Integration of western Balkan countries into the European system of digital data archives in social sciences: the case of Serbia

Aleksandra Bradić-Martinović [1+], Aleksandar Zdravković [2]

[1] Institute of Economic Sciences, Belgrade, Serbia

[2] Institute of Economic Sciences, Belgrade, Serbia

Abstract. *Science has a very important role in the development of any society. A special dimension of social phenomena has been analyzed by social sciences like anthropology, economics, education, communication studies, sociology etc. Despite that this type of research should provide solid basis for solving many social problems, the researchers in this area in the Western Balkans are facing many problems. The main limitation is the lack of funding for quality research, and the consequences are reflected in the small number of scientific papers published in internationally recognized journals and studies without connection to real problems.*

Regarding the fact that process of collecting primary data is most expensive and time consuming phase in the research process, establishment of national digital data archives for research data in social sciences and their integration into the standardized system for data sharing on the international level is considered as cost savings solution. Digital data archive (DDA) is technique that would deal with issues of preservation and archiving primary research data. In this paper we analyze the concept of DDA and its significance for researchers in social sciences, give an overview of EU effort in this field, reflected through the work of Council of European Social Science Data Archives (CESSDA), organization that gathers social data archives across Europe. In addition, we analyze the current practice of data archiving in Western Balkan countries and possibilities to build the prototype of DDA in Serbia according to the European standards.

Keywords: social sciences, database, data archive, research, DDA, CESDA

JEL Codes: 03, 031, 034

1. Introduction

Science has a very important role in the development of any society. A special dimension of social phenomena has been analyzed by social sciences like anthropology, economics, education, communication studies, sociology etc. On the other side, revolution in information and communication technology (ICT) has enabled companies to use scientific and technical knowledge much easier in order to achieve competitive advantage on the market. Knowledge becomes primary driver of a firm's value (Bock, Zmud, Kim & Lee, 2005) and has been recognized as one of the key sources of growth in the global economy (WB, 2012). Ability of storing, analyzing and sharing data and information through networks, Intranet or Internet, gives a completely new dimension of knowledge used in the companies. Arzberger et.al. (2011) point out specificity of knowledge as an economic and public good, since knowledge does not diminish by using, but increases by sharing. Kline and Rosenberg (1986) revealed a central place of knowledge and research (knowledge discovery) in innovative process that enables organizations to achieve competitive advantage.

[+] Corresponding author: abmartinovic@gmail.com .

Research has indicated (UKDA, 2002; Corti, et al., 2011) that there has been a sharp increase in collecting data that has been used in studies of economic, political and other social issues, over the last decades. Regarding the fact that process of collecting primary data is most expensive and time consuming phase in the research process, establishment of national digital data archives for research data in social sciences and their integration into the standardized system for data sharing on the international level is considered as cost savings solution. Digital data archive (DDA) is technique that would deal with issues of preservation and archiving primary research data. The practice of archiving data in the social sciences is extended in the European Union. Umbrella organization of all EU archives in social science is Council of European Social Science Data Archives – CESSDA.

Currently, opposite to EU countries, primary data collected through research in social sciences in Bosnia and Herzegovina, Serbia, and Croatia, remain mainly in possession of researchers or research institutions that have conducted specific research projects. Major step in the process of solving this problem, with the key objective to improve the research infrastructure of Western Balkan countries is FP7 – SERSCIDA project. The project addresses the issues of potentials of usage of information-communication technologies for the benefits of scientific research and exchange of knowledge in Western Balkan countries.

In this paper we analyze the concept of DDA and its significance for researchers in social sciences, give an overview of EU effort in this field, reflected through the work of Council of European Social Science Data Archives (CESSDA), organization that gathers social data archives across Europe. In addition, we analyze the current practice of data archiving in Western Balkan countries and possibilities to build the prototype of DDA in Serbia according to the European standards. The paper is organized as follows. In the first section we described role of data collection within research process. Second section deals with the issues of data archiving and DDA in Europe. Third section presents field research on current practice of data archiving in Serbia in the context of DDA establishment.

2. Research process as a base of knowledge

Kline and Rosenberg (1986) revealed a central place of knowledge and research (knowledge discovery) in innovative process that enables organizations to achieve competitive advantage. Figure 1 presents the relationship between research, knowledge and phases in the design and implementation of innovations. Numerous arrows indicate the complexity of their interactions and interdependences. Without the need for detailed analysis of individual bonds, we can point out a direct connection of research with all phases of practical implementation and indirect connection with creation of knowledge base.

Research, generally speaking, is a complex process made up of several major phases. Regardless of research, common phases, presented on Figure 2 are:

1. **Study planning** - phase of conception and according Brink at al. (2006) this phase of research involves activities with a strong conceptual element. In this phase researcher formulate the research problem, hypothesis, methodology and define the necessary data,

2. **Data collection** - expensive and very important phase of research often associated with extensive field work for interviews (face-to-face, telephone, computer assisted personal interviewing), questionnaires (paper-pencil questionnaires, online surveys), etc,
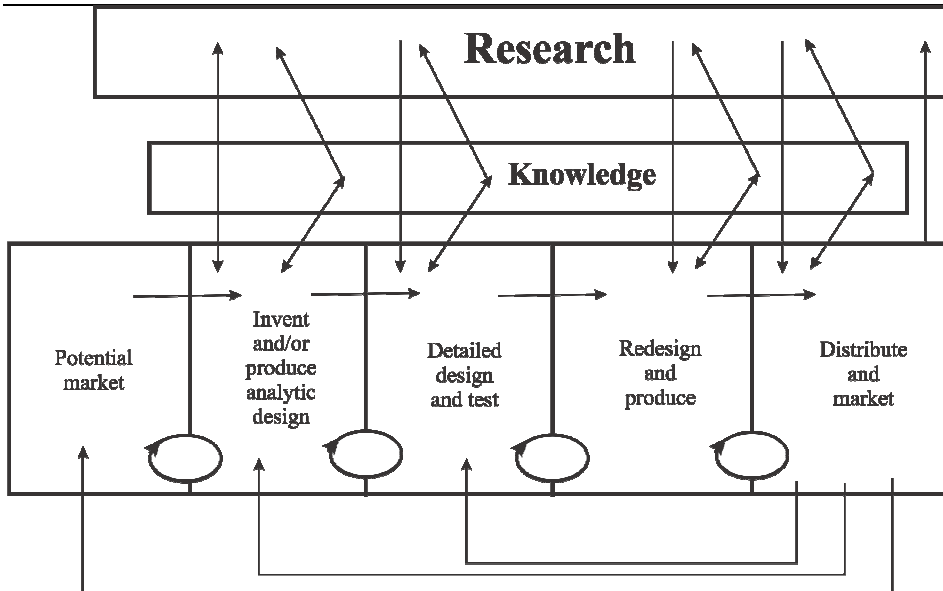
Fig. 1: Chain link model of innovation.
Source: Kline, S, J, Rosenberg, N, (1986), p. 290.

3. **Data analysis** – data must be *organized* (e.g. classification of data according characteristic properties or features), *observe* for possible existence of relationships among the two or more recorded data or parameters (e.g. correlation), *compare* with the work of other researcher and *try to find a generalised conclusion* based on limited amount of results
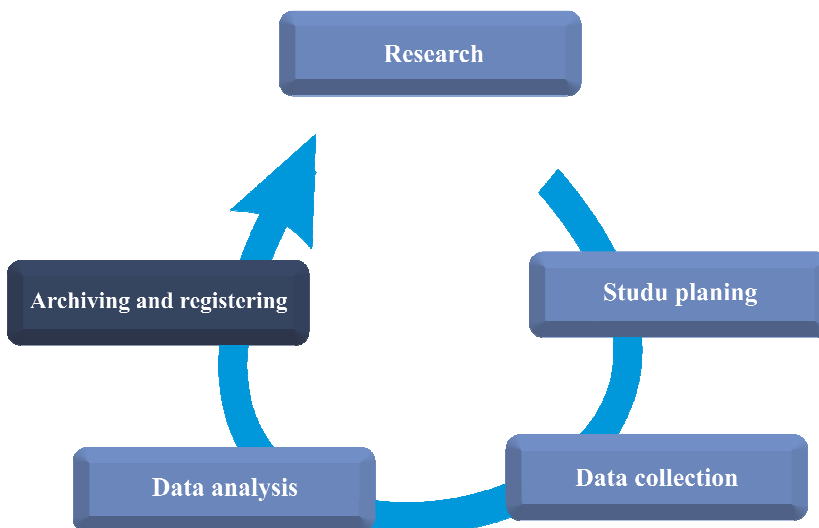

Fig. 2: Phases of research projects.
Source: GESIS, http://www.gesis.org/en/services/

The rapid development of ICT has caused significant changes in the research area. Researchers in contemporary environment can keep data in place of their origin or can easily perform data analysis at

various levels and multidimensional, due to the digital form of data they work with. This opportunity can give them a completely new insight into the research problem. NSB (2005) believes that digital collections of research data "*provide new phenomena for study*". Therefore, long-live digital data collections and digital archives are gaining great importance for researchers. However, researchers in research and scientific (i.e. think-tank) institutions must make additional efforts and adapt to changes, with the goal of full exploitation of all offered possibilities. It is obvious that "*digital information compiled for research purposes is playing an increasingly important role in today's knowledge economy*" (NDAC, 2002). Consequently these changes have created a need for open access to digital data archives (DDA) which contain research data and institutional repositories which contains scientific and research papers. As a result of these processes we can add one more phase in the research process:

4.   **Archiving and registering the data** – is the last phase of research process when there is a DDA to preserve, store and disseminate the data together with research results and make them available for research community for further analysis.

3. Archiving and sharing data

The practice of archiving and data sharing began in the fifties. There are several reasons for data sharing and Fienber (1994) argues that the archiving and sharing process:

1. Reinforces open scientific inquiry. When data are widely available, the self-correcting features of science work most effectively.
2. Encourages diversity of analysis and opinions. Researchers having access to the same data can challenge each other's analyses and conclusions.
3. Promotes new research and allows for the testing of new or alternative methods. Examples of data being used in ways that the original investigators had not envisioned are numerous.
4. Improves methods of data collection and measurement through the scrutiny of others. Making data publicly available allows the scientific community to reach consensus on methods.
5. Reduces costs by avoiding duplicate data collection efforts. Some standard datasets, such as the General Social Survey and the National Election Studies, have produced literally thousands of papers that could not have been produced if the authors had to collect their own data. Archiving makes known to the field what data have been collected so that additional resources are not spent to gather essentially the same information.
6. Provides an important resource for training in research. Secondary data are extremely valuable to students, who then have access to high-quality data as a model for their own work.

Research has indicated (UKDA, 2002; Corti, et al., 2011) that there has been a sharp increase in collecting data that has been used in studies of economic, political and other social issues, over the last decades. At the same time, there has been significant progress in computer technology and networks, statistical methodologies and techniques. It makes possible to collect, store and share data in completely new way. Modern technology has enabled the research data to be stored in digital form and to be used and reused by numerous of researchers. Today, the process can be completed directly or via Web, very easy and with small costs. Despite that, many research projects rely on primary data, even when the research on specific phenomenon has already been done by another research team. In a small country like Serbia high costs for data collecting can be big obstacle for researchers, especially in social science. For instance, in order to get personal opinions in some matter, social science researcher must provide sufficient number of questionnaires to obtain good statistical sample. That process requires time and money. The situation is particularly difficult if one takes into account the fact that Serbia in 2001 allocated 27.9 million euro or 0.16% of GDP from the

state budget for scientific research compared with Finland that allocated 3.41% of GDP at the same year. In the meantime the situation has improved, so in 2011 Serbia has allocated 110 million euro or 0.33% of GDP, but it is still far below the requirements set by the Lisbon Declaration, which proclaims that funds for this purpose should to be 3% of GDP, of which 1% would be allocated directly from the state budget. One way to improve the situation and facilitate the researchers is to create a national digital data archive which would store research data created in social sciences with a view to their reuse in the form of secondary data. Arbor (2012) points out that archives and domain repositories that preserve and disseminate social and behavioural data perform a critical service to the scholarly community and to society at large, ensuring that these culturally significant materials are accessible in perpetuity.

There are many reasons in favour of the concept of DDA, but researchers must be aware that the process of archiving requires additional knowledge and effort to prepare the data by standards. Data archiving is a process and researcher must learn how to follow all the phases of the data lifecycle.

The data lifecycle includes the following phases: creating data, processing data, analysing data, preserving data, giving access to data and re-using data
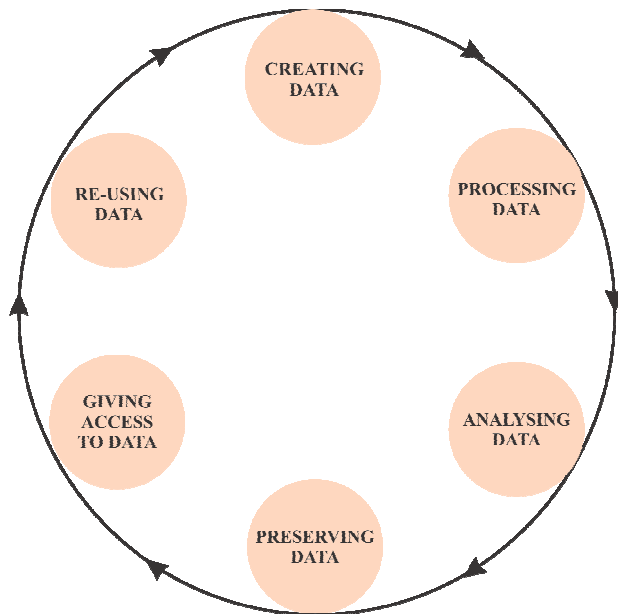

Fig. 3: The data Lifecycle.
Source: Corty, et al. 2011, p.15

Each phase of data life cycle is equally important, and must be conducted in accordance with the standards. UK Data Archive (Corty, et al. 2011, p.15) describes the stages of archiving (Figure 2):

- Creating data - plan data management (formats, storage, etc.), plan consent for sharing, locate existing data, collect data (experiment, observe, measure, simulate) and capture and create metadata.

- Processing data - enter data, digitalization, transcribe, translate, check, validate, clean data, anonymise data where necessary and manage and store data.

- Analyzing data - interpret data, derive data, produce research outputs, author publications and prepare data for preservation.

- Preserving data - migrate data to best format, migrate data to suitable medium, back-up and store data, create metadata and documentation and archive data.

- Giving access to data – distribute data, share data, control access, establish copyright and promote data.

- Re-using data – follow-up research, new research, undertake research reviews, scrutinise findings and teach and learn.

In developed countries, worldwide DDAs function for decades and one of the biggest is in the USA and UK. The history of DDA in Europe can be seen thru the history of United Kingdom Data Archive (UKDA). The first steps in the field of digital data archives in Europe have been made in United Kingdom by Social Science Research Council (SSRC). The SSRC established in 1967 SSRC Data Bank in University of Essex. During the seventies Data Bank was faces with the problem of *resistance* to the archiving data, due to the standard which had been hard to fulfilled by the researchers. The turning-point came in the early 1970s when the Government Statistical Service enabled government surveys to pass to the Survey Archive, as the Data Bank had been renamed in 1972 and became SSRC Data Archieve (UKDA, 2002). During the 1980s Data Archive put the *focus on data from empirical research* and from research that had been considered to be of *public concern*, although this period can be seen as a low point for the social science in UK. The nineties bring the need for *closer and deeper partnership among researchers* within academic institutions and users of their researches. Rapidly expanding of the Internet services has enabled a deepening of this cooperation. New millennium has brought *new standards and technology solutions*, and greater *international cooperation* among digital archives in Europe.

The practice of archiving data in the social sciences is extended in the European Union. Umbrella organization of all EU archives in social science is Council of European Social Science Data Archives – CESSDA. It was founded in the seventies and today gathers digital archives of twenty one European countries. Key objectives of CESSDA are: to ensure that the data which researcher collect today can be used in the future, i.e. help research institutions to make a plan for the archiving process; to provide a gateway to social science data (some 30,000+ social science and humanities researchers and students within the European Research Area - ERA each year, providing access to 25,000 data collections, delivering over 70,000 data collections per annum and acquiring a further 1,000 data collections each year), to organize education and knowledge sharing thru expert seminars and conferences and to cooperate with other international organizations sharing similar objectives. Today, CESSDA is in the process of transformation to CESSDA ERIC (European Research Infrastructure Consortium).

4. Data archiving practice in Serbia – field research

Although majority of EU countries establish DDA in the field of social sciences, in Western Balkan countries there are no existing or sustainable social science data archives in these countries that would deal with issues of preservation and archiving of those primary data. Thus, research institutions in WB countries are currently faced with the problem of possible loss of valuable research data forever due to the lack of infrastructure and capacities for archiving such data in an adequate manner in line with existing European standards.
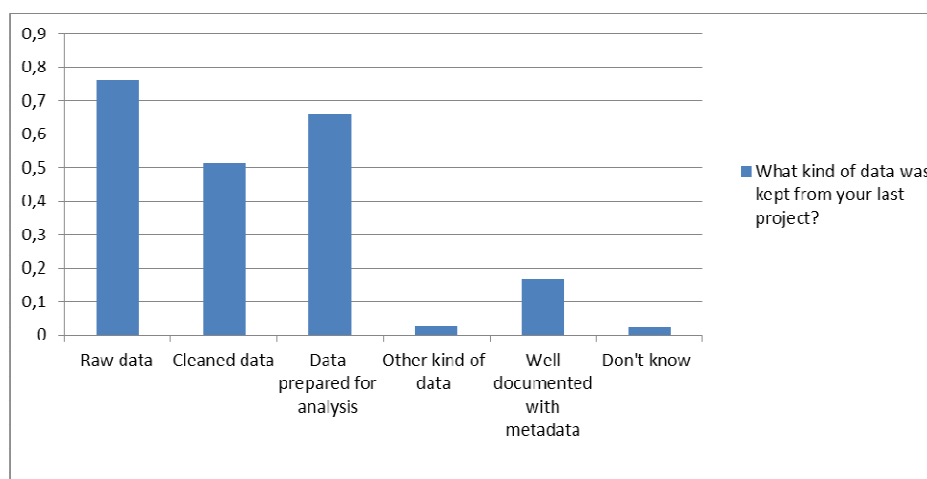
Generally speaking, during the establishment of the national DDA scientific community meets the similar problems: a) *technological issues* related to the establishment appropriate standards for data

collection and preparation for the preservation and providing compliance with those standards[1]; b) *institutional and managerial issues* that arise as a result of different forms of organizing the research process and policy; c) *financial issues* regarding funds that has to be provide for the functioning of DDA; d) *legal and policy issues* regarding the copyright and open access to the research data and e) cultural and behavioural issues are very important to those produce and those who manage research data, because they need a proper technique for data promotion according to sharing practice (Arzberger et.al, 2011).

In this section we try to address some of these issues for WB countries, more specifically institutional and managerial, as well as cultural and behavioural issues. We illustrated these issues using the survey of current practice in data archiving in Serbia, but it could be extended to other WB countries having in mind numerous similarities among them as a consequence of similar level of economic and social development. According to our opinion, these issues are the most critical obstacles in acceleration of the process of DDA building. Here we present results of our field empirical research, as the starting point for the future development of DDA and its integration in European DDA system. This survey is a part of broader research within FP7 – SERSCIDA project.

One part of our survey was loosely designed to follow scheme of the data lifecycle, which is given in previous section. We realized survey by sending the questionnaires to the social sciences researchers' community in Serbia and with response rate around 36%. Our objective was to investigate how much the practice of data preserving is widespread among the researchers producing the primary research data and how the data are preserved.

Generally, 38% of respondents confirmed that they or their research teams preserved data after research project was completed. Figure 4, in regard to the data processing and analyzing phases within data lifecycle, reveals usual forms of data kept. The majority of researchers have kept their data either in raw form or cleaned form to certain level suitable for the analytics[2]. However, practice of data documentation with metadata which is crucial for further secondary analysis is very low widespread, i.e. around 17% of researchers have accompanied documentation to the data in some occasions.

---

[1] This is particularly true in countries with the existence of DDAs at the level of individual institutions (institutes, university research centers, NSI archive, etc).

[2] This question allows multiple choices. Vertical axis shows separated relative frequency of each answer relative to number of total positive responses on data preserving practice. This also holds for the following three issues discussed.

Fig. 4: Forms of data kept

The following figure further investigates issue of data documentation. Not even that researchers in Serbia do not practice to document the data, but when they do it, they do not apply some international standards like DDI, DC etc.
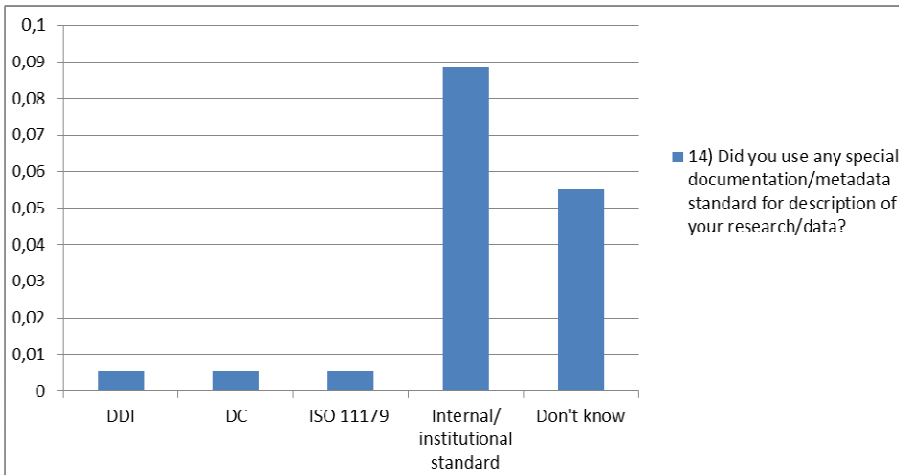

Fig. 5: Standards applied to data documentation

Figure 6 presents usual choice of mediums where data are preserved, in line with following face of data lifecycle phase. The majority of researchers used to store data at personal computers' or computers of other members of research team, while few of them in some occasions put it on servers or archives, which is an assumption for broader availability of data for secondary analysis.
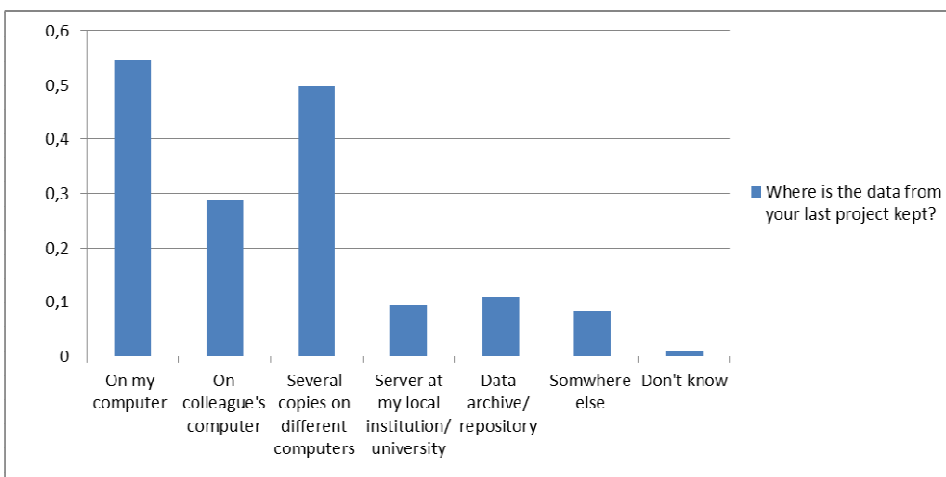

Fig. 6: Mediums where data are preserved

Finally, figure 7 gives an insight into the issues of access to data. As it is expected in regard to the low level of data standardization and choice of local mediums for data preservation, just a few of researcher made their research data available for use of wider research community. Unfortunately, the majority of datasets remains available for secondary analysis only for the members of research team.
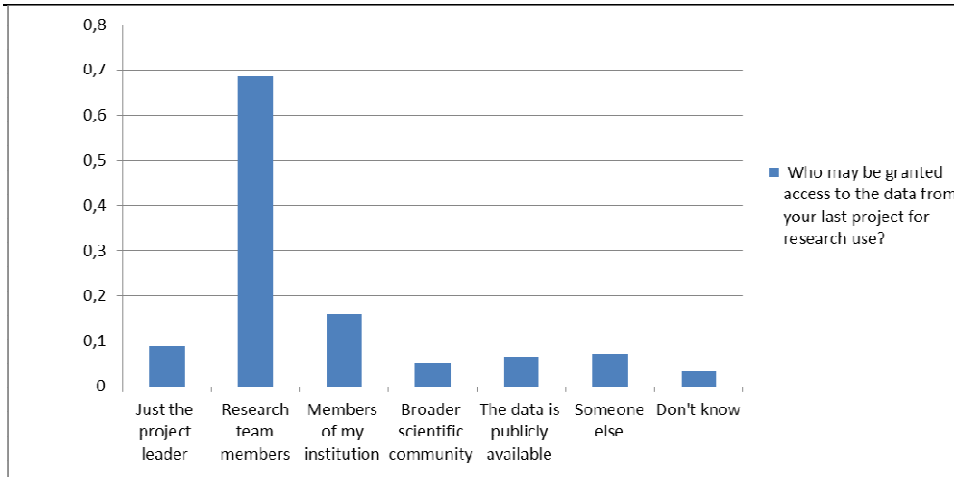
Fig. 7: *Mediums where data are preserved*

To summarize, analysis of responses to these several questions clearly give evidence that researchers in Serbia as the main producers of the data in social sciences have very low level of knowledge on the issues of data preservation. Data are mostly preserved in raw form, without appropriate documentation and standardization according to the international standards. In addition, data are mostly kept on local mediums with very limited access, usually only to the members of research teams. This altogether represents the main obstacle to the major development of the DDA and availability of data for secondary analysis.

5. Conclusion

Digital data collections are powerful catalysts for progress and for democratization of the research and education enterprise. Proper stewardship requires effective support for these essential components of the digital research and education environment of the 21st century. The need for digital collections is increasing rapidly, driven by the continuing exponential increase in the volume of digital information. The number of different collections supported by the NSF is also increasing rapidly.

Currently, primary data collected through research in social sciences in Bosnia and Herzegovina, Serbia, and Croatia, remain mainly in possession of researchers or research institutions that have conducted specific research projects. Data collected through such research, although of high value for social scientists (both in these countries and internationally) for any further exploration, remain unavailable. There are no existing or sustainable social science data archives in these countries that would deal with issues of reservation and archiving of those primary data.

We illustrated these issues using the survey of current practice in data archiving in Serbia. Our analysis shows that researchers in Serbia have very low level of knowledge on the issues of data preservation. Consequently, data are mostly preserved in raw form, without appropriate documentation and standardization according to the international standards. In addition, data are mostly kept on local mediums with very limited access, usually only to the members of research teams.

Ongoing FP7 - SERSCIDA project aims to overcome all of these obstacles and to provide solid base for establishment regional or national digital data archives in social sciences in Serbia, Croatia and Bosnia and Herzegovina.

6. References

[1] Arbor A., Colyer C., Donakowski D., et al., *Guide to Social Science Data Preparation and Archiving – Best Practice Through the Data Life Cycle,* 2012, Michigan, pp. 1-43

[2] Bock,G. Zmund R., Kim Y., Lee J., *Behavioral intention formation Knowledge Sharing: Examining the Roles of Extrinsic Motivators, Social-Psychological Forces, and Organizational Climate*. Management Information Systems Quarterly, 29(1), 2005., pp. 87-111.

[3] Brink H., Van der Walt C., Van Rensburg G., *Fundamentals of Research. Methodology for Health-care Professionals*, Second Edition, Juta and Co (Pty) Lt, 2006.

[4] World Bank, Knowledge for Development – K4D, http://web.worldbank.org/WBSITE/EXTERNAL/WBI/WBIPROGRAMS/KFDLP/0,,contentMDK:20269026~menuPK:461205~pagePK:64156158~piPK:64152884~theSitePK:461198,00.html (last visit 24 April 2012)

[5] Corti L., Van den Eynden V., Bishop L., Morgan-Brett B., *Managing and sharing data*. UK Data Archive, Colcehester, Essex, 2011.

[6] NDAC - National Data Archive of Canada, *Building Infrastructure for Access to and Preservation of Research Data – Final Report*, Social Science and Humanities Research Council of Canada, 2002.

[7] NSB, *Long-lived digital data collections: enabling research and education in the 21st century*, National Science Foundation, Report, 2005.

[8] OECD, *The Knowledge-based Economy*, Head of Publications Service, OECD, Paris, 1996.

[9] Arzberger P., Schroeder P., Beaulieu A., Bowker G., Casey K., Laaksonen L., Moorman D., Uhlir P., Wouters P., *Promoting Access to Public Research Data for Scientific , Economic, and Social Development*. Data Scence Journal, 3 (November), 2004., pp.135-152.

[10] Feinberg S.E., *Sharing Statistical Data in the Biomedical and Health Sciences - Ethical, Institutional, Legal and Professional Dimensions*, In Annual Review of Public Health, 15, Palo Alto, CA, Annual Reviews, Inc., 1994, pp. 1-18

[11] Kline S.J., Rosenberg N., *An Overview of Innovation, monograph* The Positive Sum Strategy, Harnessing Technology for Economic Growth, 1986., pp. 274-305

[12] UKDA, *Preserving and Sharing Statistical Material*, The Royal Statistical Society & the UK Data Archive, University of Essex, 2002.