

Legal Issues in Managing and Sharing Research Data

Jelena Banović¹, Aleksandra Bradić-Martinović²

^{1,2}Institute of Economic Sciences-Data Centre Serbia for Social Sciences, Belgrade, Serbia

¹jelena.banovic@ien.bg.ac.rs, ²abmartinovic@ien.bg.ac.rs

Abstract—Scientific research is directly related to the collected and processed data. Data manipulated within the natural and technical sciences are not subject to the same restrictions as data in the social sciences. Domestic researchers who collect data from individuals must be aware of the limits imposed by the Law on Personal Data Protection of the Republic of Serbia. This topic is particularly relevant to the Open Science and Open Access principle movement, for which the Ministry of Education, Science and Technological Development of the Republic of Serbia has dedicated itself through the adoption of the Open Science Platform in 2018. The paper aims to analyse and brings closer to researchers the restrictions within the two most important laws that are directly or indirectly implemented in our country - the Law on Personal Data Protection of the RS and GDPR, which is applied in the EU. Domestic researchers also must respect them in projects funded by EC or other EU financiers. The topic is covered through the presentation of the Data Center of Serbia for Social Sciences case study, a national infrastructure that deals with collecting, archiving, and sharing research data in the social sciences.

Keywords - data, open science, open access, legal framework, data centre serbia for social sciences

I. INTRODUCTION

Data has always been a valuable resource, and efforts have been made to systematise and preserve it appropriately. With the development of digitalisation, they have become an integral part of all areas of human existence. Data sources can be different, and they also include personal data about one's identity, property, social and health status, but also data on business subjects, their balance sheets, clients, data on national and

supranational levels, open or protected, as well as data on our environment and much more.

However, a particular dimension is research data in science. As part of scientific exploration, researchers collect various scales and types of data. Some sciences, especially those which use measuring instruments, generate a massive amount of data, classified as Big Data. Based on the various data, science draws significant conclusions and achieves excellent discoveries, and very often, cooperation is the key to progress. A movement that advocated open communication among scientists appeared to open the possibility of the broadest scientific cooperation. OECD observes Open Science as a process that “encompasses unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives” [1]. Despite the benefits that this movement has already brought to science and scientists, many issues still exist that prevent its spread and strengthening. Bearing in mind that this paper focuses on open research data, our goal is to point out a significant dimension of this problem - legal issues and considerations in the process of data management and sharing, with a focus on social science. Many researchers, especially in Serbia, are unaware of data's determinants - they are not informed about rights or obligations toward data, legally and ethically. A unique dimension of the problem can be seen in the social sciences that collect, process, deposit and share data on individuals. This field is subject to additional laws that do not affect data management and sharing within the natural and technical sciences.

Serbian researchers should realize that the future situation can only be more complicated. The Ministry of Education, Science and

Technological Development of the Republic of Serbia is increasingly putting pressure on implementing Open Science following the Open Science Platform adopted in 2018. In the first phase, it is insisted on implementing digital repositories in which all scientific publications will be stored and publicly available. In the next step, the same treatment is expected for research data.

With an aim to bring researchers as close as possible to the limitations in data management and sharing, together with their rights and possibilities, we will present elements upon the main document of Serbian national research infrastructure Data Center Serbia for Social Sciences - Policy and Procedures document is based.

II. DATA CENTRE SERBIA FOR SOCIAL SCIENCES

Data Center Serbia for Social Sciences (DCS) [2] is the national data centre of the Republic of Serbia established in 2014 as an organisational unit of the Institute of Economic Sciences in Belgrade to identify significant data sources within a wide range of disciplines of social sciences, collect and prepare data for the long-term preservation and enable dissemination for scientific research and academic education [3]. The Center has been developed since 2012 with the support of the European Research Community and international projects funded by the European Commission [4] and the HORIZON2020 project - CESSDA SaW (Strengthening and Widening) [5], as well as by the Swiss National Science Foundation, SCOPES project - SEEDS (South-Eastern European Data Services) [6] DCS became a national research infrastructure in 2019 when the Ministry joined the Consortium of European Social Science Digital Archives – CESSDA ERIC [7] and assigned the service provider's role to the Institute of Economic Sciences. CESSDA is a leader in data curation and sharing with great expertise. CESSDA ERIC supported DCS from the first day until the present time.

Today, DCS is available to all Serbian researchers in the field of social sciences and fulfils its mission through activities of collecting, depositing, and sharing research data, representing Serbia in the European Research Area (ERA), training researchers in this field, and participating in the Open Science Team of the Republic of Serbia.

III. LEGAL ISSUES IN DEALING WITH RESEARCH DATA IN SOCIAL SCIENCES

A. Data Protection Laws

Legal restrictions are, in essence, highly formal and include penalties in case of non-compliance. Therefore, researchers must have clear guidelines regarding the data generated during the research process, sharing them among themselves or in Open Access data repositories such as Harvard Dataverse, Figshare, Zenodo, etc. However, the situation is very complex due to several problems. The first one is the lack of a unique definition of the term - "research data". According to the website Law insider [8], there are over 50 different definitions, and the three central covers research data as "information or data which is not Personal Data that is collected or generated in the performance of the Research and includes (but is not limited to) information that is collated or stored in the searchable form". Other consider it as "the result of observations or experimentation that validate research findings and that are published separately to the Article, which can include but are not limited to raw data, processed data, software, algorithms, protocols, and methods." and the third explains it as "documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results". Based on only three definitions, it is clear to what extent the approach and the way of observing this vital research category differ. The ignorance and imprecise determinants of the basic concept are confusing and put researchers at risk.

Case study example. DCS collects, without a precise definition, quantitative and qualitative data in a wide range of social sciences and humanities (economics, education, employment and labour, ecology, conservation and land use, health, history, industry and management, law, criminology and legal systems, media, communication and languages, political science, psychology, sociology, society and culture, social welfare policy and systems). The data should be harmonised with one of the listed formats: spreadsheet and similar formats as the quantitative data and text and audio images/ photos and videos as the qualitative data.

Many projects implemented in Serbia are funded by the European Commission or other

European Union funds. Within these projects, personal data protection is predominantly based on the principles of the General Data Protection Regulation – GDPR, introduced in May 2018. It was not the first law on personal data in the EU, and the GDPR replaced the previous one – the Directive on the use of personal data in the EU, which was in charge for almost 20 years. The main reason for changes was technological development and adjusting regulations to a contemporary, digitised economy and society. Reference [9] offered the systematisation of the main GDPR principles regarding scientific research data:

- the processing of personal data must follow the law and be entirely transparent for the person(s) who made the data available to researchers. All respondents involved in scientific research have the right to know exactly how their data will be handled during and after the study. In cases where the survey is not anonymous (e.g. focus group or interview), researchers are obliged to obtain Consent from respondents. Respondents have the right to withdraw their Consent at any time without explanation.
- if there is the respondents' Consent, the researchers must not go beyond the respondents' agreement. Without Consent, researchers are obliged to use them only for scientific analysis. There are exceptions when data are taken over by the state archives, national statistics, etc.).
- to reduce the probability of misuse of the amount of data collected from the respondents, it should be minimal and fully harmonised with the goals of scientific research.

Special attention to the place and time of data storage is also necessary. Data must be stored safely (protected from malicious attacks or negligence) and must not be kept longer than the period required for research. However, archiving or long-term data storage is also desirable under the GDPR, and it requires protected infrastructures, such as trustworthy digital repositories.

Following the example of the European Union, Serbia has also improved its legislation and, in November 2018, adopted the Law on

Personal Data Protection of the Republic of Serbia [10]. The law is basically in line with the GDPR and has been in force since August 2019. Unlike the legal delimitation of the broad term “research data”, the domestic Law on Personal Data Protection very unambiguously defines the term “personal data” as “any data relating to a natural person whose identity is determined or identifiable, directly or indirectly, mainly based on an identity mark, such as name and identification number, location data, identifiers in electronic communications networks or one or more features of his physical, physiological, genetic, mental, economic, cultural and social identity” [10]. In addition to being aware of the definition, researchers also must be informed of the differences between direct and indirect identifiers. In case they overlook the indirect identifiers, revealing a private person's identity is possible, which carries additional risks (more on this topic [11]). Personal data protection is critical when sensitive data is involved. According to the law, this category includes racial or ethnic origin, religious affiliation, political and philosophical beliefs, health status, trade union membership and sexual orientation and sexual life. Researchers who manage and share data on their respondents must be familiar with these elements of the Law, as well as with the numerous restrictions it imposes. Article 92 of the Law stipulates that restrictive measures regarding the collection of personal data shall not be applied if data processing is performed for scientific or historical research or statistical purposes. In addition, related to the deposit and storage of data from scientific research in the social sciences, it envisages deviation from the set guidelines for archiving in the public interest.

Anonymisation is the primary technique available to researchers that allow data to be securely manipulated in an international environment. Anonymised data are not subject to legal restrictions, but only if the process is performed professionally and reliably. Anonymisation is the process of wholly and permanently removing identification from data by converting personal information into aggregate data. Anonymised data can no longer be linked to individual data [11].

Case study example. *DCS operates under the following laws:*

- *Law on Personal Data Protection of the Republic of Serbia (“Official Gazette of the RS”, No. 87 of November 13, 2018);*

- *General Data Protection Regulation Directive - GDPR, 95/46/EC;*
- *Law on Copyright and Related Rights (Official Gazette of RS No. 104/2009);*
- *EU Copyright Directive in the Digital Single Market, EU, 2016/0280(COD);*
- *Law on Science and Research (“Official Gazette of RS”, No. 49 of July 8, 2019) and*
- *Law on the Science Fund of the Republic of Serbia (“Official Gazette of the RS”, No. 95/2018).*

In many cases, research funders understand the position of individual researchers and seek a professional assessment of the vulnerability of respondents within the research. To ensure compliance with laws and regulations, they require the decision of the Ethics Commission or at least a person qualified to assess the risk of manipulating specific data (Data Protection Officer – DPO). Decisions are mainly based on a clearly defined procedure - Data Protection Impact Assessment (DPIA), covered by Serbian Law in Article 54).

Case study example. *DCS decide on data compliance based on a Questionnaire for assessing the impact on the protection of personal data, which covers several topics:*

- *Basic information about the project or study within which the personal data will be processed (name, client, duration) and describe the life cycle of the data to be processed;*
- *Data anonymisation level (original form, partially or entirely anonymised);*
- *Data collection method (survey, interview, focus group);*
- *The level of compliance of the data management process with certain principles set out in the Law;*
- *Is the data in the sensitive data category;*
- *Will the data leave (physically or virtually) the borders of our country and*
- *Whether the data will be archived following Article 6 of the Law.*

Bearing in mind that institutions (universities and scientific institutes) in the social sciences do

not formally appoint Ethics Commissions, such as in medical research institutions, domestic researchers can address DCS for assessment. These assessments are internationally accepted.

Consent is one essential element that enables researchers in the social sciences to share their data as Open Access. It is required throughout the entire data lifecycle (collecting, analysing, archiving, sharing, re-using). “Informed Consent” for personal data is a voluntary, determined, informed and unambiguous expression by which that person, by a statement or explicit affirmative action, gives consent to the processing of personal data relating to him. Consent needs to fulfil several requests. It needs to be:

- **Voluntary:** the person giving it must not be influenced, have the right to refuse or revoke the decision without any consequences and must not be in a dependent relationship;
- **Certain:** it must have clear information about the scope and consequences;
- **Informed:** the content and requirements of consent should be easy to understand, easily accessible, transparent and written in simple language, especially when information is given to children;
- **Active:** “silence as a sign of approval” is not taken into account, and
- **Revoking consent** must be as simple as providing it.

Adding a clause on data sharing and archiving allows the respondent to decide on a specific notice of purpose. This empowers them and allows them to choose whether or not to contribute to a research project and make their data available for future research projects. The best way to obtain the informed person’s consent to data sharing is to identify and explain all possible future manipulation of their data and offer the participant the opportunity to consent or refuse.

Formally speaking, Consent can be oral and written, but also one-time and permanent. Written Consent is in accordance with the law and is usually a request of the Ethics Committee or similar body and provides better protection to the researcher. However, in some cases, it is impossible to obtain it when the respondents are, for example, helpless people or felons. Also,

seeking this type of written consent may frighten the respondent. Verbal Consent is only valid if recorded on audio or video recording. Its disadvantage is that it is often difficult to cover all activity elements in video/audio format, which causes distrust among respondents. One-time Consent is simple and non-threatening to respondents but restricts data usage to specific research and prevents archiving and sharing. Permanent Consent is most appropriate for covering the whole data life cycle. Researchers should keep in mind the advantages and limitations of each category to avoid turning to their respondents (again) if they have not received their permission to deposit and share data in Open Access.

Case study example. *DCS provides several levels of access to data in order to protect researchers and respondents (regardless of whether the data is anonymous or not):*

- *Public-Use-Files (PUF) are data and documents that can be accessed through the dissemination portal without any restrictions;*
- *Free-Use-Files (FUF) are data and documents that can be used exclusively for the purposes of scientific analysis in academic research and teaching and which can be accessed without the permission of the depositor. Through the ordering system, either by e-mail or online form, the user can gain access and free download files from this category. Prior user registration is required, and*
- *Scientific-Use-Files (SUFs) are data and documents that can only be accessed with the written approval of the depositor. The user must familiarise DCS with the details and specify their research objectives. DCS sends a request for approval to the depositor and allows the user access only if the depositor submits written approval. In order to access these files, the user must sign an agreement on the terms of use.*

Depending on the specific research, the data collected, and the accompanying documentation, researchers in the social sciences in Serbia, when depositing data sets, can choose one of the offered levels of protection. However, it should be noted that the goal of DCS is to make the data

“Open as Possible and Protected as Necessary” to get as close as possible to full Open Access.

B. Data as Intellectual Property

Technological progress has enabled the opening of scientific papers and data collected within research. Retaining copyright in researchers’ work in the new environment has become challenging. There are many examples of abuse and appropriation. The situation is further complicated when it comes to published data sets. Researchers should be familiar with licensing, ownership, copyright, and intellectual property to protect their work and avoid threatening other people’s rights. Many academic institutions offer their researcher advice to protect intellectual property and copyright, but data on this topic are still insufficiently clear. However [12] drew attention to this very complex issue, noting that copyright protection regarding data (regardless of form - quantitative or qualitative) depends on the local regulations of each country.

One of the most common ways to protect intellectual property concerning data is licensing. A license is a legal instrument for a copyright holder or content producer to enable a second party to use their content and apply certain conditions and restrictions to those uses. When it comes to the UK, researchers can use one of three groups of licenses. In the case of public-funded research, researchers have a formal obligation to make their data open after a period of the embargo (if needed). OGL (Open Government License for Public Sector Information) [12] is typically used in this case. If the research project is funded by a third party, e.g. business sector, usually a full license applies, at least for a while. Finally, researchers mainly use Creative Commons licenses in the case of other funders.

Creative Commons is an international organisation [12] that provides free licensing through public copyright licenses. Trademark is CC completed with main additional terms: BY for attribution - implies the recognition of intellectual property rights of data creators in the process of copying, distribution, presentation, performing, doing derivative work, or remixing; SA for share alike derivative work - those who use data and publish the results of research obtained based on other people’s data must ensure that their research results have the same or higher level of restriction as to the original, not less; NC – non-commercial is used to exclude the possibility of commercial use and ND – non-

derivative work, since 4.0 version that allows derivative work, but without the possibility of sharing it.

Case study example. *When it comes to intellectual properties, DCS advises its depositors to:*

- *Decide, at the beginning of the research project, the person who will own the copyright on data. It is not always straightforward, especially in large teams from different institutions or, even, countries. In most cases, Principal Investigator (PI) or the project leader is the copyright owner, but in multinational studies, local team leaders can obtain copyright on country-specific data. It is very important to make a clear decision about this subject to avoid possible conflict after the project.*
- *Suppose the data is at a higher level of protection than Open Access (restricted use or embargo). In that case, researchers can use depositing options listed in Illustration 4, and in other cases, use Creative Commons licenses unless the contract with funders provides otherwise. DCS uses CC 4.0 version for publically available data.*
- *DCS provides guidelines and examples for data citation. Citation ideally needs to contain author(s) of the data set (in accordance with the agreement with the funder or within the research team, as explained), publication date, title, edition, version, feature name, resource type (for example database or data set), Unique numeric fingerprint (UNF), identifier (in most cases DOI or Handle) and location. An example of data set citation in APA style is:*

Smith, T.W., Marsden, P.V., & Hout, M. (2011). General social survey, 1972-2010 cumulative file (ICPSR31521-v1) [data file and codebook]. Chicago, IL: National Opinion Research Center [producer]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. doi: 10.3886/ICPSR31521.v1.

CESSDA ERIC put an additional effort to provide learning materials for researcher in social sciences – Data Management Expert Guide – DMEG [13]. The main purpose of DMEG is to support researchers with knowledge of how to make their data FAIR which basically mean to be findable, understandable, sustainably accessible and reusable. Experts from all over Europe join the effort to provide easy-to-use guide with numerous examples and country specific approaches.

IV. CONCLUSION

Researchers in science must have a clear vision of their rights and obligations. The introduction of the concept of Open Science and Open Access to scientific publications and data collected within the research further complicates the situation. In some research areas, reliance on codes of ethics is sufficient. However, when it concerns data in the social sciences, which are collected from private individuals, researchers must be familiar with the laws related to this area. Endangering privacy can have catastrophic consequences for an individual, especially if a person belongs to a vulnerable group.

ACKNOWLEDGEMENT

The authors acknowledge support from the Ministry of Education, Science and Technological Development of the Republic of Serbia and CESSDA ERIC.

REFERENCES

- [1] *Organisation for Economic Co-operation and Development (OECD)*. (January 2022). Accessed: 15.February 2022 <https://www.oecd.org/sti/inno/open-science.htm#:~:text=Open%20science%20encompasses%20unhindered%20access,by%20ICT%20tools%20and%20incentives>
- [2] *Data Centre Serbia for Social Sciences (DCS)* (2022). Accessed: 2.3.2022 <https://datacenterserbia.com/sr/>
- [3] *Data Centre Serbia for Social Sciences (DCS)*. (2022). Accessed: 1.3.2022. <https://datacenterserbia.com/en/basic-information>
- [4] *FP7 SERSIDA project - Support for the Establishment of National/Regional Social Science Data Archives*. Accessed: 1.3.2022. <http://www.serscida.eu/en/>.
- [5] *HORIZON2020 project CESSDA SaW (Strengthening and Widening)*. Accessed: 1.3.2022. <https://cessdasaw.eu/>
- [6] *SCOPES project - SEEDS (South-Eastern European Data Services)*. Accessed: 1.3.2022. <https://seedspjproject.ch/>
- [7] *CESSDA ERIC*. Accessed: 1.3.2022. <https://www.cessda.eu/>

- [8] *Law Insider*. (2022). Accessed: 16.2.2022. <https://www.lawinsider.com/dictionary/research-data>
- [9] Greene, T., Shmueli, G., Ray, S., & Fell, J. (2019). Adjusting to the GDPR: The Impact on Data Scientists and Behavioral Researchers. *Big Data*, 7(3), 140-162.
- [10] Zakon o zaštiti podataka o ličnosti Republike Srbije, "Službeni glasnik RS", 87/13/2018.
- [11] Bradić-Martinović, A., & Zdravković, A. (2013). Zaštita privatnosti – anonimizacija podataka. In: *V naučni skup USPON – Preduzetništvo kao šansa: zbornik radova*. Valjevo: Univerzitet Singidunum (pp. 206-213).
- [12] Hrynaszkiewicz, I., Cockerill, M.J. (2012). Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals. *BMC Research Note*, 5(494), 1-12.
- [13] CESSDA ERIC. *Data Management Expert Guide – DMEG*, Accessed: 20. February 2022. <https://www.cessda.eu/Training/DMEG>