National Development Plan

CROATIA

November 2017

Introduction

In this document, we present a development plan for establishing a national data archive and services for the social sciences in Croatia. The plan describes larger strategic considerations in conceiving and establishing a national data service as a first step in the process of setting up the service. In the first two sections we develop a "concept" for a future national data service, that is, the model and key features that are to be put into place. This includes defining: the overall mission; the nature of the organisation; the scope of the collection of data; the offered services and activities; the beneficiaries; and the governance structure. In the third section, the previously developed concept is evaluated in relation to realities on the ground, including probable resources and challenges for setting up and maintaining a viable data service over time. This includes subsections on identifying the host institution for the data service, defining the human resources required (number and types of staff, internal structure), and identifying potential partners among the existing network of relevant national and international organisations.

The goals of this document are:

> » to promote the development of research data infrastructure and services for the social sciences as one of the buildings blocks of Croatian scientific information system;
> » to place the data services in the context of pan-European research infrastructures; and
> » to prepare Croatia for CESSDA-ERIC membership.

The importance of investing in research infrastructures is already recognised in all relevant Croatian strategic documents (Strategy for Education, Science and Technology, Smart Specialisation Strategy) and plans (Croatian Research and Innovation Infrastructure Roadmap). The development of data archive services has to be put in that context and also in a broader European context which is strongly supported by the strategic documents. By investing in the development of data archive services for the social sciences, Croatia can make the first steps towards satisfying conditions for membership in CESSDA-ERIC[1], recognised as a ESFRI Landmark project in the ESFRI Roadmap 2016. One of the intentions of this document is to demonstrate that Croatia is participating in CESSDA as well as in DARIAH, C-ERIC, CLARIN-ERIC, SHARE, and ESS, which are currently specified in the Croatian Roadmap, and that CESSDA-ERIC should be considered for future versions of the Roadmap.

---

1   CESSDA, Consortium of European Social Science Data Archives, became CESSDA ERIC in June 2017, after the ESFRI Roadmap was published.

This plan was created within the project CESSDA Strengthening and Widening[2] by the Croatian team members, based on experiences and knowledge gained while working on the past and current projects aimed to establish a data service for social sciences in Croatia. The plan was created following the *Guide for the elaboration of national data service development plans*, produced for the CESSDA SaW project by FORS - Swiss Centre of Expertise in the Social Sciences[3]. This plan can be further developed by any agency or a group of relevant stakeholders, organizations and/or individuals who are interested in establishing data services for the social sciences on a national level. This plan, together with the Guide, can also be used as a template for developing similar national level infrastructures for other areas of science. Involvement of policy-making and funding institutions, primarily the Croatian Ministry of Science and Education, is crucial in developing this plan further in the future and to put it into action.

The benefits of establishing a national data service

Empirical research in the social sciences often involves extremely extensive and financially demanding endeavours, especially if they aim for analysis of complex phenomena on a representative sample. During such research, large amounts of data are collected, which can not be reproduced because phenomena that are studied by social sciences are changing over time.

Because data are often not appropriately preserved by researchers, they may be lost or become unusable over time. This represents a loss of potential knowledge and cultural heritage. The data service, through appropriate procedures, policies, tools and infrastructures, ensures long-term preservation of data.

From a research policy perspective, there are benefits of national data services to scientific communities that justify the needed investments. First, through their core activities - data preservation and dissemination - data services make possible long-term access to and wider and more effective use of existing data. This means that publicly-funded data are used more fully beyond their original purposes, and that research funders can concentrate on financing fewer new data collections. Indeed, data gathered for original research can be further exploited by "secondary" users for new insights and scientific contributions. Recent work by Beagrie et al.[4] demonstrates the economic benefits of national data services.

---

2   CESSDA Strengthening and Widening project (http://cessdasaw.eu) is financed from Horizon 2020 programme and coordinated by CESSDA.

3   Some parts of this text were taken directly from the Guide.

4   Beagrie Neil, John Houghton, Anna Palaiologk, and Peter Williams. (2012). Economic Impact Evaluation of the Economic and Social Data Service. Available http://www.esrc.ac.uk/files/research/research-and-impact-evaluation/economic-impact-evaluation-of-the-economic-and-social-data-service/

Second, data services strengthen research practice and quality by rendering research more transparent and open for replicability. By making data and related documentation available for scrutiny, original research can be tested and evaluated, a key pillar of the scientific method. Also, knowing in advance that the data could be used by others in the future is a way to ensure better quality in methods.

Third, more and more funders and journals are requiring that data be deposited at and available to researchers for secondary use and replication - having archives and repositories as places to store and disseminate data makes possible these requirements.

Finally, the easy availability of secondary data brings value to university teachers and students, who benefit from having data to train and illustrate methods and concepts in their courses and projects.

Mission

The mission of the **Croatian Data Archive Services for the Social Sciences** is to provide a foundation for social science research. The data archive is committed to providing support to researchers during the entire lifecycle of the project (from hypothesis development and grant preparation to data collection - original or from secondary sources, and data analysis), and in the end it ensures that the data are preserved and reusable for the long-term after the project ends. The data archive promotes data sharing and secondary analysis to achieve wider and more effective use of existing data in research and education.

The mission is realised by acquiring and managing digital objects and by ensuring that the data are seamlessly accessible in open formats and well documented for efficient re-use. In its work, the data archive follows international professional standards and procedures, and it cooperates with similar institutions around Europe and the rest of the world.

General features of the data service

In this section, we outline the general desired features of the data service by describing the main objectives and structure of the data services. The resulting concept will define the foundation and help to plan further and more concretely (in section 2).

Definition of organisation

The Croatian Data Archive Services for the Social Sciences is the national infrastructure public

service whose role is to ensure long-term preservation and dissemination of social science research data. It is an expert centre in data curation. The data archive should be established as a unit within a larger existing organisation, but the data services should be available to the whole community, national and international. The purpose of the data archive services is to provide a vital research data resource for researchers, teachers, students, and all other interested users.

Scope of collection

The data service will curate social science research data produced by researchers and research organisations during their research projects. The primary focus is on quantitative data in the disciplines of sociology, psychology, education science, information science, political science, and economics. Other disciplines can be included as well, as far as they produce data by using methodologies of the social sciences and cover issues on society and economy. The data service will also collect qualitative data, but with a more careful selection and with consideration of available resources (e.g., it will not deal with time-consuming qualitative data management if there are quantitative data still waiting for processing). Both contemporary and historical data will be acquired but with higher priorities for data management for contemporary data. Considering the size of data, mainly traditional-scale data will be processed for now, and not large-scale or "big-data". Curating data from administrative registers can be considered in the future in cooperation with other projects and initiatives, and with sufficient resources.

Services and activities

The core services of the data archive are:

- » Solicitation and acquisition of data from data producers;
- » Curation and long-term preservation of data;
- » Maintaining a catalogue of the archived data;
- » Enabling access to data and access rights administration;
- » Support, assistance, and advice for data depositors during the entire lifecycle of the project;
- » Support for data users;
- » Promoting a data sharing culture and secondary analysis.

All services can be scaled and/or developed progressively, but these core services must be implemented first. Some additional services could also be offered, such as:

- » Promotional events and outreach to users;
- » Promotion of Open Access to research data;
- » Creation of materials and training for data depositors and for data users;
- » Implementation of tools for data discovery and other tools to improve the spreading of information about available data.

Some optional services can also be offered in the future, e.g.:

- » Providing online tools for data analysis and variable exploration - creating tables, graphs and other analytical results;
- » Enabling secure access to sensitive data (safe room).

Moreover, the data archive should establish and coordinate a local network of data service users (data depositors, data users, and librarians), and it should also be extensively included in regional and international collaboration projects.

Beneficiaries

The data archive offers services for researchers who want to deposit their data and for others who need to use existing data in new research projects or for education. The primary recipients of the data archive services are university researchers and teachers, university students (doctoral, master, and/or bachelor) and researchers from other public research institutions (e.g., scientific institutes, government organisations). Another category of users could be policy makers and funding bodies who can develop data management requirements and criteria for evaluation of research once they have proper infrastructure for such requirements to be put in force. The service can also be used by researchers from NGOs and private research companies, secondary school teachers and students, journalists, businesses, professional organizations, etc. The services are free for all users from the public sector for non-profit use, but fees may be applied for commercial use.

Governance structure

The archive will be led by the Head of the data archive, who reports to the hosting institution and to the funding institution.

An Oversight Board will have the responsibility of reviewing the work of the data archive. The Board will review annual plans and annual reports and will ensure that the work of the data archive is in line with goals and strategies of scientific and other institutions to which they

belong. It will consist of representatives of major stakeholders: Ministry of Science and Education, Croatian Science Foundation, the data service host institution, major research and education institutions (universities and institutes), representatives of the Croatian Bureau of Statistics and other government organisations that are potential major data depositors and/or data users.

A Scientific and Expert Board has an advisory role on issues related to social science methodology and data archiving procedures. It includes researchers from various fields, data experts/archivists, data librarians, IT experts, and a representative of a CESSDA service provider. Their purpose is to ensure that the needs of the scientific community and other interested communities are adequately addressed.

Financing schemes

The formation phase of the data archive can be divided into two different stages: the start-up phase and the operational phase in which the cost profiles will change over a period of months or years. The start-up phase reflects the run up of activities such as the recruitment of staff and specific start-up activities such as training and further development of policies and procedures for the archive. The costs in terms of staff time in the start-up phase can be quite substantial, whereas the operational phase will reflect increasing efficiency as procedures become increasingly established and normalized. The outline of the financial requirements can be based on the KRDS Activity Model put forward by Beagrie et al[5]. It describes the full range of activities required to support long-term preservation of research data, and supports the allocation of costs across these activities. Final calculations for costs will depend on the following key variables: Staffing (based on roles and responsibilities framework of the charter document of the data centre); Software, hardware, servers and other equipment; Economic adjustment such as rates of inflation.

Ideally, the data service will be financed from the state budget as a regular activity (permanent employment status and yearly budget). Alternatively, development and functioning of the data service could be financed through 2-5 year projects, through national calls which follow national strategies for science, research and education and which are related to implementation and development of research infrastructures. In any case (and in the meantime), the data service should participate in international and regional projects (EC calls and others) to support further development, education and cooperation with relevant communities. The host institution

---

5   Beagrie, Ch. et al. (2011). User Guide for Keeping Research Data Safe. Available http://www.beagrie.com/static/resource/ KeepingResearchDataSafe_UserGuide_v2.pdf. An extensive list about digital preservation and data curation costing and cost modelling of which the report by Beagrie is part of can be accessed under the following link: http://wiki.opf-labs.org/display/ CDP/Home.

contribution consists of the provision of administrative resources and workspace facilities.

Resources and challenges

In this section we consider how the concepts and general features of the data service outlined in section one can be implemented in practice, taking into consideration real local and institutional resource conditions.

Host institution of the future data service

The Faculty of Humanities and Social Sciences, University of Zagreb (FFZG) has been recognised as an appropriate hosting institution for the new data service since 2006 when a group of scientists and information experts from various disciplines and institutions gathered in the Initiative for Social Science Data Archive. The Initiative participated as a project associate in the CESSDA Preparatory Phase Project[6] (2008). Further efforts directed to the establishment of the data archive in Croatia started through the FP7 project *SERSCIDA - Support for Establishment of National/Regional Data Archives*[7] (2012-2014). Efforts continued in 2015 when FFZG became the partner in two new international projects: *SEEDS - South-Eastern European Data Services*[8] funded through the SCOPES programme by the Swiss National Science Foundation and the Swiss Agency for Cooperation and Development; and *CESSDA SaW - Strengthening and Widening* financed through the Horizon 2020 programme.

The FFZG, as the largest research and education institution in Croatia in the area of social sciences and humanities (HSS) offers more than 100 BA and MA programs as well as 19 PhD programs, and has more than 500 academic staff and around 7,000 students. The FFZG is well integrated into the national research network and it has the biggest academic library in Croatia for the areas of HSS, which offers public services to all users interested in its collections.

The prototype data archive - repository and website were developed at the Library during the SERSCIDA project and then improved during the SEEDS project[9]. The newest prototype is built

---

6 "In 2006 CESSDA was identified as a social science research infrastructures (RIs) of excellence by the European Strategy Forum on Research Infrastructures (ESFRI). Consequently, CESSDA was awarded FP7 funding to undertake a Preparatory Phase Project (PPP) lasting two years. The PPP will facilitate CESSDA's transition to a fully integrated and formally constituted European Research Infrastructure (ERI)." https://ppp.cessda.net/doc/CESSDA_PPP_ECRI2008.pdf

7 SERSCIDA - Support for establishment of National/Regional Social Science Data Archives, http://www.serscida.eu/

8 SEEDS - South-Eastern European Data Services, http://seedsproject.ch/

9 Prototype data archive website: https://hr.seedsproject.ffzg.hr/?lang=hr; Prototype data repository: https://dataverse.ffzg.unizg.hr/.

on the Dataverse platform and is completely ready to become a production system with a persistent identifier (DOI) assigned to each dataset. Also, data archive policies and procedures were developed during the work on both projects.

The FFZG has reliable existing technical infrastructure - proper servers and hardware, software, network and telecommunications. This infrastructure is currently used for FFZG Library operations and for providing digital services to users, and could support some of the needs of a data archive and its services. But more sustainable measures have to be implemented to ensure a long-term preservation of data collections. This can be achieved in cooperation with the University of Zagreb University Computing Centre (Srce). Some preliminary communication and discussion between FFZG staff and Srce management have already been established in that direction. Also, formal cooperation between these two institutions has been established around the Dabar[10] infrastructure, which is perceived as an important component of the future data service.

Human resources and internal structure

There is a minimum of four dedicated staff roles needed for a data service. One should be the role of the Head of the data service - a person responsible for financial and executive management, communication with stakeholders and regional and international partners. The Head of the data archive must be a person with a full understanding of the research data management workflow, as well as of issues and challenges associated with these processes, and should also be familiar with technical aspects of research data management. He or she has to have good management skills and a strong national and international network of stakeholders. At least an MA degree in a field of the social sciences is needed, while a PhD would be an advantage.

There should be two data specialists. Data specialist roles can be divided into junior and senior positions, based on previous work experience. A senior data specialist should have demonstrated knowledge in social science methodology, statistical data processing, and data management. An MA in a field of the social sciences and an additional degree in information science are desirable. A junior-level data specialist should hold at least a BA in a social science field. This does not require previous extensive experience in research data management, but sufficient knowledge of statistical operations and data processing is needed for that position.

A fourth role should be an IT expert. Systems administration, software programming and database management skills are necessary for a position of the IT expert in the data archive.

---

10   "Dabar provides technological solutions that facilitate maintenance of higher education and science institutions' digital assets, i.e., various digital objects produced by the institutions and their employees." https://dabar.srce.hr/en/dabar

The position also requires an overall understanding of data exchange protocols in a networked environment and security issues associated with it. He or she has to have experience in software applications architecture, development, and maintenance. A BA in the field of informatics and computer science or equivalent experience is required, and previous experience with research data management and processing is desirable.

For the normal operation of the service, administrative support from the hosting institution is necessary. The following services are crucial: financial administration, accounting services, HR department services, legal department services, and IT department services, together with all other workspace maintenance services.

At the beginning, the bare minimum number of dedicated people to establish the Croatian Data Archive Services for the Social Sciences is two. Both persons should be data experts and should be able to undertake several roles. At least one of them should also have a good understanding of the data service IT needs to be able to implement new technical solutions in cooperation with other partners. One of them should take on the role of the Head of the data service.

For the last 5 years, existing FFZG staff, together with external collaborators from other institutions (prominently Centre for Scientific Information, Ruđer Bošković Institute and University of Zagreb University Computing Centre - Srce), have come through various training activities related to establishment and day-to-day operation of a data archive and services for the social sciences. This experience has to be utilised in the process of establishment of the national data services. There are currently no fully dedicated staff to support the full establishment and operations of the data service, and so during the establishment phase new staff has to be employed to take on the activities.

Partner support and cooperation

On the national level, it is very important for the data archive to be recognised as a valuable national service by all relevant institutions in the area of research and education. Stakeholders and partner institutions should include:

» The Ministry of Science and Education (MSE) as the main science policy institution. The data archive services and activities have to be recognised in national strategies related to science and education and supported in action plans. Establishing and sustaining a national service is not possible without the strong commitment of the Ministry. Membership in Pan-European research infrastructure CESSDA-ERIC is only possible through the Ministry and one of the conditions for membership is to have an operational service provider who is able to meet all professional requirements.

» The Croatian Science Foundation (CSF) as the main funding body for competitive research projects. The Foundation can develop criteria for project calls and evaluation which include data management plans, data archiving issues, and open access to data. The Foundation is a possible funding source for some of the activities of the data service. Also, it can finance data management activities in research projects and in that way promote the importance of good data practices in research.

» The University of Zagreb University Computing Centre (Srce) as the provider of crucial technical infrastructure for long-time preservation of the digital assets collected and managed in the data archive. Staff employed at Srce was involved in project activities about establishing a data archive and services for the social sciences. Srce implements and maintains technical aspects of the Dabar infrastructure, which is important in the context of data archiving. The Dabar is developing in cooperation with several other research and education institutions.

» The Centre for Scientific Information, Rudjer Boskovic Institute as the institution with experience in building and maintaining information systems for researchers. Staff employed at Srce was involved in project activities about establishing a data archive and services for the social sciences. The Centre is also involved in development and implementation of the Dabar infrastructure.

» All research and education institutions in Croatia in the area of social sciences, including faculties (constituents of universities) and public research institutes, should be in close cooperation with the data services. Researchers employed at these institutions are the main provider of data for the data archive, and also the primary users. These institutions can have their own policies about data management and data sharing. Also, they usually have their local information specialists/librarians who can promote and facilitate data sharing and data reuse practices.

» The Croatian Bureau of Statistics (CBS), which produces national statistics, and other valuable data heavily used by social science researchers. A formal cooperation with the CBS could be established in the future in the area of managing, documenting and promoting the use of microdata produced by CBS which are of the most interest to researchers.

» Government bodies, which produce data and statistics of value for some researchers. These institutions can also benefit from having a secure and reliable place to store and document the data they produce. Discussion about possibilities for data linking should be started.

» The National Archive, which can play a role in the long-term preservation of some kinds of data (e.g., publicly available research data as a national heritage).

» The National and University Library (NUL) and all other university and academic libraries in Croatia as the main partner in the dissemination of data and other information managed by the data archive. Data archive operations are in large part very similar to library operations so that the same operational models can be applied to both institutions. Thus, the close cooperation of the data archive and libraries is necessary.

Cooperation with other research infrastructure projects, especially the ones in the area of social sciences which are recognised by ESFRI Roadmap and which are active or were active in Croatia

- namely DARIAH-HR, SHARE, and ESS, is important because there are some similarities between these infrastructures which could be synergised.

Regional and international cooperation is also necessary for the service to improve its relevance and competence. Participation in regional projects and continuing cooperation in the region is interesting because of linguistic and cultural similarities, as well as common history. International cooperation is achieved through membership in CESSDA-ERIC, which represents a comprehensive integrated social science research infrastructure.

During the last five years, while working on projects related to the establishment of data services (SERSCIDA, SEEDS and CESSDA-SAW), most of the listed stakeholders were contacted and the importance of data services for the social sciences was explained to them. Representatives from various institutions were included in project activities - conferences, training and visits to European data archives.

Local challenges

The need for national data archive services for the social sciences has yet to be clearly recognised in relevant strategic documents and in supporting action plans in Croatia. The most important documents for the development of research infrastructures in Croatia are the Strategy of Smart Specialisation and Croatian Research and Innovation Infrastructures Roadmap, as they are the foundation for future investment in national and international infrastructure projects. The current version of the Croatian Roadmap (June 2016) does mention the importance of large repositories of data for the social and life sciences, and the need to include Croatia in Pan-European research infrastructures is expressed in the document as well as support for ERIC memberships, but CESSDA-ERIC specifically is not yet listed as a research infrastructure project that could be financed from European structural funds[11]. The reason might be that in the time when the Roadmap was developed CESSDA still did not have ERIC status.

The strong challenge for the last few years was how to communicate with the Ministry of Science and Education and other relevant policy makers and financing bodies the need and importance of establishing the data archive service for the social sciences. The government was not stable and it was not clear in which direction the implementation of the relevant strategies would go, so

---

11   Croatian Research and Innovation Infrastructures Roadmap (June, 2016). https://mzo.hr/sites/default/files/migrated/plan_razvoja_istrazivacke_infrastrukture_u_rh.pdf (Croatian only)

Older version of the document from April 2014 is available in English at the following URL: https://mzo.hr/sites/default/files/migrated/croatian_research_and_innovation_infrastructures_roadmap.pdf

no one at the Ministry was able to make any promises or commitments. It was not clear about the sources of funding from which the establishment process and day-to-day operations of data services could be financed.

Afterword

The NDP for Croatia has been drawn up by Marijana Glavica from FFZG, mglavica@ffzg.hr.

Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3
10000 Zagreb, Croatia
Phone: +385 1 4092 393
Email: data.archive@ffzg.hr

Design: Open Concept AS
openconcept.no