PODRŠKA STVARANJU DRŽAVNIH/REGIONALNIH ARHIVA
ISTRAŽIVAČKIH PODATAKA IZ OBLASTI DRUŠTVENIH NAUKA
SUPPORT FOR ESTABLISHMENT OF NATIONAL/REGIONAL
SOCIAL SCIENCES DATA ARCHIVES

Analysis of existing potentials for the establishment of a social sciences digital data base archive in Croatia

SEVENTH FRAMEWORK
PROGRAMME

Disclaimer

The information in this Report does not necessarily represent the view of the European Commission (EC).

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288985.

Preface

SERSCIDA is an EU funded project, funded under FP7 Capacities work programme, Science in Society thematic area. It is a strategic project, designed to support cooperation and knowledge exchange between those EU countries that are members of the Council of European Social Sciences Data Archives (CESSDA) and the Western Balkan (WB) countries in the field of social science data archiving. The project addresses the existing potential for use of information-communication technologies for the benefit of scientific research and exchange of knowledge, as laid down in the call for proposals topic. It aims to produce tangible results and improve the capacities for exchange of knowledge and data collected through research in social sciences between the European countries and WB countries involved.

The four main objectives of SERSCIDA are:

- to support the establishment of social science data archives in order to increase the level of sharing and preserving data collected through research in WB countries;
- to identify potentials and infrastructures for establishment of social science data archives in WB countries involved;
- to enhance the exchange of knowledge and sharing of data collected through research in social sciences in the WB countries and EU – CESSDA countries;
- to increase the level of understanding and support for open access policies through involvement of policy and decision makers in WB countries;

This Analysis is a mapping exercise of the existing capacities and scope for the establishment of a digital data archive in Croatia. The same exercise simultaneously took place in Bosnia and Herzegovina and Serbia. The Analysis will serve as a baseline survey for this project and as the foundation from which other SERSCIDA activities will be developed.

The tasks foreseen in the preparation process of the Analysis included:

- development of methodology for analysis of existing potentials for establishment of social science digital data archives in the WB region,
- analysis of existing potentials and infrastructure for data archiving in the WB countries involved and
- provision of recommendations for making use of the existing potentials.

Acknowledgments

The Analysis of existing potentials for the establishment of a social sciences data archive in Croatia is prepared by Faculty of Humanities and Social Sciences, University of Zagreb. The Analysis preparation has been lead in co-operation with the SERSCIDA partners from *Fondation Suisse our la Receherche en Sciences Sociale* of the University of Lausanne, Switzerland, *Goeteborgs Universitet*, Goeteborg, Sweden, *University of Essex*, Colchester, UK, *Univerza v Ljubljani*, Ljubljana, Slovenia, *Institut ekonomskih nauka*, Beograd, Serbia, and *Centar za ljudska prava*, Univerzitet u Sarajevu, Sarajevo. Special thanks for support during the preparation process, vision and suggestions go to Brian Kleiner and Hilary Beedham. Setting the scene for methodology has been done by Hans Joergen Marker and Mattias Persson, Goeteborgs Universitet. The authors of this report are Marijana Glavica, Teo Matković and Alen Vodopijevec.

Table of contents

Preface 3
Acknowledgments 4
Introduction and methodology overview 7
Policy and Legal Frameworks 9
Survey on production, preservation and use of research data among researchers 14
Survey on existing infrastructure and interest for data archiving 31
Summary and conclusions 43
Recommendations and future steps 45
Annexes 46
Annex 1. List of interviews held 46
Annex 2. Questionnaires 48

Abbreviations

CESSDA - Council of European Social Sciences Data Archives

DOI - Digital Object Identifier

EIB - European Investment Bank

FP7 - Seventh Framework Programme for Research and Technological Development

GDP - Gross domestic product

ISBN - International Standard Book Number

ISSN - International Standard Serial Number

MOST - Management of Social Transformations Programme

PASW - Predictive Analytics SoftWare

RRPP - Regional Research Promotion Programme

RZS - Statistical Office of the Republic of Serbia

SAS - Statistical Analysis System

SCDA - Social Sciences Data Archives

SERSCIDA - Support for Establishment of National/Regional Social Sciences Data Archives

SPSS - Statistical Package for the Social Sciences

WBC - Western Balkan Countries

WP2 – Work package 2

Introduction and methodology overview

This is the Croatian report of work package 2 (WP2) which focused on mapping the existing potentials for establishing a social science data services in WB countries. This mapping will help to identify possible participants for the conference planned within the WP3, as well as inform the CESSDA member institutions about current situation in WB countries which will help them to develop training modules to be organized within WP4. By taking stock of the existing potentials for the establishment of a data archive the report provides recommendations for the establishment of social sciences research data services in Croatia.

To collect the information which could illustrate the existing potentials and areas to be improved, three main groups of stakeholders who can play important role in establishing, maintaining and using a data archive were identified. These three groups are:

- The research policy and funding bodies
- The research/data service infrastructures institutions
- The researchers

Considering the research policy and funding bodies the survey aims to assess the support for the establishment a data archive both in financial terms and in a research policy setting. The research infrastructures can provide both technical infrastructure such as servers and data capacity, and more service oriented infrastructure such as e.g. survey collection software. When addressing the research infrastructure, institutions undertaking research in the social sciences were included as well. The researchers will provide and use data, as well as be a potential knowledge resource for the archive.

Due to the difference between the three groups, a general survey which would cover all groups was not practical. Instead, the survey instrument for each group was designed to fit the specific characteristics of that group. In producing the survey instruments SERSCIDA has benefited from experience from a similar survey conducted by the CESSDA PPP project and from national surveys conducted by CESSDA members in their own nations.

Each survey instrument is presented briefly below, with main areas of interest and how the data collection was done.

The Interview guide for Research Policy and Funding Institutions

In this case the ministry which handles the research policies in Croatia, as well as the major research funders were approached, either face to face or via phone. The purpose was to establish contact and determine the possibilities for financing a data archive both in the establishment phase and in the long term. This also included questions and discussions about if the funders had any requirements related to research data e.g. archiving of data or open access. To these questions of requirements on archiving of research data, information was also sought as to whether the current science policy includes anything regarding archiving research data. Another crucial question in this case were the ministry’s and funders view of whether data archiving should be carried out at national or local level.

The Questionnaire for Researchers

The instrument which was designed to collect information on production, preservation and use of research data in the field of social sciences is an online survey questionnaire. The online survey was prepared using LimeSurvey software hosted by University of Zagreb University Computing Center. The questionnaire collects information about the researchers' experience of documentation, re-use and disseminating of research data, but also about which type of statistical/analytical software packages,

methodology and data are primarily used in their research. In addition to these areas the researchers were also asked if their institution had any policy regarding long term preservation and/or documentation of data. This gives a brief overview of the demand for data, training and support and the supply of data and of experience with documentation.

The Interview Guide for Research Infrastructure Institutions

The instrument which was designed for the research infrastructures was a semi-structured interview/survey. It was conducted either face to face or via phone with representatives of research infrastructures and research institutions with in the social science. The National Statistical Institute was also contacted. The respondents were asked to answer questions about whether they could provide technical infrastructure or research service. They were also asked if the institution had any policy or competences for long-term preservation and documentation of data. The institutions and infrastructures were also asked if they had any available technical capacity or data collecting services which could be valuable for the establishment of a data archive. The survey was also a way of getting in touch with potential future collaborators with the data archive.

Part I: Policy and Legal Frameworks

1. Basic features of science system in Croatia

Research activities in Croatia are mostly conducted at universities (higher education sector) and in public research institutes (government sector), overwhelmingly financed from the state budget. Research and development activities (R&D) are also conducted in the busine sector, mainly funded from its own resources, but in terms of social science, the number of such research studies compared to the government sector and higher education sector is negligible[1].

The Government of the Republic of Croatia sets the basic principles and guidelines for the development of science and higher education. The main administrative body responsible for the planning, financing and monitoring of the entire science and education system is the Ministry of Science, Education and Sports (MSES). There are different professional bodies that contribute to the development and quality of science and science system which are active in the Republic of Croatia. The highest professional body in the field of science is the National Council of Science (NCS), which appoints the science and art councils and the main committees from the ranks of distinguished scholars, artists and professors. The National Council for Higher Education operates in the field of higher education. The Science and Higher Education Funding Council is the professional body of the NCS and the NCHE, setting the criteria for the allocation of budget funds for science and higher education and monitoring those criteria. To perform the evaluation of institutions and programs in science and higher education, the Croatian Government established an independent and autonomous Agency for Science and Higher Education. As regards the funding for research activities, in addition to MSES, the Croatian Science Foundation plays an important role, funding a variety of programs, research projects and university reforms.

Detailed information about the science system in Croatia can be found on the website Erawatch - information platform on the European, national and regional science research systems and policies[2] while here we feature the elements relevant to the aims of the SERSCIDA project regarding the current situation in the system of science financing in Croatia and shaping of science policies.

The current situation regarding the financing for scientific activities is extremely poor, while the global economic crisis certainly contributed to this. Investment in research and development is below the EU average and has been dwindling in recent years.

A new government was elected in late 2011 and consequently the current Croatian policy is outlined in the Croatian Government Program for the 2011-2015 mandate. This program does not explicitly mention development of research infrastructure, which is a topic of relevance to the SERSCIDA project. The emphasis is placed on closely linking the science and industry, in other words, establishing a strong science commercialization, wherein the scientists perceive a risk of further cuts in financial resources for social science research, as the research in this area often has no commercial value. Viewed from the other side, social science research can inform government and other institutions about how funds could be spent more effectively. In the Government program the emphasis is placed on the importance of inclusion into the European Research Area, although it is not fully operationalized, thereby implicitly giving importance to research infrastructure and international knowledge sharing. A new strategic document for science, education and technology is under development. In April 2012, the MSES issued the Guidelines for the Strategy in Education, Science and Technology, in cooperation with the Croatian

1 Croatian Bureau of Statistics (2011). Research and development, 2010. http://www.dzs.hr/Hrv_Eng/publication/2011/08-02-01_01_2011.htm

2 Erawatch Croatia. http://erawatch.jrc.ec.europa.eu/erawatch/opencms/information/country_pages/hr/country

Academy of Science and Art. A project entitled the "National Strategy for the Development of Croatian Innovation 2013-2020" is also planned. The objectives set in the 2006-2010 Science and Technology Policy of the Republic of Croatia will be in effect until the adoption of the new strategy. In addition to the new strategic documents, the plans include amendments to the legal framework governing the system of science and higher education.

Essentially, the system of science in Croatia is in a period of a major change, while the current financial situation prevents the key institutions from pledging, in one form or another, funding for future activities to be established within the SERSCIDA project. However, the current situation regarding the relevant policy documents can be seen as a good opportunity and the right moment to incorporate the developing guidelines into strategic documents and thus ensure a steady sustainability of activities related to data archiving, preservation and dissemination of the research data in social sciences.

2. Current policy related to science infrastructure

Existing policy of the Republic Croatia related to science infrastructure of the Republic of Croatia is given in the Program of Activities of the Republic of Croatia for the term 2011-2015 dated December 2011 and in the Guidelines for strategy of upbringing, education, science and technology dated 13 April 2012.

On 28 May 2012 the Commission for Science Infrastructure was established with the aim of defining operative policies related to science infrastructure. The Commission's task is to propose strategic guidelines for:

- construction and development of national science-related infrastructures in the Republic of Croatia and establishment of closer links with the European science-related infrastructures with the aim of better usage and development of science-related infrastructures and better access of Croatian scientists to the European scientific capacities;
- policy related to usage, direction and maintenance of national scientific infrastructures including related human resources;
- support to particular national infrastructure projects;
- European cooperation in the field of fundamental science-related infrastructures;
- investments required for development of national infrastructural projects (together with the representatives of coordination boards);
- enhancing use and integration of science-related infrastructures with the wider community, including economy and analysis of research infrastructure return in terms of science and economy;
- informing the wider public about the importance of research infrastructures, including education of children and the young

3. Key components of the research infrastructures and public institutions taking part in development of research e-infrastructure

Through its past activities and budgetary investments Ministry of Science, Education and Sports has developed key components of the research e-infrastructures composed of the following:

- network infrastructure

- computer and storage infrastructure
- authentication and authorization infrastructure - AAI@EduHr

The Ministry's key partners in the development of e-infrastructure are CARNet and the University Computer Center of the Zagreb University – Srce, as well as all other institutions part of the system of science and education that can and are willing to contribute to the development of the common research e-infrastructure in the Republic of Croatia.

Croatian academic and research network CARNet ensures interconnectivity of all participant the research process so as to ensure communication, data exchange and access to resources and se of science system services. CARNet is part of the Pan-European research network GEANT.

Srce computing centre ensures stable, reliable, high-quality and advanced e-infrastructure, closely linked with the European one, for the need of academic and research community in Croatia as a whole.

4. Support for Open Access to scientific information

In October 2012, with the aim of raising awareness and responsiveness of all those that take part in creation, publication, usage and preservation of science-related information in Croatia "Croatian Open Access Declaration"[3] was published and promoted. The Declaration underlines fundamental importance of scientific information, need to make them available to everyone and its mandatory preservation. It was made by a group of professionals who have vast experience in working with, encountering and use of information with open access with a desire for its adoption by the key bodies, institutions and organizations and to act decisively and without any reserves in the spirit that is behind it. Ministry of Science, Education and Sports (MSES) expressed its support to the Declaration: *"The principles and practice of open access to scientific information and open access to scientific information and open educational contents in the system of science and education of the Republic of Croatia will be actively supported by means of public policy. The public policy will be used to define the measures and recommendations for creation of systematic and/or institutional pre-requisites and favorable environment for creation and usage of open access contents, and/or open educational contents, as well as ensure sustainability of the achievements in the field.'"*[4] MSES has also instigated activities related to the development of public policy of open access to scientific information and open educational contents, by establishing Commission for Open Access.

5. Legal framework regarding data protection and dissemination

Copyright and Related Rights Act (Official Gazette no. 167/03, 79/07, 80/11) regulating the copyright and related rights, their implementation and protection. As one of the "related rights" the law makes reference to the "rights of data producers to their databases." The content of research data archives is subject to the Copyright and Related Rights Act, while Article 7 of the foregoing Act regulates the data pools and databases. In line with this, it will be necessary to regulate the access to data based on consent of the author (owner) of the data.

3 Croatian Open Access Declaration (2012). http://www.fer.unizg.hr/oa2012/declaration

4 Ministry of Science, Education and Sports of the Republic of Croatia (2012). *Support for "Croatian Open Access Declaration"*. http://www.fer.unizg.hr/_download/repository/HDOP_-_pismo_podrske_-_MZOS.pdf

Personal Data Protection Act (Official Gazette no. 103/03, 41/08, 130/11, 106/12) regulates the protection of personal data of individuals and the control over the collection, processing and the use of personal data in the Republic of Croatia.

Law on Archive Records and Archives (Official Gazette no. 105/97, 64/00, 65/09) regulates the protection and conditions of use, storage, use and processing of archive records, public archive service and responsibilities and tasks of the archives. Archives shall be the records or documents that are established by the action of legal or natural persons in carrying out their activities and have a permanent significance for culture, history and other sciences, regardless of the place and time of their creation, irrespective of the form and the medium that keeps them.

6. Interviews with research policy and funding institutions

Initial contacts with institutions dealing with designing and implementation of scientific policies, financing scientific area and in general, dealing with development and quality of scientific area were made at the end of September and at the beginning of June 2012.

The following institutions were contacted: Ministry of Science, Education and Sports, Croatian Scientific Foundation, Agency for Science and Higher Education, and National Council for Science. In the letter sent to the above-mentioned institutions the project SERSCIDA was presented, as well as issues that the project covers, specific objectives to be accomplished by the project, and the manner of accomplishing those objectives. The institutions have been asked for an interview in order to touch upon the issues relevant for successful accomplishing of the project objectives and to set foundations for establishing adequate and permanently sustainable infrastructure for social sciences data archive. Topics announced for discussion included analysis of the needs for such activity, existing policies on science infrastructures and legal framework, as well as possible institutional solutions for the services to be shaped within this project. Representatives of the stated institutions were invited for a round-table session on research data, which was organised on 14 November 2012. In addition, a conference was announced, to be organized within SERSCIDA project in March 2013 in Belgrade.

Ministry of science, education and sports replied to the invitation with a letter explaining the activities conducted so far by the Ministry on developing the fundamental components of research e-infrastructure, they listed the key partners of the Ministry in building the e-infrastructure, master documents defining the existing policy related to scientific infrastructure, and they also mentioned the information on establishing the Commission for scientific infrastructure. The Ministry expressed the need for activities and projects, which result in building research e-infrastructure, particularly data infrastructure. Given that the Ministry recognises holders of such activities and projects as its partners, University of Zagreb Faculty of Humanities and Social Sciences is identified as one of the partners of the Ministry in building such infrastructure and support to SERSCIDA project was expressed.

An interview was conducted with the Croatian Scientific Foundation director whereby the main topic was introduction to the conditions that are being incorporated into Scientific Projects Agreements financed by the Foundation. The Agreement sets the requirements, such as: the project must have a web page, number of scientific publications is being evaluated, as well as holding public lectures, and/or dissemination of results in any form whatsoever, including workshops, publications in professional journals, press, radio and TV shows, participation in international conferences, and lectures for students. Recommendation is that the publications are being issued in an open access approach, although this is optional. For the time being, it is hard to talk about incorporating obligations about research data

archiving in the Agreements, because it is a sensitive topic due to the issues with confidentiality and possibilities of patenting and commercial use of data. The Foundation could start considering encouraging scientist for data archiving if there was a reliable infrastructure, still this would have to be decided by the Board of Directors of the Foundation. The Foundation is currently in the process of reform and they are soon expecting the appointment of the new Board of Directors. Director of the Foundation considers that it is important to preserve the research data. The biggest encouragement for the scientist can be the cooperation, which can be accomplished once the data are shared. The discussions also covered the possible financing of the data archiving activities in the future, after project SERSCIDA ends. The Foundation is exclusively financing scientific projects and thus can ensure employment of scientific novices/young researchers.

We did not manage to arrange any meetings with the representatives of the Agency for Science and Higher Education and the National Council for science, but they have agreed to take part in the round table discussion titled “Archiving and Using of Research Data in Social Sciences - scientists' needs, existing infrastructure and shaping of scientific policies", on which we report latter in this document.

Part II: Survey on production, preservation and use of research data among researchers

1. Methodology

We have used an on-line questionnaire to gather information on the practices related to production, preservation and use of research data as well as to assess the interest for the establishment of the research data archive.

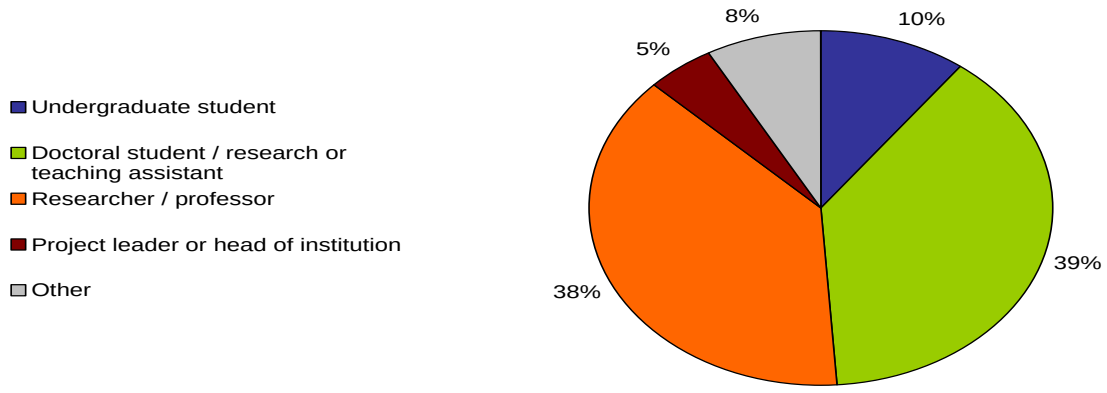
The invitation for participation was sent in June 2012 towards 31 social science institutes, faculties and departments (specifically in the fields of economy, politology, psychology, teaching, rehabilitation-educational science and social work and policy), as well as towards three national professional associations in fields of sociology, political science and psychology.

The access to the questionnaire was not restricted and did not require authentication, allowing everyone who was informed and who followed the link to participate in the survey. This is why the set of participants (and, consequently, the answers) is not likely to be representative for the entire population involved in social science research, as respondents were self-selected both at the level of institutions (which forwarded the invitation to their members with various amounts of enthusiasm) as well as at the level of individuals.

2. Survey participants

The total of 307 respondents accessed the Croatian section of the questionnaire, but 121 failed to complete the questionnaire by failing to provide substantial amount of answers about their research or data use. In general, those gave up filling their answers at an early stage starting with the questions pertaining to the details of their own research. 186 respondents included in the analyses include ten persons who failed to complete the questionnaire, but have provided substantial amount of answers. Thus constructed, the completion rate stands at 61%.

Figure 1: Current principal activity of respondents

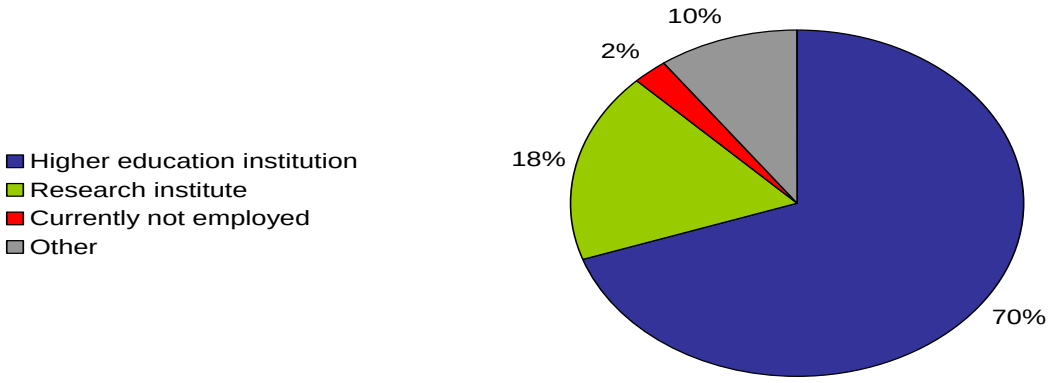


N=186

Out of 186 participants, 19 were undergraduates or lower level students (not explicitly targeted, but who got the information through the departmental or association mailing lists), 72 assistant teachers, scientific novices/young researchers and other doctoral candidates, 71 were tenured researchers or university professors, and 9 identified themselves as project or institution leaders. Fifteen participants specified their status as other (teachers, interns, advisors, self-employed researchers). As for the

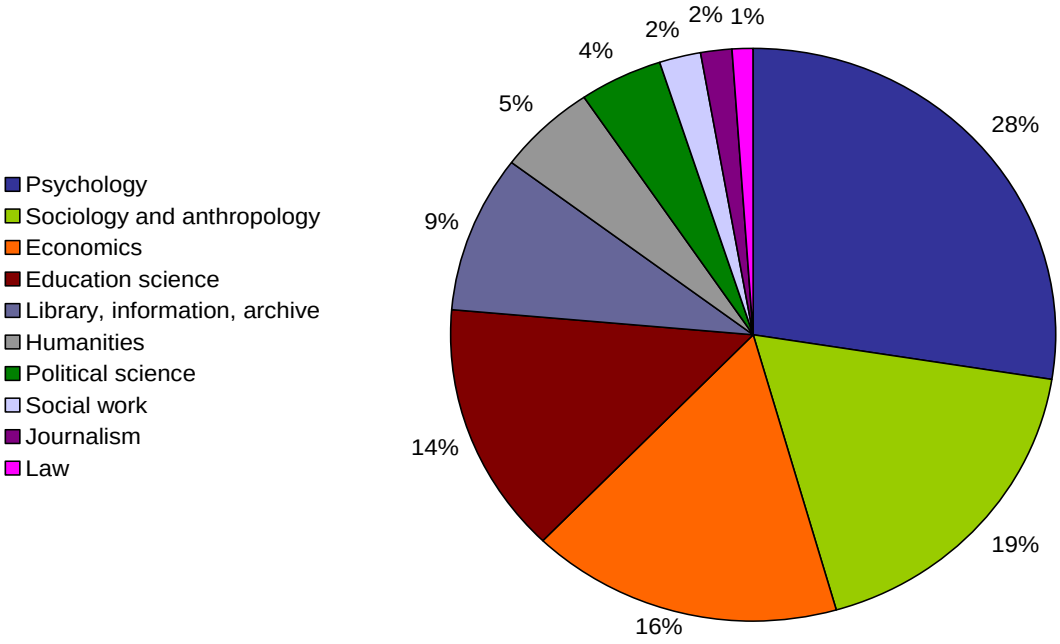
institutional affiliation of the participants (not counting students), 70% were working in higher education institutions, 18% in research institutes, total of 98% worked in the public sector,, and only 2% in the private or non-governmental sector (as those haven't been targeted with invitation letters).

Figure 2: Institutional affiliation of respondents (excluding students)



N=168

Figure 3: Principal research discipline of respondents



N=182

As for the field of education of the participants (according to ISCED criterion), the psychology is most prevalent (27%), with considerable presence of sociologists (18%), economists (16%), educational (14%) and information sciences professional (9%) politology and journalism (total of 7%). Five percent of participants were from the humanities.

The data of the Croatian Bureau of Statistics based on the Annual reports on research and development (table 5.1)[5] show that in 2010 there were 1537 researchers in the social science field, 258 in the

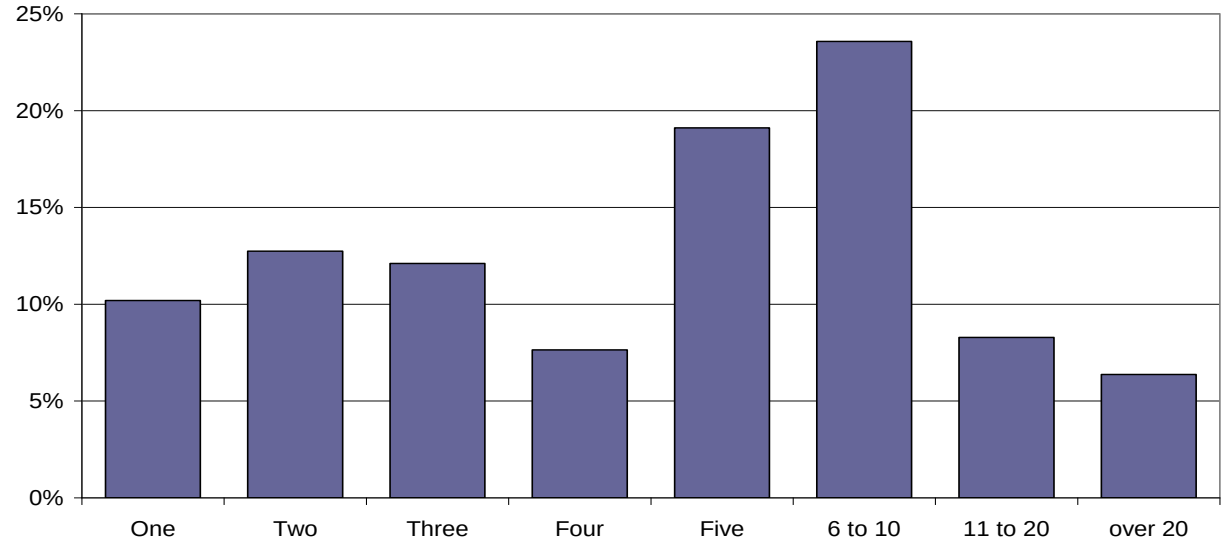
5 Research and development, 2010. http://www.dzs.hr/Hrv_Eng/publication/2011/08-02-01_01_2011.htm

government sector, 1316 in the higher education sector. Given the fact that 155 come from the social science field (excluding students and humanities) we can conclude that survey coverage stood at about 10% of the population.

3. Production of data

83% of the research participants have produced some sets of data in the past five years, and are thus to be considered to be active research data creators, that is persons of interest to the data archive. Given self-selection at recruitment, this share is to be considered as upper estimate, most likely exceeding the prevalence of data creators among the entire population of the scientists and university teachers in the field of social science. Within this and next section, we will first analyse frequency and characteristics of the research data collection efforts among this set of respondents, to be followed by inquiry in their practices of research data preservation and sharing,

Figure 4: Number of datasets the researcher had produced or helped to produce during the past 5 years



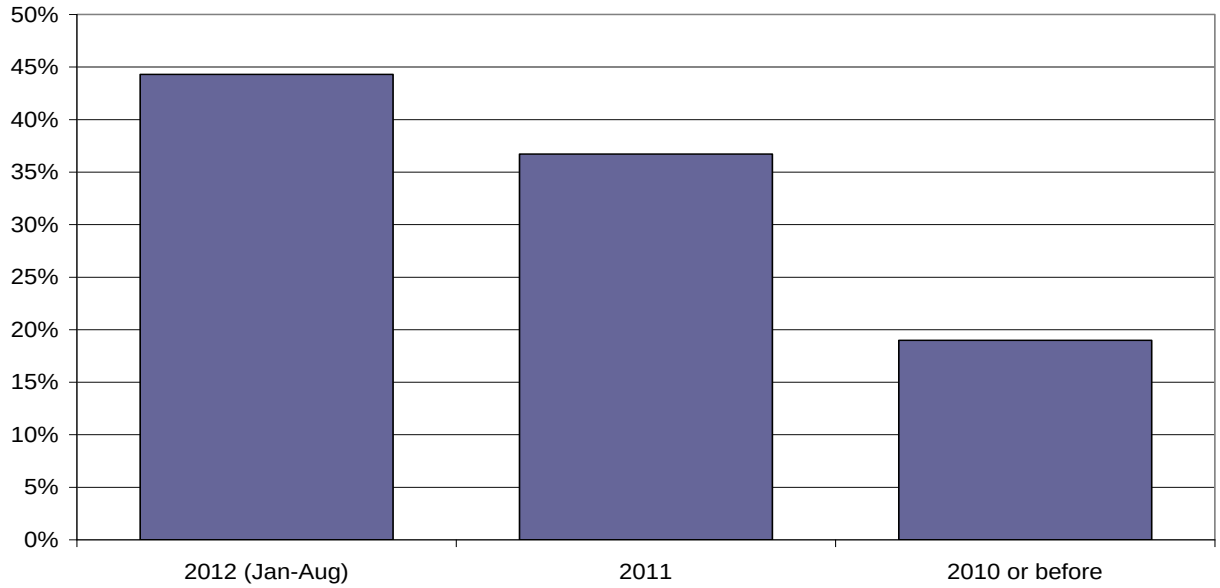
N=157

Out of all respondents who have produced research data within the past five years, 10% have participated in collection of just one set of research data, about one fourth collected 2-3, 4-5 and 6-10 each, whereas one in seven participants indicated participation in more than 10 research projects resulting in data sets creation. Median and modal number of data-yielding research efforts was 5, which is tantamount to collecting about one data set per year.

This is consistent with the following findings that 80% (128 participants) completed the field of the latest research within the past 20 months, meaning that presently they are active data creators. It is

interesting to see this level of research activity happening despite the fact that original time frame of the major ministerial (Ministry of Science, Education and Sports) scientific project cycle had expired in 2010.

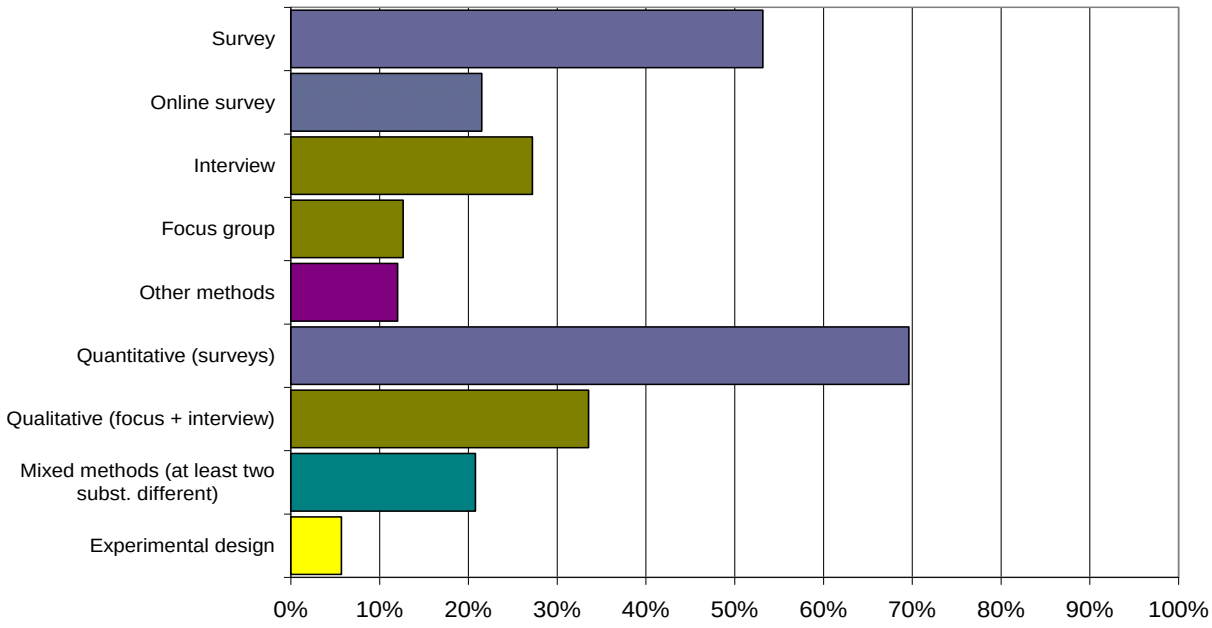
Figure 5: Year the most recent fieldwork (or data collection phase) completed



N=158

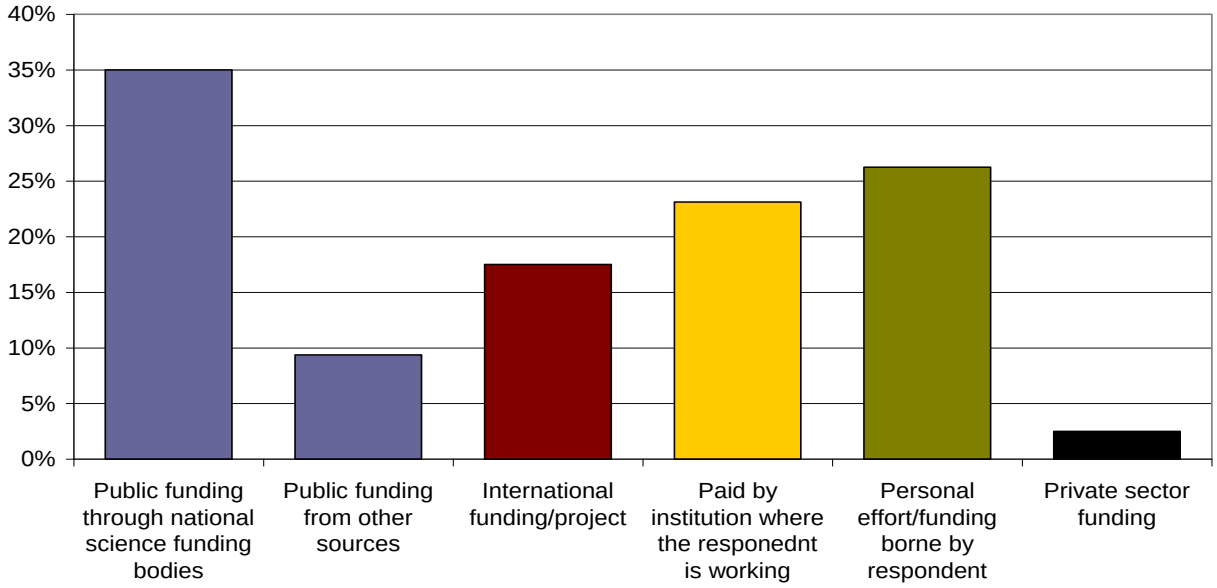
When it comes to the data collection methods used in the latest research (in several cases multiple research methods were used), quantitative data dominates (70%), out of which in 22% of the research projects on-line questionnaires were used. In about 33% of the research projects personal qualitative data were collected; either through interviews (27%) or focus groups (13%). 12.5% of the research projects used other data types (collection of texts, video recordings, ethnographic recordings, case studies, analysis of content/print, analysis of documentary material, expert reports, quotation analysis, observation with participation, archives research, observation, automatic text analysis). 20% of the projects saw a complementary use of substantially different methods (quantitative and qualitative). This indicated that along with the need for archiving of quantitative research, there is need to consider services for archival of qualitative social science data pertaining to interviews and focus groups as well as other research artefacts.

Figure 6: Data collection methods used in the last research



N=158

Figure 7: Funding sources of last research



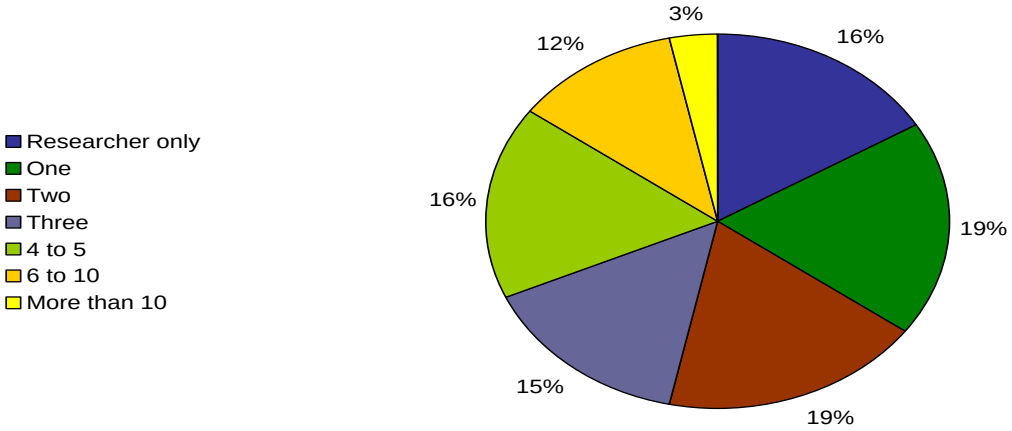
N=159

As far as the research funding is concerned in 35% of the cases the most recent research was financed from the national bodies within the scientific system (Ministry of Science, Education and Sports; National Council of Science), and in other 9% from other public sources. In 18% of the cases the funding come through the international projects. It is worth mentioning that 23% used the resources from their own institution , while in 26% cases the respondents have explicitly stated that the research was carried

out in absence of any funds, based on researchers' own efforts alone (this applies both to qualitative and quantitative research efforts). Only 3% of the reported projects were funded from private sources. This is at odds with official social science funding breakdown for research and development in the state-owned institutes and universities in 2010, which include institutional running costs and salaries. The later source reports that 80-84% of the investment came from the state and local government, 6-7% from own funds, 3-5% from funding abroad, and 5-10% from companies (private and public).

The information on size of project teams and number of associates is important in understanding the initial practices of data archival, as well as observed patterns of data use and sharing within and without the project team. One sixth of researchers did the last research on their own. However, small teams dominate – 55% of participants worked on projects that included two to four researchers. Only in one third of the cases five or more researchers worked on a project and in their latest project just 3% of respondents cooperated with more than 10 persons. The structure of research projects is evidently predominantly dispersed, with many small research teams, which is congruent with the organizational and funding practices of previous public-funded project cycles, and number of collaborators on state-funded projects in the field of social sciences.

Figure 8: Number of co-researchers involved in the most recent project



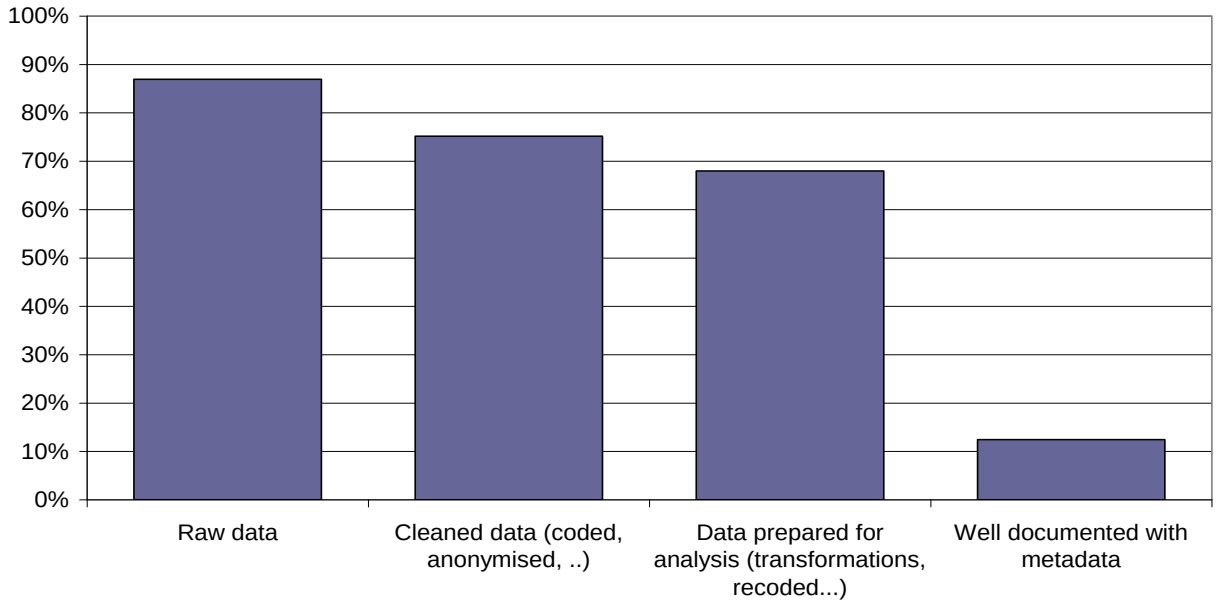
N=152

4. Archiving practices and preferences

The data relative to their (own) latest research is kept by 99% of the parties interviewed, but this is rarely done in a systematic or standardized manner.

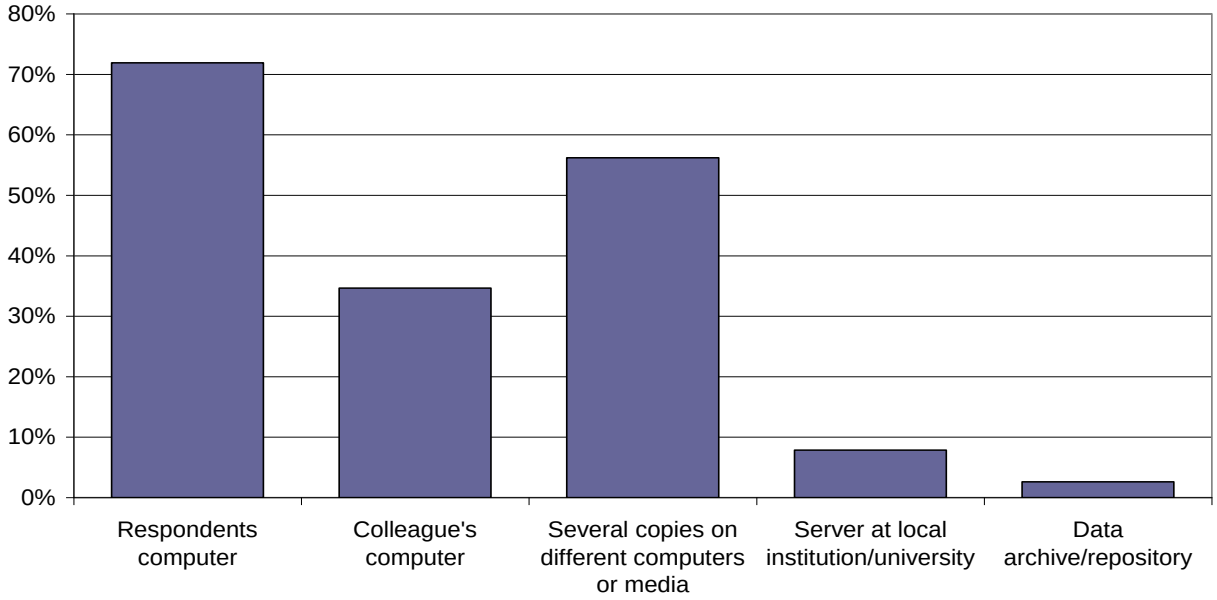
However, the standards of archiving are neither harmonized nor comprehensive. Existence of raw data was acknowledged by only 87% of respondents, "clean" and anonymized data sets in 75% of cases, and use of transformed data was reported by 68% of data producers. The existence of well documented datasets, described by metadata was mentioned by only 12% of participants. Out of those, most could not name the standards used for documenting and there was only one person that mentioned one international standard (DDI, DC, ISO 11179), while some them mentioned using in-house or institutional standards for data documentation.

Figure 9: What kind of data was kept?



N=153

Figure 10: Where is the data kept?



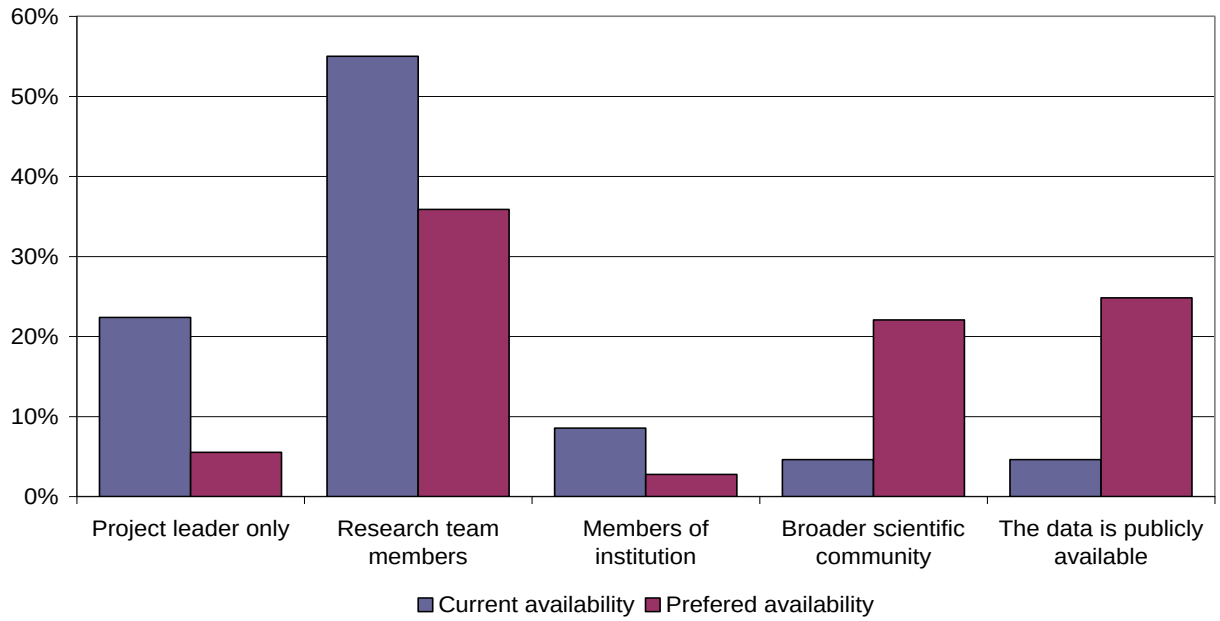
N=153

And where is data kept? The data are most often kept by the respondent himself/herself on his/her PC (in 72% of the cases). In 35% of the cases the researchers have identified that there is a copy on the other researcher's PCs (most often along with the copy on his/her own computer), and in 56% cases they stated that the copy is saved on several computers or media. In mid-term run these approaches may turn out to be very problematic and likely result in loss of data, particularly when considering that the data were systematically documented in just about one of eight cases. With respect to data keeping,

there is a similar pattern in terms of keeping data on the institution's server or in special data archives, as only 11% of participants kept their data in the same manner. In short, in only about one of eight cases the most recent research data is kept well documented or kept on some kind of repository.

As for the access to data for secondary analytical purposes, in 20% of the cases only a team leader had such privileges. However, the dominant scope of access, stated by somewhat more than half of the respondents, is that data access is limited to the research team that ran the project. It must be born in mind that this, is usually a very small group comprised of 2-4 researchers, meaning that the scope and potential for data sharing (and capacity for secondary analysis) is very limited. Institutions to which researchers belong are seldom perceived as appropriate not only for archiving but for the data sharing as well. Only in 10% of the cases did other researchers in the institution have exclusive access to data. Finally, another 10% of respondents stated that the data are available to the academic community or the public. It is indicative that this practice was more frequent among the internationally financed projects.

Figure 11: The extent of access to the data of the most recent research project



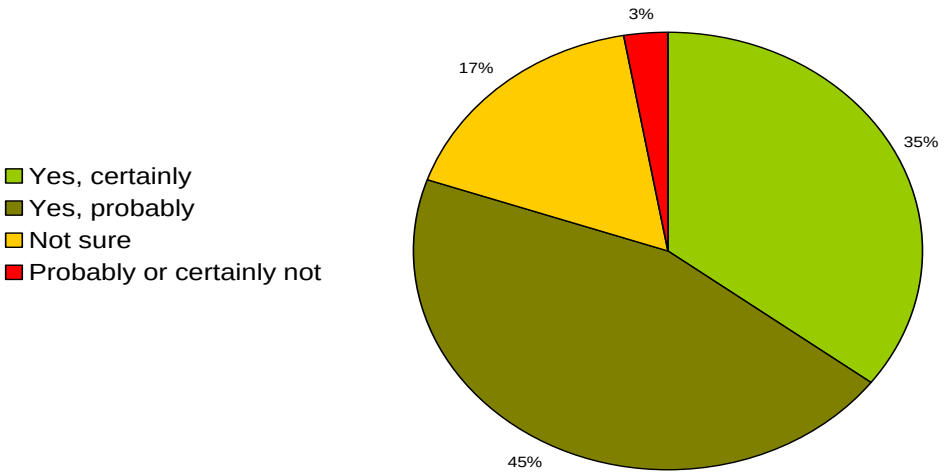
N=152 (current), 145 (preferred)

However, for about 56% of data creators, the current state of play was not considered desirable and they considered wider access to their own data to be a preferable outcome. First, only one in twenty respondents finds the data appropriation and/or control by the research project leader as unacceptable. The research team is still the modal preferred scope of access, but only for about 33% of researchers. Only a few see their own institutions as preferable boundary of data sharing practices, and almost 50% of data creators stated that their preference is for all interested researchers or the wider public to have access to data (both options are equally represented).

In case that safe archives with regulated access are up and running, 80% of participants believe that they would keep their data there. On the other hand, only 3% participants would not be willing to do it. Of course, such an assessment that is probably higher than real disposition of population of social scientists

in Croatia, primarily due to self-selection of survey participants based on their interest in the subject of research, but a data-archival institution would obviously address the reported existing needs and help in overcoming of current issues identified with respect to safe archival and enabling of access to the research data.

Figure 12: Willingness to provide research data to a social science data archive, provided that data would be safely preserved and access regulated



N=148

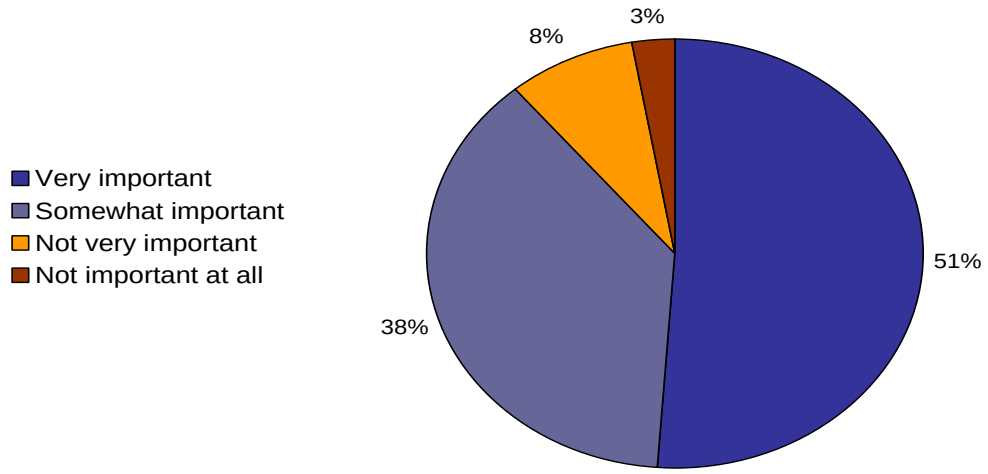
5. Use of data and secondary analysis

Half of the research participants (including those who did not produce any data) have stated that the access to research data is of paramount importance in their branch and the total of nine tenths finds it to be important.

However, implementation of that access in Croatia is significantly restricted. On the one hand, only every fifth researcher could state with certainty that the data from his research has been ever analyzed outside of the team that produced it, and only every tenth can recall this happening during past twelve months (implicating his/her data is used by other researchers on regular basis). About one half is convinced that no one has ever used their data, and about one third does not actually know with certainty whether the data they produced were used by other researchers – which would not be the case if the archives existed.

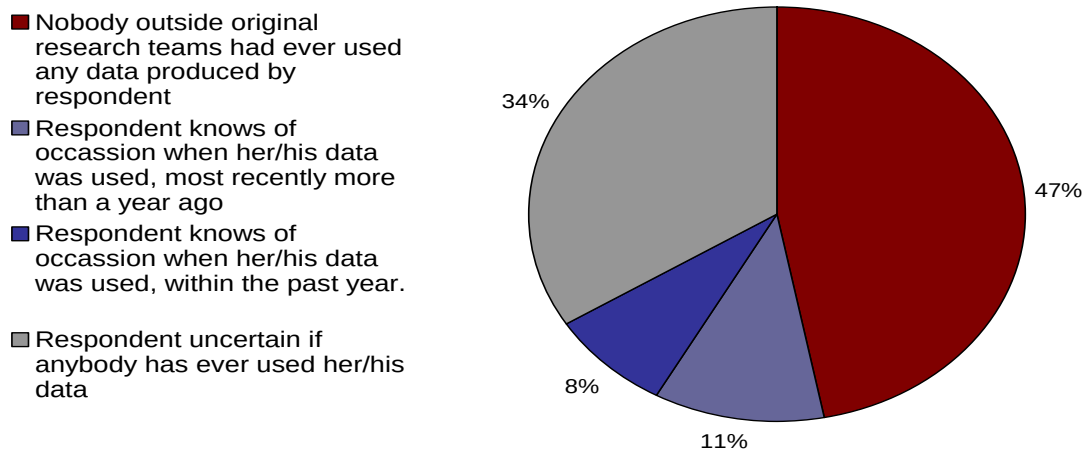
As for the secondary analysis of other researcher's data, the situation is slightly better, at least with respect to quantitative data, where one fifth of participants regularly uses such data (most recently less than two months prior to survey), and another fifth also reported that they have resorted to secondary analysis of other researchers' data at least once. Yet, 60% of the research participants did not ever analyze the data of quantitative research they did not produce themselves. As far as qualitative research is concerned, only 20% of respondents had experience in using secondary data.

Figure 13: Importance of sharing of research data within respondent's discipline



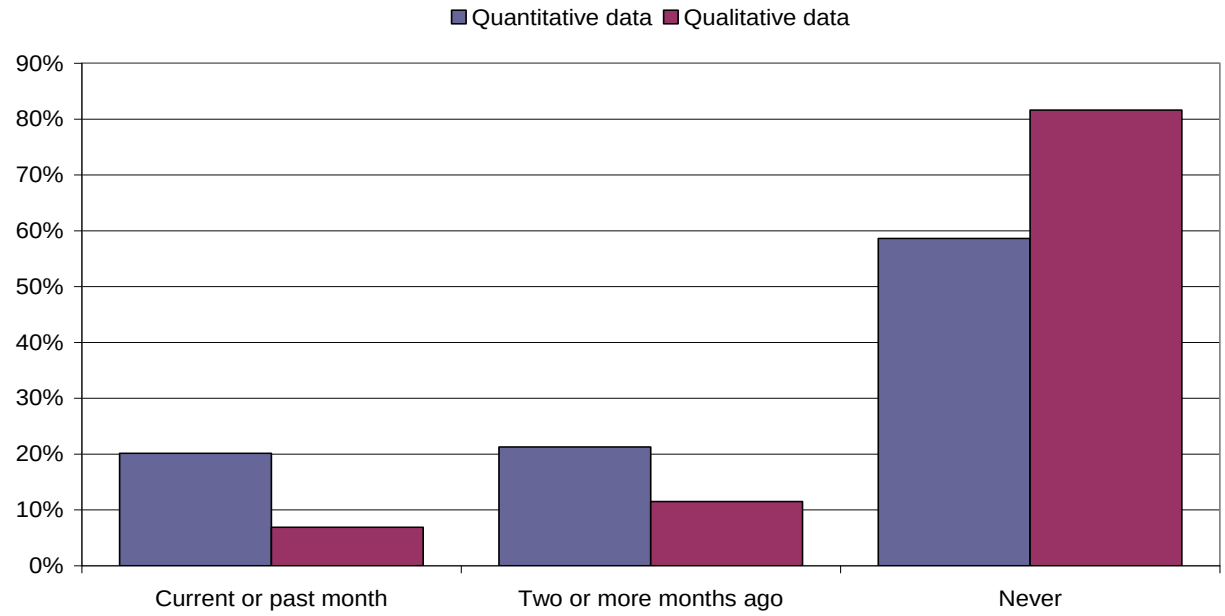
N=176

Figure 14: Knowledge about respondent's research data being ever used for secondary analysis by someone external to the original project team



N=145

Figure 15: Respondent's secondary use of research data produced by others



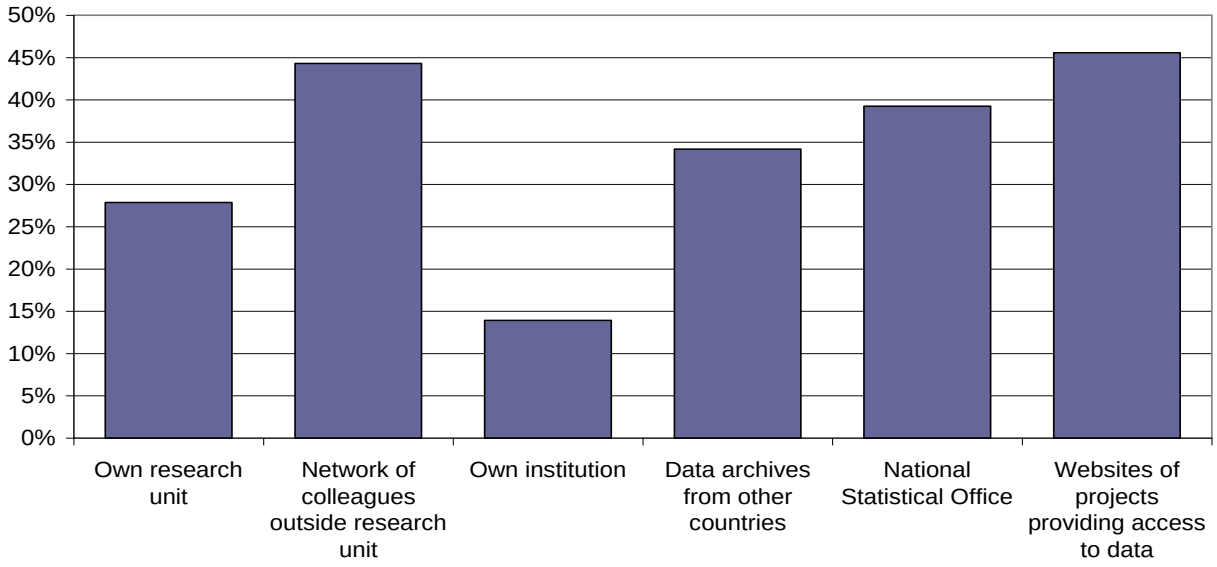
N=174

Currently, data for secondary analysis is obtained in various ways. Project websites or data repositories abroad were mentioned quite frequently as a source, and 33% of respondents mentioned using Croatian Bureau of Statistics to obtain data for secondary analysis, as this is an institution where a significant amount of research is conducted, resulting in macro and micro-data relevant for social science (e.g. LFS, HBS, SILC)[6]. Therefore, despite lack of national data archive, access to secondary data through formal channels is already common. The data is seldom obtained through colleagues working on the same project as researcher, or from the institution in which the respondents works; the most common way of obtaining the data for secondary analysis is to rely on one's own professional network outside of institutions.

As central obstacles for wider application of secondary analysis the respondents have most often cited poor availability of data, which was reported by 66% of participants, as well as the research culture in which sharing of data is not encouraged. Non-existence of adequate data and/or research was mentioned by 25% of the research participants only. Obstacles such as own methodological deficit and poorly documented data for secondary analysis were recognized by a smaller but substantive share of respondents. Though, these issues can be widely recognized only once greater availability of research data is achieved, so the frequency with which individual issues are recognised by respondents should best be understood as a list of current priorities, as it is important first to establish access and then pursue additional services aimed at building capacity and excellence in documentation.

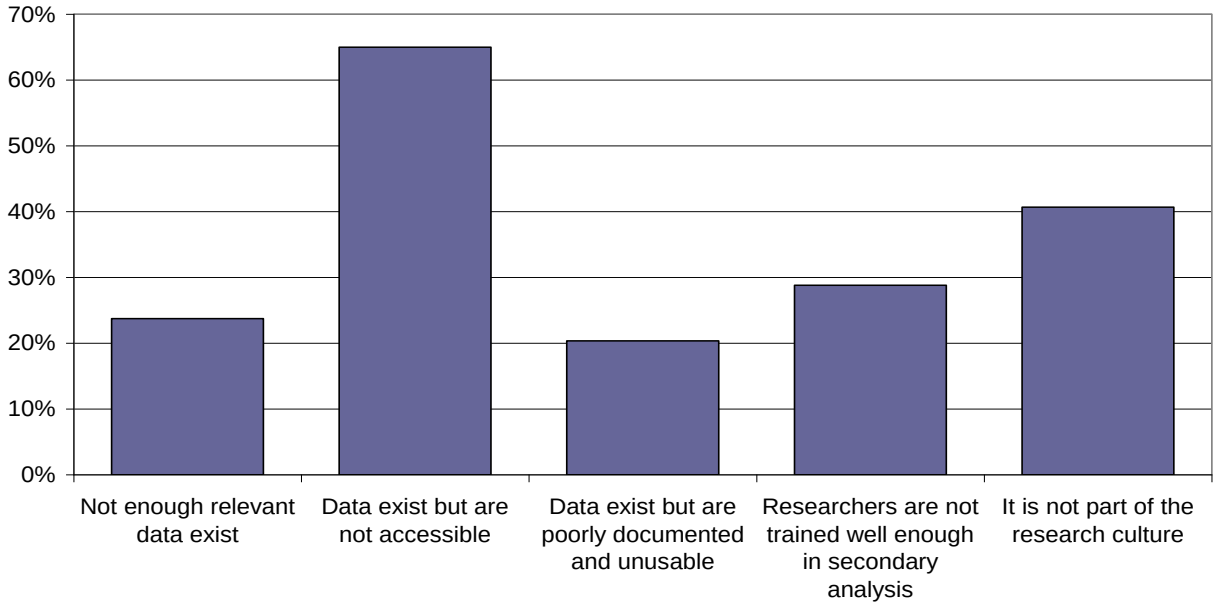
Figure 16: Sources used by respondents in order to obtain data for secondary analysis

6 The microdata access from CBS is contractually regulated for specific purposes, pending obligatory fees, and is allowed only for researchers registered in national register of scientists, working in institutions registered in national register of scientific institutions. The data is delivered in anonymised form, without territorial identificators, as a raw text data without labels or documentation (apart from the pain copy of survey pages). On some occasions, some users are consulted with CBS about implementation of changes or survey design, but not in an institutionalized fashion.



N=79

Figure 17: The barriers to conducting secondary analysis

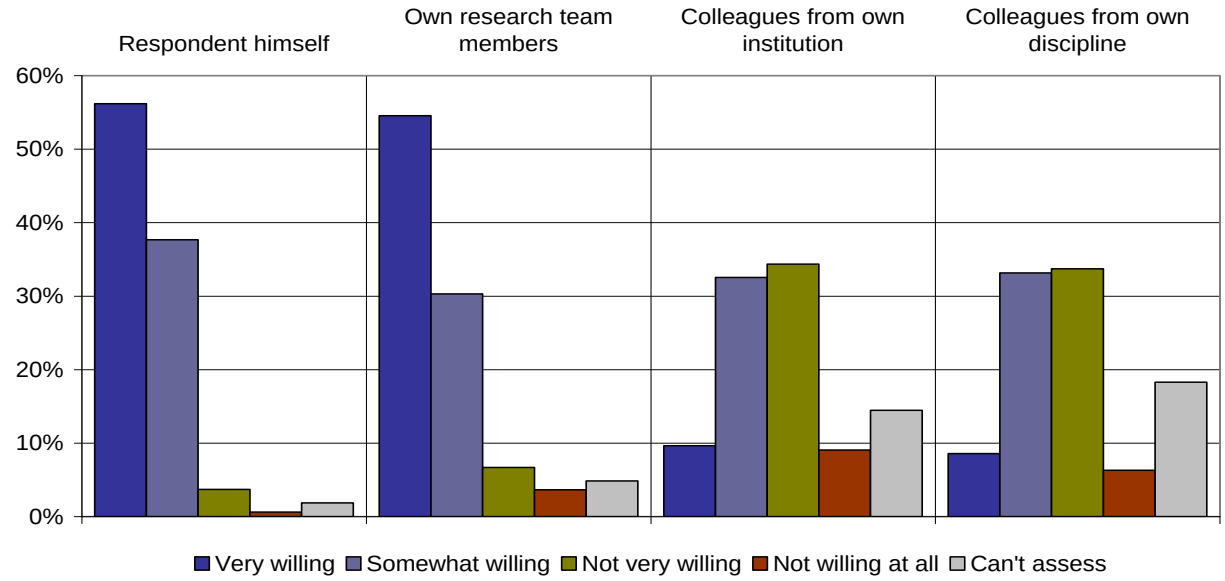


N=177

The respondent assessments of dispositions towards data sharing are also indicative. Most participants report being strongly inclined towards making their data available for the secondary analysis, and they assess dispositions of their project-mates as similarly open. However, when asked about the dispositions of their colleagues, both within the institution and their own field in general, the reported situation

becomes considerably ambivalent and assessments normally distributed, with similar number of respondents judging their colleagues to be and not to be disposed to making data available for secondary analysis. There is no overall difference between the assessments for the institution and their own academic field. There are several possible interpretations of such findings. The first one is that our respondents and their project teams are indeed more prone to data sharing (as those scientists have, in fact, decided to take part in this survey), and so provide a "core group" intent on good practice, but one which is so far without institutional support for data sharing and no means of enabling data access that they find preferable - thus their experience in sharing of their own data is less frequent and more narrow in extent than they would prefer. The second interpretation could be associated with lack of trust, as everyone sees the inclination of the others in a more negative light than their own - and such perception emerges from rarity of occurrences where such cooperation is realized. This is best fought through data sharing practice, as respondents with some experience of using others' data for secondary analysis rate the willingness of colleagues from the field to share data way better (52% affirmative), than respondents who have never done secondary analysis (34% of whom asses their colleagues in field are willing to share data). Development of a data archival and distribution service would be expected to improve data availability to researchers, and with that lead to improvement of their assessment of others' willingness to share the data they have produced.

Figure 18: Respondents opinion of the the prevalent attitude with respect to sharing one's own research data among...



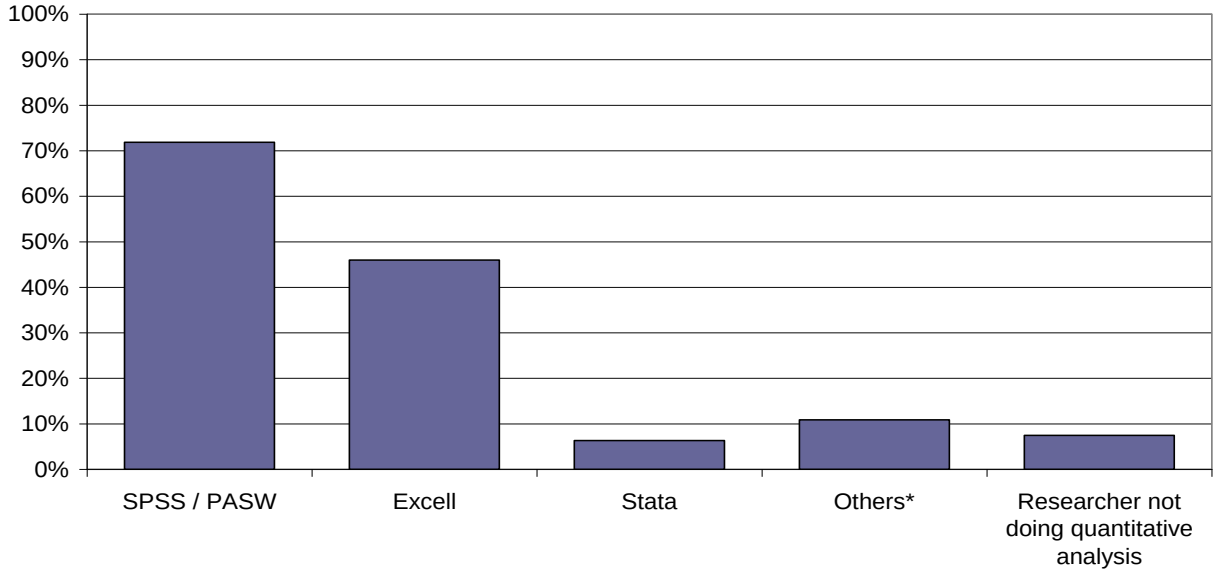
N= 162 (self), 165 (team), 166 (institution), 175 (discipline)

6. Data analysis software

The current usage of software platforms used for data analysis must be taken into account when organizing data archives and supporting services. In this respect, the current practices in Croatia for most part seem to be lagging in terms of international trends and standards.

Although only 7% of respondents have replied that they do not use quantitative analysis in their work, almost three fourths uses SPSS/PASW package. Further, almost 50% stated that they use Excel in their statistical analysis. Use of other packages is also reported, but quite sporadically: some participants use R (3), SAS (2) Statistica (3), Eviews (3) or Amos (1), whereas only Stata users are somewhat bigger in number, with 11 participants reporting using it.

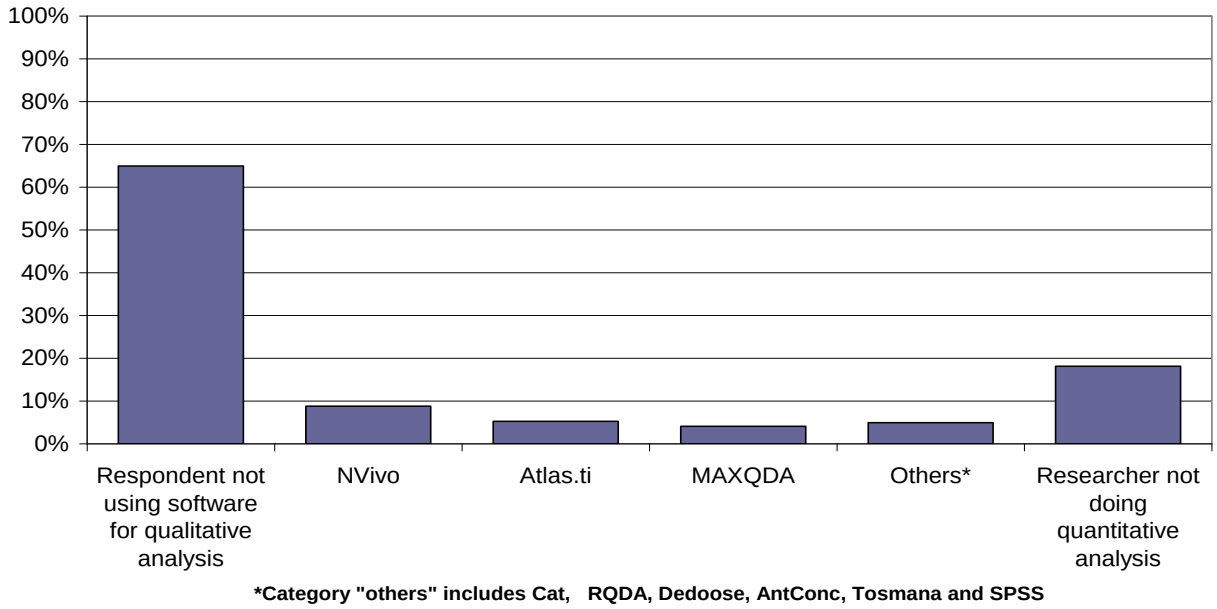
Figure 19: Software commonly used for quantitative data analysis



N=174

As for the qualitative analysis, 20% of respondents reported that they do not resort to such analysis, and additional two thirds stated that they do not use any software packages when performing qualitative analysis. The rest of the participants are, as a rule, competent in use several packages. Among this minority, largest group uses NVivo (15), but Atlas.ti (9) and MAXQDA (7) are also moderately represented, whereas CAT is used by three and AntConc and Tosmana is used by one participant of this research.

Figure 20: Software commonly used for qualitative data analysis



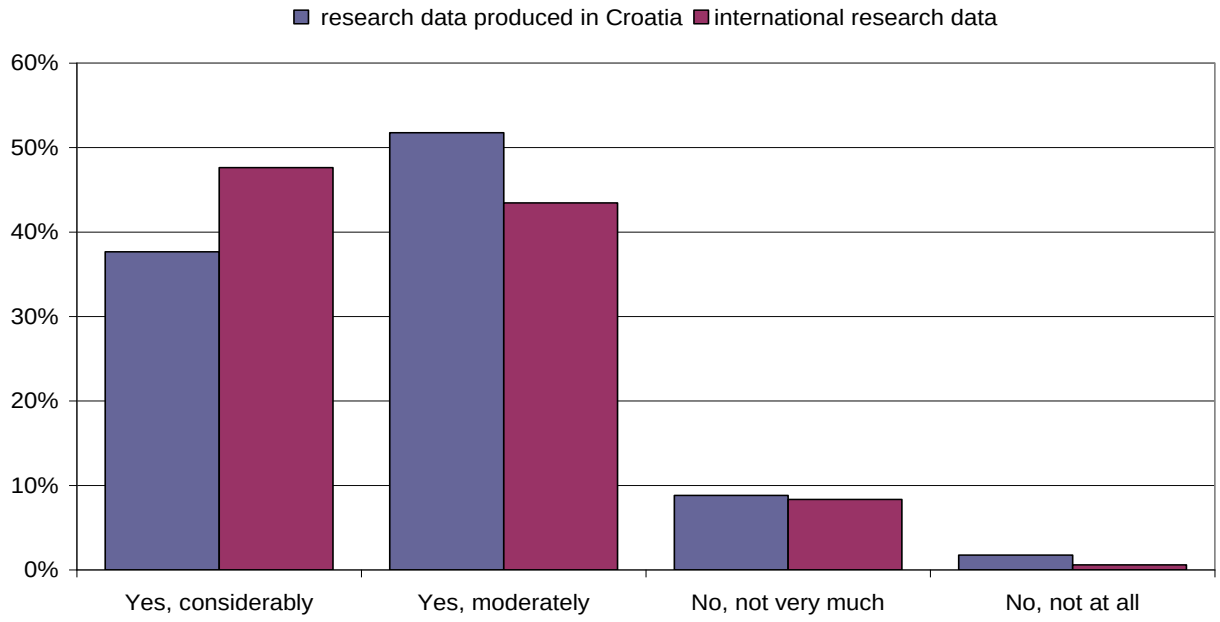
N=171

7. Need for data archive and initial support

Nine out of ten respondents think that their productivity would be significantly or somewhat improved by greater availability of data collected in Croatia, and are even more enthusiastic about prospect of availability of international research dataset. Provision of both falls within the domain of the data archives mission. Additionally, four out of five respondents that are active in teaching thinks that better availability of research data would enhance quality of their classes.

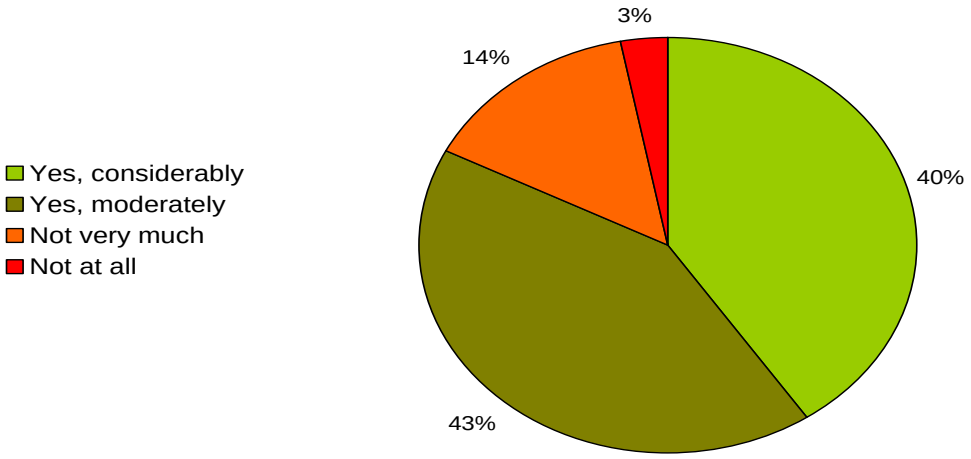
Generally, 60% of respondents have stated that having an institution that specializes in social science data archiving in Croatia would be exceptionally useful and another 33% believes that it would be useful to a certain extent. The opinions that an institution of that kind would make little or no contribution at all were extremely rare. Even if this level of support is overestimated in comparison with one in general population of social science researchers due to self-selection of participants in our survey, it can be hardly disputed that there is a significant need for this type of service among the social scientists in Croatia and a substantial shortage of options for a high-quality archive and access to research data. This is why 73 (39%) participants of this survey have expressed their wish to be contacted again in relation to this project and establishment of the data archives.

Figure 21: Would respondent's scientific work benefit if there was greater accessibility of...



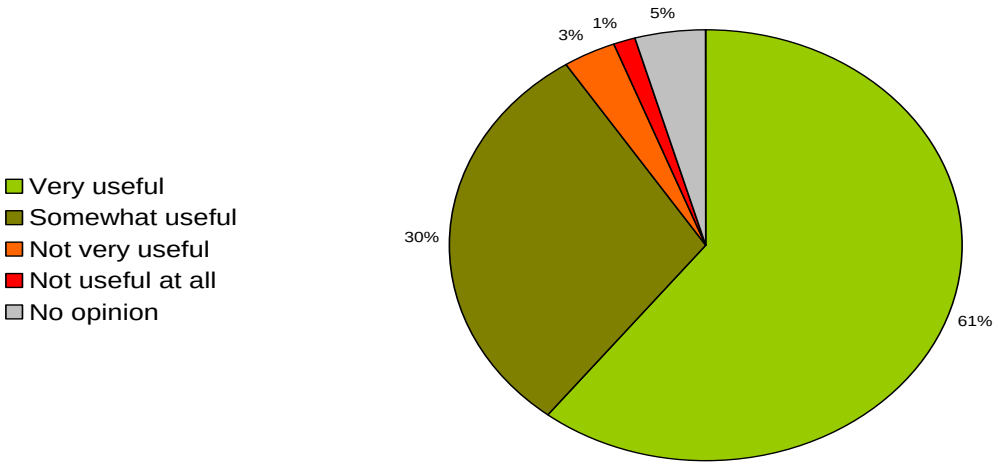
N= 170 (domestic), 168 (international)

Figure 22: Would respondent's teaching benefit if there was better access to national or international data?



N=126

Figure 23: Assessment how useful could be an institution that specializes in data archiving in Croatia


Very useful
Somewhat useful
Not very useful
Not useful at all
No opinion
61%
30%
3%
1%
5%

N=175

Part III: Survey on existing infrastructure and interest for data archiving

The primary goal of this project stage was to get ourselves familiarized with the infrastructure and tools that are already in place and could be used in the establishment of data services, so as to make the best of the existing infrastructural elements.

In early September (institutions in Zagreb) and early October (institutions outside Zagreb) 2012 we have contacted different institutions that may be interested in this project from different perspectives. The list of institutions we have contacted is given in Appendix 1.

In the letter we sent to the relevant institutions we first gave a brief introduction of SERCIDA project and a clear and succinct explanation as to the importance of systematic and organized data archiving. We have also mentioned the possibility of secondary data use and the potential for joining the European network of research data archives (CESSDA), which will enable Croatian scientists to access data created in other countries. We have explained that the goal is to get familiarized with the practices related to archiving and use of research data already in place in Croatia and that these goals should be achieved through use of the existing capacities. As the reason why we have contacted them, we stated that they have been identified as one of those that may participate in the process of establishment of social science data infrastructure. The institutions were asked to reply if interested in this form of cooperation.

The letters were e-mailed and sent in hard copy with the attention to the heads of respective institutions. The exhaustive list of institutions to which the letters were sent is given in the Appendix X. Twelve institutions responded to the letter, of which nine have agreed to be interviewed. We arranged appointments and interviews guide with those institutions that responded in line with the developed interview guide.

One institution has been identified - University Computer Center, University of Zagreb (Srce) as an institution that deals with systematic collection of research data and basically, we identified the elements of required infrastructure in this institution. This was the only institution to which we were in a position to put the questions from the interview guide.

No other institution has instigated similar services or engaged itself in a systematic collection of research data. Interviews in those institutions were conducted for promotional purposes, meaning that our contact with those institutions was seen as an opportunity to inform them in detail about the activities and goals of SERSCIDA projects, as well as about practices, purpose and importance of preserving data and their potential secondary usage. Those institutions were unable to answer most of the questions from the interview guide, which we practically used as a rough outline for the interviews. We tried to identify and instigate interest among those institutions in the data archiving and use of research data. We have also tried to identify modes of cooperation between the future (national) service research data center and institutions dealing with social science research.

For this reason, this report will be divided into four sections. In the first section, the initiative for establishment of data archive in Croatia will be presented. The second section will deal with presentation of the information collected during the interview conducted with the representative of the University of Zagreb Computing Centre - Srce. Most important information obtained at the institutions involved in social science research, or specifically in those institutions where such data are produced and used, will be presented in third section. In the final section we will give a description of several national and international projects dealing with scientific e-infrastructure with which we have established informal contacts and which should become our formal partners so as to explore potential for our future cooperation.

1. Initiative for establishment of national social sciences data archive in Croatia

The initiative for the establishment of research data archive for the social sciences in Croatia emerged at the beginning of the year 2006. The initiative was launched by a few of enthusiasts from various social science fields, who saw the importance and the need to establish the infrastructure for research data services. The initiative was led by a team of people employed in various institutions within the system of science and higher education.

Regardless of the failed initial attempt to institutionalize data archiving services and an unaccepted proposal to the Ministry of Science Education and Sports, through continued activities, the team worked to popularize the idea of systematic research data archiving among local researchers, by presenting the idea of data archive at local conferences[7], and has established important international contacts. With the help of colleagues from University of Ljubljana Social Science Data Archive (ADP), they established connections with Council of European Social Science Data Archives (CESSDA) which resulted in a contribution to CESSDA-PPP project[8]. During that time, the team also managed to gain a small and one-time monetary donation from one commercial donor, and team members had big plans for the future. But the circumstances back then were not suitable, so the data archive was actually never established.

On a practical level, experience has been gained in the preparation of data sets for archiving in a standardized form, based on example of pre-election polls for party elections in 2000 and 2003. These dataset were prepared for GESIS - Leibniz Institute for the Social Sciences.

During the time when the team was active, responsible ministry supported in principle the establishment of social sciences data archives. The team tried to identify appropriate institution where the data archive could be located. After exploring different options they agreed that University of Zagreb Faculty of Humanities and Social Sciences (FFZG), with the new library which was back then in the process of construction, could host the future data archive. For that reason, FFZG is now one of the partners in SERSCIDA project.

2. Infrastructure available at Srce computing centre

Srce ensures stable, reliable, high-quality and advanced e-infrastructure, closely linked with the European one, for the need of academic and research community in Croatia as a whole. There are two elements of the infrastructure that are particularly important for SERSCIDA project:

- authentication and authorization of infrastructures that are part of the system of science and university education in the Republic of Croatia (AAI@EduHr) and
- "Public data collections" service is intended for storing different data collections that are of the public interest and which scope requires considerable data resources.

AAI@EduHr is part of eduGAIN service which ultimate goal is the establishment of the Pan-European web "Single Sign On" (Web SSO) mechanism for GEANT services as well as for web services of individual institutions that are members of partner federations. In line with this, eduGAIN is one of potential

7 Henjak, A., Potočnik, D., Matković, T. i Vodopijevec, A. (2007). Zasnivanje arhiva podataka društvenih i umanističkih znanosti. 9 CARNetova korisnička konferencija. http://cuc.carnet.hr/cuc2007/program/radovi/b2-7-bio.html

8 Widening CESSDA: Inclusiveness andComprehensiveness of the upgraded Europeanresearch infrastructure. http://www.cessda.org/project/doc/D7.1_Report_and_Recommendations.pdf

authentication-authorization solutions for the future national/regional archives that will enable regulation of access to information at the level of the overall European research community.

"Public data collections" service provided by Srce

Existing infrastructure and data capacities

Srce offers a public data collections service. This service is used for storing primary scientific data collected as part of measurements and research, including different educational multimedia contents, archived data collections and similar and its amount requires considerable data resources. This type of data is considered to be of public interest for the scientific and educational community and the public as a whole. Public data collections are accessible to all scientific fields, meaning that they are not restricted to a specific field. All members with AAI@EduHr user account and those gathered around the so-called virtual organization are free to use this service. They presently have 6 data collections.

Data preservation

Srce enables and is responsible for permanent (long-term) availability and/or archiving of data stored in the Public data collections. Data archiving is done over semi-annual intervals and two copies are made for the archives and stored at separate locations where they are kept for 10 years. In case the user deletes his/her data, the data are kept on the tape for 10 years. Srce ensures that a proper environment is in place for safekeeping of data and regulates the preservation of users' data pursuant to the "Rules on terms of use of data services". There are internal rules that govern back-up procedure of Srce data. These internal rules are neither public nor available.

Data documentation and standards

Srce does not deal with data processing of any kind (including documentation making, data description, modifications or dissemination), Srce does not regulate data formats – this is all left to the data owners. The data owner must describe the data by metadata sets defined by Srce. Dublin Core is used as standard for metadata. Data are not kept based on a certificate, meaning that no standard is used to specify the data protection level, instead this is done by internal rules and procedures. No OAIS or Data Seal of Approval are used. Srce does not take part in international projects requiring harmonization of data groups enabling their cross comparison. Srce does not use Thesauri for conversion/mark up/indexing of data.

Tools and technology

For their data services, they use a tool developed by Greek academic community, but are in the process of identifying a new tool, given that the Greek model is no longer being developed. Tools like NESSTAR, FEDORA or Dataverse are not used.

Data discovery and dissemination

Data are accessed via web interfaces and different clients (WebDAV, MS Windows, mobile phone). Data access is regulated by each individual data owner. Collections are organized as virtualcommunities where physical users can read and write documents. One of the virtual organizationmembers is responsible for the collection contents and its representation. One virtual organization can be the owner of several data collections. Only metadata on research data must be public. This description is mandatory present on a public service portal, thus enabling it to become 'public', in other words, its existence and description becomes revealed to all those that may be potentially interested in its use. Srce encourages data collection owners to enable access to their collection wherever possible. Possession of AAI@EduHR user account is a pre-requisite to access such services and data. There are no

specific restrictions on different levels. The data can be accessed from abroad only if researchers have an AAI@EduHR account and if the data are public. Data are disseminated via web interface only.

Data policy and service funding

All users must agree with the rules prescribed in the Rulebook on Terms of Use of Data Services[9] before they start using the service. Srce does not assume the ownership over data. The public data collection service is financed from Srce funds, specifically its current business operations earnings are used to ensure continuity of service. The issue of service development and expansion has not been systematically resolved, although they plan to work on it. Srce has a small number of users who are interested in storing their data. Their tendency is, of course, to increase the number of its users.

Staff capacities

Srce has one person that is in charge of system maintenance for all data services – a system engineer and head of data services. Both persons perform other tasks from within their purview. The staff did not undergo proper educational training in data preservation and dissemination relative to those services. The party interviewed recommended that experienced archives officers working abroad propose the type of training they should undergo. They also need to be trained in Croatian and the EU legislative framework for creation of such services.

About possible National Data Service for the social sciences

Our interviewed parties were of the opinion that there is a need for storing structured data in Croatia, providing that they did not give consideration to any features specific to social sciences or other fields for that matter. Our interviewed party was unable to give us a rough estimate as to how much data collected in Croatia ends up lost due to a failure to preserve such data in a safe environment. The interviewed party was also not familiar with any features specific to social sciences but was satisfied that is important to establish national data archives for social sciences. Our interviewed parties were of the opinion that the national archives should enable data storing, implementation of software application at a national level and make use of the existing resources. Srce has experiences in several projects where one person could work for Srce or some other institution and cooperate on joint projects.

3. Interest for data archiving and use of research data among the institutions involved in research activities

Nine interviews were conducted with representatives of eight institutions (five faculties, two institutes and national library) involved in research activities. Those institutions are involved in research activities in the fields of economics, econometrics, international relations, culture and communication, resource economics, politics, media studies, information sciences and educational sciences. Five interviews were conducted on the premises of the interviewed parties and the rest of the interviews were conducted by telephone. The greatest turnout was recorded among the institutions dealing with economics related sciences (four institutions).

The responses obtained at the institutions that agreed to be interviewed were grouped and summarized under the following headings.

Methodology and type of data produced and used

9 The Rulebook on Terms of Use of Data Services http://www.srce.unizg.hr/fileadmin/Srce/proizvodi_usluge/podatkovne_usluge/JKP/Pravilnik_Podatkovne_Usluge_v1_4.pdf

In the social science field quantitative methodology is mostly represented, whereas qualitative methodology is also present but in a considerably lesser degree or only as of quite recent date.

Institutions in the field of economics use much more ready-made data in their research than they produce their own and they constantly voice their need for relevant international and national data bases that contain such ready-made data. They have access to such data either by being a subscriber as an institution or by procuring a set of data required for a particular project. They use ready-made data by supplier, which, naturally, is in the ownership of the ordering party. The ordering parties can come from a commercial sector or public institutions like ministries, national institutes and other institutions.

Some of the data are not even raw data but simply reports and partially processed data. Institutions in other areas mostly produce their own data.

Data preservation

Researchers mostly keep data on their personal computers. Some institutions have some space on their own server where they keep a back-up copy of the data. It is difficult to assess how much data is lost both on institutional level or Croatia as a whole, but in the long run all the data are probably lost given that improper way of data preservation. At times, not even much thought is given to this issue and some do not find it to be important because the data become outdated. Not much effort is made on preserving data collected in social science research.

About a possible National Data Service for social sciences

Majority of institutions have agreed that it would be useful to establish the social science data archives although not all of them have recognized the need for such services. The following was offered in support of establishment of the national data archives:

- there is high demand for data;
- there is awareness about secondary use of data: some data are useful for several research projects (avoiding duplication of work) or for comparison with previous research;
- availability of research data would enable greater credibility of research work and growth of professional and scientific ethics;
- would be useful for educating students;
- data should be kept in standard form;
- all countries in Europe have such archives, so should Croatia;

Regarding the key functions and services of the data archive, the following responses were given:

- receiving and managing data;
- enable the option of regulated access to data;
- data should be searchable;
- ask questions option (user supprot)
- download of data in different readable (usable) formats;
- organization of data by subject fields, browse data by fields (make the entire content easily visible and user-friendly); thesaurus of good-quality
- describe data by metadata;

- usable software for data preservation and access;
- the data services staff should work on popularization of data archiving (lobby the researchers);
- the data services staff should educate researchers on the topic of data management (an author himself should draft data-related documentation);

As for the mode of organization of the future data archives, our interviewed parties provided the following responses:

- a central point is required where data would be collected;
- it is not advisable for the archives to be at the level of one institution because this would be counterproductive in terms of data exchange;
- some institutions lack capacities for setting up of the archives; the data should be taken care of and such systems should be regularly maintained;
- there is a need to create a network, co-operation with institutions at which data are produced; need to interconnect field-specific institutions (economics);
- experts should be put in charge (computer, information experts ...); librarians have potentials for future cooperation; there should be one person to act as a link between scientists in a particular institution and the central archives and as support to local scientists; employees of NUL have experience in development of different standards for documenting of archives material/data;
- some form of an advisory expert group that would gather experts in different fields is required; the group would be in charge defining standards and rules, scope of the archives, follow-up of global trends
- there should be staff permanently employed; the state must pledge to finance this activity; avoid current practices associated with national projects – where a service is instigated as part of the project but no one is there to take care of it afterwards;

No repeated invitations were sent to the non-responding institutions because we were operating under the assumption that at this juncture those institutions are not interested in this issue for different reasons. It could be that our expectations were not clearly communicated, but it is more likely to assume that at this point some institutions simply have no sufficient interest, information or knowledge on this issue or lack capacities for cooperation we mentioned in the letter. This assumption is to a certain point supported by the fact that two institutions contacted us after they received the letter asking to be informed in greater detail what this is about. We have not heard from one of them since and the other one informed us that they have no person that we could talk to about potential future cooperation, and/or that they have no capacities for such cooperation, but that there are several persons employed with the institution that could be given the task of filling the data base in future.

Those non-responding institutions will be contacted again and informed about this project stage results. Those institutions will also be invited to participate in all future activities related to data archiving and secondary use of research data that will take place within or outside the project. The first activity of such kind was a round table discussion on the topic "Archiving and Use of Research Data in Social Sciences - Scientists' Needs, Existing Infrastructure and Shaping of Science Policies", that is organized as part of SERSCIDA project on 14 November 2012.

4. Potential cooperation with other institutions and existing e-infrastructure projects

The European Commission is currently financing several projects that are to a greater or lesser degree related to the research data archiving and open data access. It is of paramount importance that SERSCIDA project team also gets actively involved in the follow-up of the activities undertaken as part of those projects so as to avoid duplication of efforts in the planning and development of particular infrastructures and enabling interoperability of new systems.

In addition, there are several data bases and e-infrastructure services (bibliography, scientific journals, and scientists) that the future system will have to bring together because our final goal is to have a comprehensive service for research community and the public in general. Cooperation in the area of interoperability with those services means connecting research data with publications (integral texts and/or bibliographical entries) and scientific profiles.

Potential cooperation with institutions, existing web services and existing e-infrastructural projects can be divided into several points:

- Cooperation with institutions dealing with similar activities (archiving of digital material that could potentially be used in social research, production and dissemination of statistical data):
    - National and University library
    - national archives
    - Croatian Bureau of Statistics
- Connection with national services for academic and research community:
    - Croatian Scientific Bibliography - CROSBI
    - Who is who in science in Croatia
    - HRČAK - portal for Croatian Open Access Journals
- Cooperation with DASISH project that incorporates 5 initiatives of ESFRI research infrastructures in order to identify potential synergies and joint planning of a functional future system:
    - CLARIN – Common Language Resources and Technology Infrastructure
    - DARIAH – Digital Research Infrastructure for the Arts and Humanities
    - CESSDA – Council of European Social Science Data Archives
    - ESS – European Social Survey
    - SHARE – Survey of Health, Ageing and Retirement in Europe
- Cooperation with project OpenAIRE Plus working on connecting the peer-reviewed publications with the original research data.

Potential modes of cooperation with the National and University Library

In the Natioanl and University Library's (NUL) collections there exists material that can be used as primary data in some social science research. Such material could be processed in the way that its research potential is stressed and the research data archives and NUL could work together on increasing the visibility of such material.

NUL maintains the Digital Academic Repository that brings a collection of digitized old PhD theses (1880-1952.) and bibliographical entries related to the theses of the University of Zagreb and available

biographical data on the holders of the referenced PhD theses. The Repository preserves contemporary PhD theses and scientific master theses. The data on publications in the Repository can be linked with the data in research data archives.

In addition, NUL also disposes with the research data left as legacy by the Croatian scientists. Such data are not digitized from paper forms. It is assumed that such data could be found in any scientific field given that NUL covers the materials from all scientific fields, but only thorough research of the scientific legacy would give precise information on the volume of such data. Primary data of this kind could also be found on CDs, floppy discs and similar that is integral part of the legacy left by particular scientists. One form of cooperation with NUK could be set up as part of a separate project which goal would be to first identify the material presenting research data among the scientific legacy material and after such material could be digitized and preserved in the research data archives.

NUL employs experts with experience in developing different standards for bibliographic records, and experts dealing with subject headings, or specifically with thesaurus development. NUL employee have demonstrated an interest in participating at the round table discussion on the topic "Archiving and Use of Research Data in Social Sciences - Scientists' Needs, Existing infrastructure and Shaping of Science Policies" that will be hosted on 14 November 2012 as part of SERSCIDA project. People from NUL are open for different forms of future cooperation.

Potential forms of cooperation with the national archives

National archives, just like NUL keep in their collections material that can be used in research and for this reason we may see it as an institution dealing with a similar activity. National archives prepare and maintain guides through holdings and collections where many collections are represented as a source and material for research. Employees of State archive in Zagreb (SAZ), which is a member of the network of government run archives, have shown interest for the activities that are ongoing as part of SERSCIDA project. They invited SERSCIDA project representatives to take an active participation at the conference "4th Zagreb Day of Archives - archival science - culture - science: partnership for development" hosted by SAZ on 30 November 2012. This conference covers topics related to partnership with the museums and libraries and cooperation between the archivists, historians and other scientific researchers. Likewise, SAZ employees have expressed their interest in taking part at the round table discussion on the topic "Preservation and Use of Research Data in Social Sciences - Scientists' Needs, Existing Infrastructure and Shaping of Science Policies" that was organized as part of SERSCIDA project on 14 November 2012.

Services of the Croatian Bureau of Statistics

The Croatian Bureau of Statistics (CBS) is one of the leaders in the statistical research and a key player in dissemination of statistical data in the Republic of Croatia. It operates pursuant to the Official Statistics Act (Official Gazette Nos. 103/03, 75/09 and 59/12).

Identical letters were sent to the CBS and other institutions where they were asked to send us a response in case they are interested in cooperation so that we can arrange a meeting. Due to other engagements, representatives of this institution were not in a position to meet with us in September or June, and they sent a written response instead with description of the activities they carry out as well as a note that these services are offered to all users interested as far as the publicly available data are concerned and they also described the requirements that a user must meet to access confidential data.

The services provided by the CBS, publication and availability of data to all interested users are carried out through a standard or special processing. Standard processing pertains to the processing of data

that are published in the publications of the CBS with online availability. The published data can be accessed online for free and on the premises of the Institute's library. Special processing entails processing that is done in line with the specific user's needs. This service is chargeable. In addition to the processing of publicly available data, the Institute can prepare confidential data at users' request, providing that those users satisfy the necessary pre-requirements. Such data can be delivered to scientific and medical institutions and users are obliged to provide a certificate from the Register of Scientists and/or scientific institutions that is kept by the Ministry of Science, Education and Sport, stating the title and purpose of the project for which the data is requested with a note that the data will be used for the purposes specified therein. Those users (physical or legal person) that satisfy the terms requested then enter into a contract with the CBS. Following the signing of the contract and payment of the service the data is delivered to the user usually by e-mail or on an optical media (CD -ROM).

To conclude, at this project stage, we have established initial contact with the CBS informing the Institute in brief about the goals of SERSCIDA project. Subsequent project stages will be used for discussing modes of cooperation between the CBS and the future social science data archives.

Round table "Archiving and Use of Research Data in Social Sciences - Scientists' Needs, the Existing Infrastructure and Shaping of Science Policies"

The first round table on research data in the field of social sciences was held on 14 November 2012, immediately after the initial contacts and discussions with relevant participants of the process of creation, usage and archiving of research data. This round table has been announced during these initial contacts and discussions. The main intention was to summarise and encompass findings from previous activities and to get more feedback from relevant stakeholders who could or should be interested in this issue from different perspectives - data producers and users (individual researchers, institutions engaged in the research), institutions that make an integral part of a scientific infrastructure (Srce computing centre, libraries, archives) and bodies dealing with a development and quality of science system. In addition, the intention was to achieve continuity in communication with them.

The subject of this first round table was rather broad. The intention was to deal with the most important issues that the SERSCIDA project dealt with during mapping of the existing potentials for establishing the archive of data for social sciences. Therefore, the round table was titled "Archiving and use of research data in social sciences – scientists' needs, the existing infrastructure and shaping of science policies".

Scientists' needs were presented based on results of a survey on production, preservation and use of research data, among scientists in the field of social science, that have been described in the first part of this report. The results of this survey were supported by a personal story of an experienced researcher in the field of sociology. Services and activities of the Srce computing centre have been presented in the part dealing with the existing elements of sub-data infrastructure and have also been previously described in this report. Experiences of creating research infrastructures in the humanities have been presented too using the CLARIN project example. Bearing in mind that systematic care for the research data and encouraging scientists to archive and exchange data can improve the quality of science and higher education, it was important to hear the opinion of representatives of central bodies in charge of development and quality of science activities. The President of the National Science Council talked about shaping of science policies. Representatives of the Agency for Science and Higher Education also participated in the meeting.

50 various participants from 15 different institutions attended the meeting (including keynoters, that is, the institutions they came from). Several relevant issues have been discussed during the conversation. The participants were given a suggestion of conclusions of the round table and in this manner the final conclusions will be reached subsequently through interaction with participants who participated in the discussion. The participants agreed that this is an important subject and that it is necessary to organize in the future several more similar meetings or workshops that could, with more focus and details, deal with subjects opened in this first round table as well as other subjects that appear as relevant.

Here we present some issues brought out by participants of the round table, and a more detailed report as well as final conclusions shall be published subsequently on the web pages of the SERSCIDA project.

An issue of a long-term sustainability of data services

A long-term sustainability of data services at the national level can be ensured in such manner that the state budget provides assets for a permanent financing of such activity. The main institution through which this can be achieved is the Ministry of Science, Education and Sports (MSES). The basis for this is recognition of the importance of data archiving into science policies via relevant documents. The MSES has founded a Commission for Science and Infrastructure that has to develop a strategic plan (roadmap) unit. The Croatian Parliament has created a national, coordinating body and a national, operational body

for creating an educational, scientific and technological strategy. These bodies can influence shaping of science policies. The National Council for Science (NCS) has also an important role in shaping science policies since it is in charge of the so-called institutional infrastructure that is, creating rulebooks and defining criteria. The NCS can provide a support for this kind of activity, for allowing scientists to access data can influence the quality of the scientific work.

Data archiving as one of the criteria for evaluation of scientific work

Without a stable, secure, clear and well-defined infrastructure for data archiving, it is not possible to prescribe a mandatory archiving of data upon a completion of a project and/or take data archiving into account as one of the criteria for evaluation of scientists or institutions. The competent authorities for now are not prone to imposing such obligations on scientists, especially because of the current situation in which scientists have problems accessing data (and literature) from other sources needed for their work. Synchronization is needed between what is mandatory and what is available to the scientists.

On the one hand, prescribing an obligation for data archiving could ask for a greater responsibility on the part of the relevant ministry, to perpetuate, that is, to provide funds for further exercise of this activity. Such an obligation would encourage more responsible behaviour towards public funds, both on the part of the scientists and competent institutions. Without the obligation for data archiving or possibility to get additional points, because of data archiving, upon evaluation, a chance to fill the archives with a critical amount of data, that would be useful for researchers, decreases and when this happens scientists themselves shall easily recognize the importance of data archiving.

A problem of research instruments protected by copyright

Participants noticed that since the data has been gathered based on research instruments, without preservation of the instruments themselves, the research data could be useless or insignificant. When talking about instruments, one often gets into the field of intellectual property. If there is a decision at the state level for mandatory data archiving, it is necessary to clearly define what to do with the instruments and to give an answer to the question if, and under what conditions can such an obligation be applied to instruments too? During this project and in the future when the archives will be established, it is necessary to know the law regulations in detail and to educate authors on the possibility of a legal data archiving and use of data and instruments. Also it is necessary to be familiar with international practices regarding research instruments and to learn about best practices.

The issue of intellectual property over the instruments is more stressed in the field of psychology. However, there are different trends which recognize the value of using the existing instrumentation distribute their instruments freely.

Access rights

A free exchange of any scientific information, data included, is important not only because this can be an expression of a responsible behavior toward money of the taxpayer, if the research is financed by public funds, but also because sharing of data represents the core of the science and development of scientific knowledge. However, it is necessary to stress that, when talking about the need to archive data, this does not have to imply a free access to that data. This means that, if the sponsors of scientific projects would oblige scientists to archive data, that does not have to be done in such a manner for the data to become available to the public immediately. A regulation of access is a special subject to be

dealt with from within the system for data archiving. The data access right is granted by the author, that is, a person that is the data owner.

Which data should be included in the archives?

One can ask a question if it is necessary to archive all produced data given the assumption that some data types are not appropriate or useful for the secondary analysis and that researches representing nations, with a clearly defined methodology and instruments, are suitable for this type of data base and for later usage.

However, upon archiving at the national level, one should not judge in advance if something is representative and useful or not. It is necessary to provide a great amount of documents for data in standard way, in order to correctly identify the type of the sample and potential problems with a certain dataset. It is then possible to decide from these documents the suitability of the data for further analysis.

Issue of archiving qualitative data

There have attempts at establishing a repository of quality data for the requirements of an international project, but such bases have been useless in generating data meaning. In spite of that, it is possible to identify a need for archiving of qualitative data and to imagine possible uses of such data, for example, together with language technologies that are being developed within projects dealing with the scientific infrastructure in the field of humanities.

Long-term data preservation and relations with the national archives

Research data that is significant for public, in terms of archival practices, is regarded as a material that shall reach national archives in the future. National archives are public institutions and therefore, all users have the right to use the archives' material under equal conditions. If the material in question is a digital material, it is necessary to consider a possibility to transfer data in order to ensure its long-term preservation.

Summary and conclusions

Interest of science policy and funding institutions for data archiving

The Ministry of Science, Education and Sports (MSES), is the main administrative body responsible for development and functioning of the whole scientific system in Croatia. Together with the Government of Republic of Croatia that sets the basic principles for development of science and higher education, it plays a major role in establishment of national social sciences data archive in Croatia. Other institutions and professional bodies such as National Council for Science, the Agency for Science and Higher Education and Croatian Science Foundation can provide support for such activities by expressing the importance of data archiving and implementing criteria related to data archiving and by defining criteria for evaluation of scientists, scientific institutions and scientific projects related to data management and archiving. All these major institutions/bodies dealing with development and quality of science in Croatia expressed their support for SERSIDA project. However, if implementation of criteria for evaluation related to data archiving is not possible without stable and functional infrastructure for data archiving.

Recently, two major bodies for the development of science policies were established. The MSES founded a Commission for Science and Infrastructure that was tasked with development of a strategic plan (roadmap) unit. The Croatian Parliament has established a national coordinating body and a national operational body to create an educational, scientific and technological strategy. The funding data archive in the future remains uncertain due to a general trend in recent years of reducing for science and projections of budget lines for next three years.

Researchers' interest for data archiving

Based on a survey conducted among scientists, it is evident that researchers in Croatia store their research data. However, this data is often very poorly documented and kept mostly on individual computers with several copies on different computers and media. While access to data is usually allowed only to the project manager and project team members, the researchers consider this not to be a desirable state of affairs, with almost half of the respondents believing that their data should be made available either to the general scientific community or publicly available. In general, researchers express their readiness to archive their data in a specialized data archive for the social sciences. Social scientists in Croatia consider important to have access to research data collected by others, but in practice, secondary analysis is rare, and researchers often do not even know whether their data were used. In those cases when secondary analysis was performed, the data sources are mostly informal and professional social networks, project websites and the Croatian Bureau of Statistics. For the main obstacle to carrying out secondary analysis, the researchers suggested the lack of availability of the existing data. The researchers also stated that data sharing was not part of the research culture. The researchers believe that better data availability, especially for international research, could contribute to their own scientific work, and can improve the quality of teaching. The need for establishment of national data archive for social sciences is widely recognised among researchers.

A social science data archive is required because the researchers in this field are interested in what happens over time. Without comparative research it is difficult to know what is really going on in society, and impossible to create a complete picture of the social phenomena. Also, the data often remain unused because the researcher who collected the data moves to new problems. Re-use of the same data allows for greater utilization of resources invested into research. The available data has great potential in the education of students as handling the real data is more motivating for students than just

listening to theory. If there was a requirement to archive data within a ceratin period after the project has ended, this would increase the level of responsibility of researchers and force them to be more systematic, both of which can indirectly improve their research quality.

Existing research infrastructure for data archiving

The Croatian academic community has a well-established network and storage infrastructure. The University of Zagreb University Computer Centre (Srce) ensures a stable, reliable, high-quality and advanced e-infrastructure, closely linked with the European infastructure, for the needs of the Croatian academic and research communities in Croatia as a whole. There are two elements of the infrastructure that are particularly important for SERSCIDA project:

- authentication and authorization infrastructures that are part of the system of science and university education in the Republic of Croatia (AAI@EduHr) and
- “Public data collection” service is intended for storing different data collections that are of the public interest and the scope of which requires considerable data resources.

AAI@EduHr is part of eduGAIN service, whose ultimate goal is the establishment of the Pan-European web “Single Sign On” (Web SSO) mechanism for GEANT services as well as for web services of individual institutions that are members of partner federations. In line with this, eduGAIN is one of potential authentication-authorization solutions for the future national/regional archives that will enable regulation of access to information at the level of the overall European research community.

Regarding the existing "Public data collections" service more work needs to be done on the software application level as well as on the user services level.

Recommendations and future steps

- Establish closer contacts with the Commission for research infrastructure as well as with national bodies for creating an educational, scientific and technological strategy. Make sure that data archiving is well recognised in relevant policy documents.
- Continue and intensify cooperation with Srce computing centre, since this institution has already started building basic blocks for the data archiving infrastructure (network, storage and authentication/authorisation infrastructure) and has been recognised as the one of key partners of the Ministry of Science, Education and Sports in the development of e-infrastructure.
- Identify people who will go to the conference in Belgrade planned within work package 3 (WP3) as well as people who will go to working visits and undertake the training planned within work package 4 (WP4). As for the conference and working visits, people from policy making institutions and people from research and infrastructure communities should be chosen. People who will continue to work on establishment of data archive in the future should attend the training.
- In order to function properly, a data service will need permanent staff, including a data services specialist and a software developer. During the SERSCIDA project, proper staff education will be provided during the work package 4 (WP4) and considerable amount of money will be invested into this area. To justify this investment, MSES should provide support for this project in the future.
- Continue to explore options for establishing the social sciences data archive. At this stage of the project, it seems like future data archive in Croatia should be organised on a national level. Social science data archive could be established as a separate organisational unit at University of Zagreb Faculty of Humanities and Social Sciences (FFZG) and should work in strong cooperation with Srce computing centre, because both of these institutions have recognised potentials and could share available resources. Keep in mind that the data archive should be defined and perceived as an institution which provides nation-wide services. Srce is already strongly profiled as an institution with influence on a national level.
- Build cooperation with all relevant higher education and research institutions in Croatia towards building a network of people who will work closely with researchers employed at these institutions. Members of this network could provide local support to scientists, facilitate data archiving and data sharing, and serve as mediators between central data archive and local institutions.
- Build cooperation with other similar projects mentioned

Annexes

1. List of interviews held

institution	person	date and time	visit or telephone
The Institute of Public Finance [Institut za javne financije], Zagreb	Martina Fabris, librarian Irena Klemenčić, lawyer	11.9.2012. 10:00	visit
Institute for International Relations [Institut za međunarodne odnose], Zagreb	Snjezana Ivanović, librarian	11.9.2012. 15:00	telephone
National and University Library [Nacionalna i sveučilišna knjižnica], Zagreb	Tinka Katić, librarian Sofija Klarin, librarian	13.9.2012. 10:00	visit
Faculty of Economics and Business, University of Zagreb [Ekonomski fakultet Sveučilišta u Zagrebu]	Silvana Brozović, systems librarian	20.9.2012.	visit
University of Zagreb University Computing Centre [Sveučilišni računski centar Sveučilišta u Zagrebu (SRCE)]	Damir Danijel Žagar, head of data services	25.9.2012. 10:00	visit
University of Zagreb Faculty of Political Sciences [Fakultet političkih znanosti Sveučilišta u Zagrebu]	Andrija Henjak, researcher	29.9.2012 15:00	visit
University of Split Faculty of Economics [Ekonomski fakultet Sveučilišta u Splitu]	Maja Fredotović, vice-dean	9.10.2012 10:00	telephone
University of Osijek, Faculty of Philosophy, Department	Kornelija Petr-Balog, head of department	11.10.2012. 13:00	telephone

of Information Science [Filozofski fakultet u Osijeku, Odsjek za informacijske znanosti]			
University of Osijek, Faculty of Philosophy, Department of Psychology [Filozofski fakultet u Osijeku, Odsjek za psihologiju]	Silvija Ručević, head of department	11.10.2012. 14:00	telephone
The Croatian Science Foundation [Hrvaska zaklada za znanost]	Lovorka Barać Lauc, executive director	17.10.2012. 10:00	visit
University of Zagreb, Faculty of Education and Rehabilitation Sciences [Edukacijsko-rehabilitacijski fakutet Sveučilišta u Zagrebu]	Danijela Bratković, vice-dean	26.10.2012. 10:00	visit

Annex 2. Questionnaires

Annex 2. A Questionnaires Researchers' questionnaire

What is your current principal activity?
1. Undergraduate student
2. Doctoral student / research or teaching assistant
3. Researcher / professor
4. Project leader
5. Head of institution
6. Other

With what type of institution are you currently principally affiliated?
1. Higher education institution
2. Research institute
3. Currently not employed
4. Other
(if not student)

Do you work in public, private or nongovernmental sector?
Public sector
Private sector
Nongovernmental sector

What is your principal research discipline?
Teacher training and education science
Humanities
Law
Psychology
Sociology and cultural studies
Political science and civics
Economics
Journalism and reporting
Library, information, archive
Social work and counselling

Reseach participation during past 5 years
Yes

Number of datasets the researcher had produced or helped to produce during the past 5 years

Year the most recent fieldwork (or data collection phase) completed?

Which data collection method was applied in this research?

How was this research financed?
Public funding through national science funding bodies (science ministry, science foundation...)
Public funding from other sources (other ministries, state agencies, cities and municipalities...)
International funding/project
Private sector
Own funding (institution you are working in paid from its own funds)
Other source
Don't know
Personal effort/funding

Number of co-researchers involved in given research project
0
1
2
3
4 to 5
6 to 10
more than 10

After the last research project, did you or your research team members retain the data?
Yes

What kind of data was kept?
Raw data
Cleaned data (coded, anonymised, ..)
Data prepared for analysis (with transformations, created indexes, recoded)
Other kind of data
Well documented with metadata
Don't know

Where is the data from your last project kept?
On my computer
On my colleague's computer
Several copies on different computers and/or different media
Server at my local institution/university
Data archive/repository
Somwhere else
Don't know

Who may be granted access to the data from your last project for research use?
Just the project leader
Research team members
Members of my institution
Broader scientific community
The data is publicly available (open access)

Someone else
Don't know

In your opinion, what would be the ideal level of access to this research data?
Just the project leader
Research team members
Members of my institution
Broader scientific community
The data should be publicly available (open access)
Someone else
Don't know

Willingness to provide your data to a social science data archive?
1. Yes, certainly
2. Yes, probably
3. Not sure
4. No, probably not
5. No, certainly not

With respect to your own discipline, how important is the sharing of research data?
1. Very important
2. Somewhat important
3. Not very important
4. Not important at all
5. No opinion

Do you know if any other researcher outside your own team had used for secondary analysis any of the research data that you produced?
Nobody outside my team had ever used research data that I/we have produced
Yes, I know of occassion when my/our data was used, but this was more than a year ago
Yes, I know that my/our research data was used for secondary anaylsis recently, the most recent occasion being within the past year
I am not sure if anybody used my/our data for secondary analysis

When was the last time that you analysed quantitative data that were not produced by yourself or your research team.
this month
last month
2-5 months ago
6-12 months ago
13-24 months ago
more than two years ago
never

When was the last time that you analysed qualitative data that were not produced by yourself or your research team.
this month
last month

2-5 months ago
6-12 months ago
13-24 months ago
more than two years ago
never

In general, what are the barriers to conducting secondary analysis in your country?
Not enough relevant data exist
Data exist but are not accessible
Data exist but are poorly documented and unusable
Researchers are not trained well enough in secondary analysis
It is not part of the research culture
Something else
Don't know

Please indicate all the sources that you ever used sucesfully to obtain such data
Your own research unit
Network of colleagues outside of your research unit
Your own institution
Data archives from other countries
National Statistical Office
Websites of projects (national or international) providing access to data
Other source

What statistical software do you commonly use for your quantitative analyses?
Excel
R
SAS
Stata
SPSS / PASW
Other (statistica, eviews, amos)
I don't do quantitative analysis

What software do you commonly use for your qualitative analyses?
I don't do qualitative analysis
I don't use any software for qualitative analyses
Atlas.ti
NVivo
MAXQDA
QDA Miner
CAT
RQDA
Dedoose
Other (AntConc. Tosmana, SPSS)

Would your scientific work benefit if you had better access to...
research data produced in your country
international research data

In your opinion, what is the prevalent attitude with respect to sharing one's own research data among...
you personally
your research team members
coleagues from your institution
coleagues from your field of science

Does your professional activity include teaching responsibilities?
Yes, No

In the context of your teaching, how often do you analyze or discuss research data?
Regularly
Sometimes
3. Rarely
4. Never

Which type of data do you use in your teaching?
Data collected by students through the coursework
Data from past projects in which you have participated
Publicly available data and datasets (e.g. international surveys)
Artificially generated data or datasets included with the textbooks/software

Would your teaching benefit if you had greater access to more national or international data?
1. Yes, considerably
2. Yes, moderately
3. No, not very much
4. No, not at all

In your view, how useful could be an institution that specializes in data archiving in your country?
1. Very important
2. Somewhat important
3. Not very important
4. Not at all important
5. No opinion

Please leave your email if you wish to participate in the SERSCIDA project activities

Annex 2. B Questionnaires Guidelines for semi-structured interviews with government counterparts and potential

Research policy setting: interview with policy makers/funders

PREPARATION

Stakeholder identification

- who is involved in making science policies?
- what are the main sources of financing scientific work?

Reading relevant documents, domestic and international

relevant documents:
- current documents o znanstvenim poliotikama
- project funding criteria and rules
- criteria for evaluation of scientific work

- identify everything concerning science infrastructure and data archiving in relevant documents
- are we one of the members of ESFRI? (Croatia and Serbia are associated members)

http://ec.europa.eu/research/infrastructures/pdf/esfri/membership/esfri_membership_july_2011.pdf#view=fit&pagemode=none
http://ec.europa.eu/research/infrastructures/pdf/esfri/home/esfri_inspiring_excellence.pdf#view=fit&pagemode=none

Good data source (not all infomation are up to date)
http://erawatch.jrc.ec.europa.eu/erawatch/opencms/information/country_pages/

Invitation letter

- maximum one page
- introduce the project and yourself, explain what is data archive
- aks for meeting to talk about relevant issues on establishing data archive
- attach topics for disscussion and questions (so that they can prepare themselves

TOPICS FOR DISCUSSION

A. Needs

B. Current policies and activities, plans for future

C. Possible institutional solutions
D. Financing

QUESTIONS

A. Have you met the need for archiving of research data in the social sciences?

Do you rely on the results of research in social sciences when proposing/designing scientific policies and for other activities? If so, can you give some examples?
In this process was there a need for examining the raw data obtained in these studies?

Have you ever used the online available source data or the results of research (eg Eurostat)? What data sources did you use?

Have you met the requirements of the users / authors of the study (scientists and researchers) to archive research data in order to preserve and use them in future research?

B. Is there something in the scope of your current policy and actions with regard to data archiving in the social sciences? What are the plans for the future?

What is the current science policy in relation to the archiving of research data?

Are there any requirements related to research data when financing projects or other activities? (eg if funding publishing scientific journals, criteria could include and requirement related to research data arhcihivng)

Is there a policy relating to open access to research results?

If there is nothing currently:
- Why is not there?
- Can this be changed and how?
(assumption is that If there is need, and no current activities, there must be plans for the future)

C. What do you think how an institutional solution for data archiving should be established?
what kind of isntitution should carry this task? (scientific instutute? faculty? university? library?)
centralized, distributed?

D. What are the possible sources of funding for activities related to the collecting, processing, use and disemnination and permanent preservation of research data?
available or prospective (EU)

Annex 2. C Questionnaires Data Services

Data Service Infrastructure Survey 2012
SERSCIDA

SERSCIDA is a project funded by the European Commission to study the conditions of social science research in Balkan countries toward establishing durable data infrastructure at the service of researchers in the region.

We are particularly interested in existing work within [*name of country*] with respect to the long-term preservation of research data and the potential for a national social science data infrastructure.

*(For the interviewer: Please instruct and remind where relevant that the term "data" refers to **raw data**, that is, information collected in specific research projects that is recorded in machine readable format and used for analytic purposes. This could be survey data, interviews in sound files, video footage, notes, images, etc. By "data", we do **not** include analyses, descriptions, statistics, facts, or conclusions that appear in reports, papers, websites, or scientific publications.)*

About your institution

1) Name of institution

2) Type of institution

   a) University
   b) Public research institute
   c) Private research institute
   d) Library
   e) National/Regional archive
   f) National Statistical Institute
   g) Other (specify)

3) What is your institution's principal research discipline? (*Only applicable if respondent answered b or c on Question 2*)

4) What is your institution's principal research methodology focus? (*Only applicable if respondent answer b or c on Question 2*)

   a) Quantitative methodology
   b) Qualitative methodology
   c) Mixed methodology
   d) Other (specify)
   e) Not applicable

5) What is the scope (discipline) of your data collection? *(Only applicable if respondent answered d, e, or f on Question 2)*

Existing infrastructure and data capacities

6) Does your institution store or disseminate data for use by researchers?

   a) Yes (*go to question 8*)
   b) No (*go to question 7*)

7) What happens to the data produced at your institution? Are they at risk of being lost forever? (*skip to question 34 after answering this question.*)

8) What discipline(s) are covered by the data you store or disseminate?

   a) Social sciences
   b) Humanities
   c) Other (specify)_____

Data preservation

9) How does your institution store research data for the long-term? Could you briefly describe this process?

10) How safe are the research data that are preserved at your institution? That is, are the data kept on servers that are protected? Are there backup or formal preservation systems? (*Interviewer to explore which.)*

11) Are the data treated in a way that assures that they can be accessed and used again in 20-50 years? (e.g., kept, with necessary documentation, in a non-proprietary machine readable format)

Data documentation and standards

12) Does your institution use any documentation standards for the research data that it preserves? (*If yes:*) Which standards does it use, e.g., DDI, Dublin Core, other?

13) Does your institution follow a particular standard for trusted digital repositories, such as OAIS or the Data Seal of Approval?

14) Does your institution participate in any international survey projects that aim to harmonise datasets from different countries for comparative purposes?

15) Does your institution make use of any thesaurus to translate and/or index your data? *(If answer is yes, interviewer to ask for details).*

Tools and technologies

16) Does your institution have any experience with particular data service tools for the social sciences, such as NESSTAR, FEDORA or Dataverse? (*If yes:*) Could you briefly describe the purpose for which you use these and your experiences of using them?

Data discovery and dissemination

17) Does your institution allow access to the research data that it preserves?
(*If no, skip to next section*.)

18) Who is allowed to have access to the data? *(Explore if respondent answers researchers, whether this includes researchers in their own organization only or also in other organisations)*

19) What are the conditions that must be met to access these data?

20) Do you have some kind of authentication system that identifies who is eligible to access the data?

21) Are all of the data equally accessible? That is, are there some data that are more accessible than others?

22) Can the data be accessed from outside of the country? (*Interviewer: If yes, probe to see whether the access is for national researchers who are abroad, or whether there are also foreign researchers who can access the data.*)

23) By what means does your institution disseminate its research data? For example, are the data sent out on CDs, or are they available on screen, or can they be downloaded from a website?

24) Does your institution have a data catalogue that allows people to find the data that they are looking for? (*If yes:*) Is the catalogue visible outside of your institution, and what software is used to enable this?

Data policy and service funding

25) Does your institution have any policy or other documents regarding long-term preservation of research data? (*If yes:*) Could you briefly describe the policy or documents? (*Also, ask if they could send it to us by e-mail.*)

26) Does your institution use any legal agreements for storing, disseminating, and/or using research data? This might be in the form of deposit contracts or end-user licenses. (*If yes:*) Could you give a brief description of these legal agreements*? (Also, ask if they could send them to us by e-mail.)*

27) Do the intellectual property rights remain with the researcher/data producer, or are they transferred to your institution?

28) Does your institution receive any external funding for archiving/data service activities? (*If no:*) How is the work of data preservation and dissemination paid for at your institution? *(If yes, ask how much, whether the income is constant and whether it is dedicated to particular activities, e.g., preservation, dissemination, user support).*

29) In your view, would your institution do more to preserve and disseminate research data if it had more resources?

Staff capacities

30) Does your institution have dedicated staff for the preservation and dissemination of research data? (*If yes:*) How many?

31) Do these staff members have specific training in data preservation and dissemination? (*If yes:*) Please explain.

32) What sort of additional training do you think would be needed for your staff to acquire sufficient knowledge about data service policies and practices?

33) What kind of statistical software package experience do they have?

a) SAS
b) SPSS
c) STATA
d) R

e) MathLab
f) Excel
g) Other (specify)

About a possible National Data Service for the social sciences

34) In general, to what extent are *social science* research data preserved for the long-term in [*name of country*]?

35) Could you estimate roughly how much research data produced in [*name of country*] are lost because they are not stored in a safe setting for the long term (in percentage)?

36) Do you think it would be useful or important to establish a national data service for the social sciences in [*name of country*]? Please explain why or why not.

37) (*If yes to question* 35) We are interested in your views about what a national social science data service might look like in [*name of country*]. Can you say what key functions such an institution should serve? How might it be structured and what kinds of relationships should it have to other institutions?

38) Are there any existing national data service infrastructures for other disciplines in [*name of country*] (for example, in the humanities, medicine, climate and environment, natural sciences, or technology)?

a) Humanities
b) Medicine
c) Climate and environment
d) Natural Science
e) Technology
f) Other (specify)
g) No