing: Whitepaper, Springer Nature survey data, 2017, available: doi: 10.6084/m9.figshare.5975011 [accessed 02.01.2019].

[12] TENOPIR, C. et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide, in: *PLoS ONE*, 2015, **10**(8), available: doi: 10.1371/journal.pone.0134826 [accessed 02.01.2019].

[13] ȚURCAN, N., CUJBA, R., Open Access Policy to research outputs in the Republic of Moldova. State of the art and perspectives, in: Hendri Hanson et al. (eds), CEE e|Dem and e|Gov Days 2017: Digital Divide in the Danube Region: Is it still significant in explaining ICT adoption in eDemogracy and eGovernment? Budapest (OCG Conference Proceedings), 4-5 May 2017, Budapest, 2017, pp. 283-293.

[14] VICENTE-SAEZ, R., MARTINEZ-FUENTES, C., Open Science now: A systematic literature review for an integrated definition, in: *Journal of Business Research*, 2018, 88, 428–436, available: doi: 10.1016/j.jbusres.2017.12.043 [accessed 30.12.2018].

[15] What is Open Science? Research Center for Open Science and Data Platform (RCOS), National Institute of Informatics, Tokyo, available: https://rcos.nii.ac.jp/en/openscience/ [accessed 30.12.2018].

[16] World Bank, Information and Communications for Development 2018: Data-Driven Development, 2018, available: doi: 10.1596/978-1-4648-1325-2 [accessed 02.01.2019].