South-Eastern European Data Services

D5 – Policy and procedures: Kosovo

Author(s):	Lirije Palushi (CPC)
	Renate Kunz (FORS)
	Irena Vipavc Brvar (UL-ADP)

Version History

Version	Date	Changes	Modified by
1.0	February 14, 2017	Draft generic version	FORS
1.1	February 17, 2017	Generic Version	ADP
2.0	April 30, 2017	Released Version	CPC

Acknowledgments

This report has been developed within the "South-Eastern European Data Services" (SEEDS) (www.seedsproject.ch) project. The participant organisations of the SEEDS project are:

Name	Short Name	Country
Faculty of Humanities and Social Sciences University of Zagreb	FFZG	Croatia
Centre for Monitoring and Research, Podgorica	CeMI	Montenegro
Centre for Political Courage, Pristina	CPC	Kosovo
Institute for Democracy and Mediation, Tirana	IDM	Albania
Institute of Economic Sciences, Belgrade	IES	Serbia
Saints Cyril and Methodius University, Institute for Sociological, Political and Juridical Research, Skopje	ISPJR	Macedonia
Swiss Foundation for Research in Social Sciences, Lausanne	FORS	Switzerland
University of Ljubljana, Social Science Data Archive, Ljubljana	UL	Slovenia

Table of Contents

1. Introduction

There is a growing awareness about data sharing among international organisations, national science organisations, research funding bodies, data services, universities and researchers, on the one hand but also the growing needs for proper tools, mechanisms and instruments aimed at providing trustworthy long-term preservation of research data. Data sharing enables the reuse of data by researchers who did not generate these data themselves, thus leading to greater efficiencies and more research. Data sharing also stimulates the usage of data beyond research in academia. Ultimately, data sharing leads to a higher return on investment. Data sharing furthermore makes science more transparent and facilitates replication of research by others.

Crucial perquisites for any existing or aspiring data service are a set of clearly defined, written down, core policies in order to achieve trust among those stakeholders.

Through their core activities – data preservation and dissemination – data services make long-term access to and wider use of existing data possible in the first place. This means that publicly funded data are used more effectively, beyond their original purposes (secondary use). It goes without saying, that data services, as a key stakeholder, have to develop a transparent set of policy and procedures that support internal data management procedures across the whole data life-cycle and ensure accountability and allow for external quality control. Accountability and transparency are key factors for creating trust by funders and researchers.

Three main models and guidelines that serve as a foundation for this policy and procedures document are outlined in chapter 2. The following chapters focus on the policies of the future data service. They are represented in a three-layered policy structure: The high-level organisational infrastructure, chapter 3; the descriptions of digital object management procedures as a data lifecycle approach, chapter 4; and the segments on technical infrastructure, security and risk management, chapter 5. The policies are for the time being described in one document. As the future data service starts to take shape, and services start being more distributed, the different policies can be developed further into separate different documents (strategies and programs).

2. Conceptual frameworks

2.1 CESSDA Maturity model

The first model that acts as a makeshift is the CESSDA SaW Capability Development Model (CESSDA-CDM)[1]. It aims to provide a structured view of processes across an organisation (data service or research infrastructure) and it can be used to set process improvement goals and priorities, provide guidance for quality processes and activities, and provide a benchmark for assessing and appraising current practices. The CESSDA-CDM was generated in the realm of the CESSDA SaW[2] project as a tool to evaluate social science data archives and services in European Research Area (ERA) countries, to identify gaps and bottlenecks in existing data services, and to produce national development plans to close the gaps and overcome present barriers.

It is a structured collection of elements that identify and describe the characteristics of effective preservation processes and activities. Building on established frameworks for trustworthy data

---

[1] http://cessdasaw.eu/content/uploads/2016/06/D3.1.pdf.
[2] http://cessdasaw.eu/.

preservation and the CESSDA community's prior experiences, the model provides both a starting point for emerging preservation initiatives and a reference tool for established data services that wants to strengthen their services. It is a model that can be used to appraise and/or improve the capability of a data service to perform and to provide services.

The CESSDA-CDM takes its cue from the Reference Model for an Open Archival Information System (OAIS) (see chapter 2.2) and the European Framework for Audit and Certification (also known as Trusted Digital Repository EU) (see chapter 2.3).

The CESSDA-CDM is hierarchical. On the highest level, the model focuses on three main subject areas, so called Capability Requirement Areas (CRA), which describe on a high-level, the main objectives and principles of a data service. Each CRA is divided in various Capability Process Areas (CPA), which each has its own purpose. Within each CPA there are several activities defined to achieve the objective(s) of that CPA.
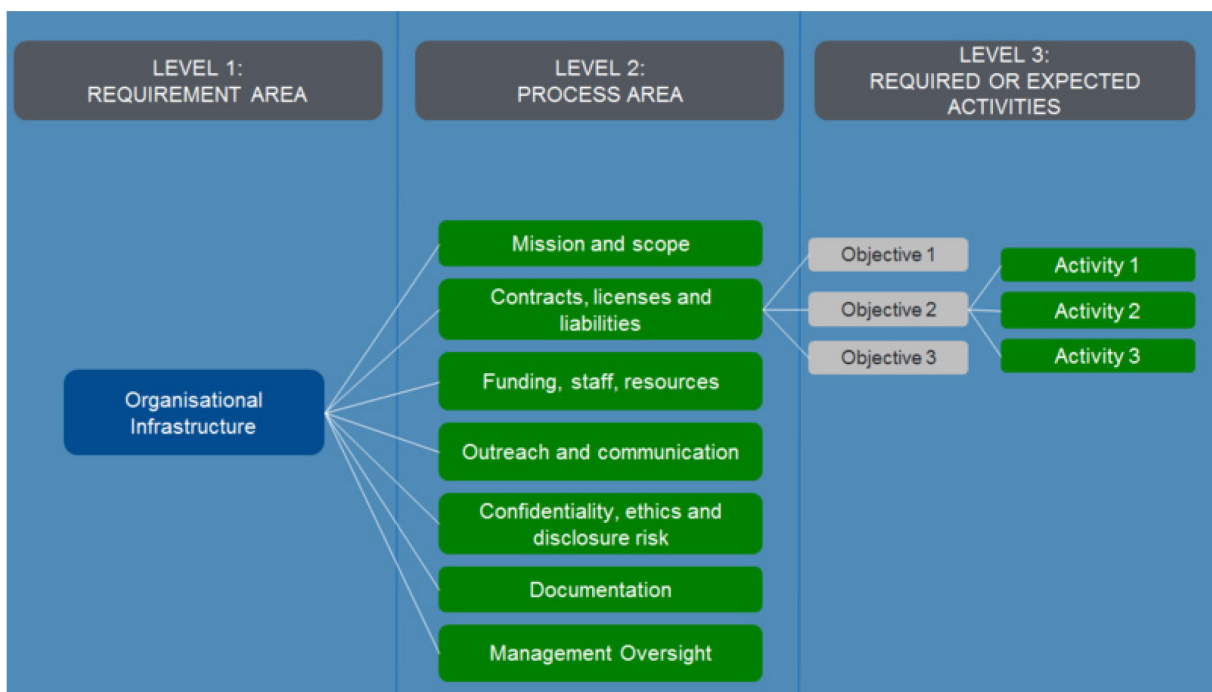
CESSDA SaW Capability Development Model (CESSDA-CDM), in: CESSDA SaW, Deliverable 3.1 Heuristic Maturity Development Model, 2016, p. 11 (http://cessdasaw.eu/content/uploads/2016/06/D3.1.pdf).

2.2 Open Archival Information System Reference Model (OAIS)

The policy domains put forward in this document correspond to functional areas within the Open Archival Information System Reference Model (OAIS). The OAIS model is a conceptual framework for an archival system dedicated to preserve and maintain access to digital information over the long term. The OAIS model specifies how digital material should be preserved for a community of users (Designated Community) from the moment digital material is ingested into the digital storage area, through subsequent preservation strategies, to the dissemination of digital material for the end user.

The OAIS model differentiates three so-called Information packages which all are connected and relate to each other. They were developed in order to better describe the different handling and

varying activities in digital preservation. The information package that is ingested into the archive is called Submission Information Package (SIP). Once in the archive, the SIP is enriched with metadata and converted into an Archival Information Package (AIP), which represents the form in which the digital information is actually stored for the long term. The AIP's are made accessible through the so-called Dissemination Information Packages (DIP), which are generated for a specific user group according to certain legal requirements. Three principal actors, known as Producers, Consumers and the Management are interacting within six functional entities, Ingest, Data Management, Archival Storage, Preservation Planning, Access and Administration.[3]
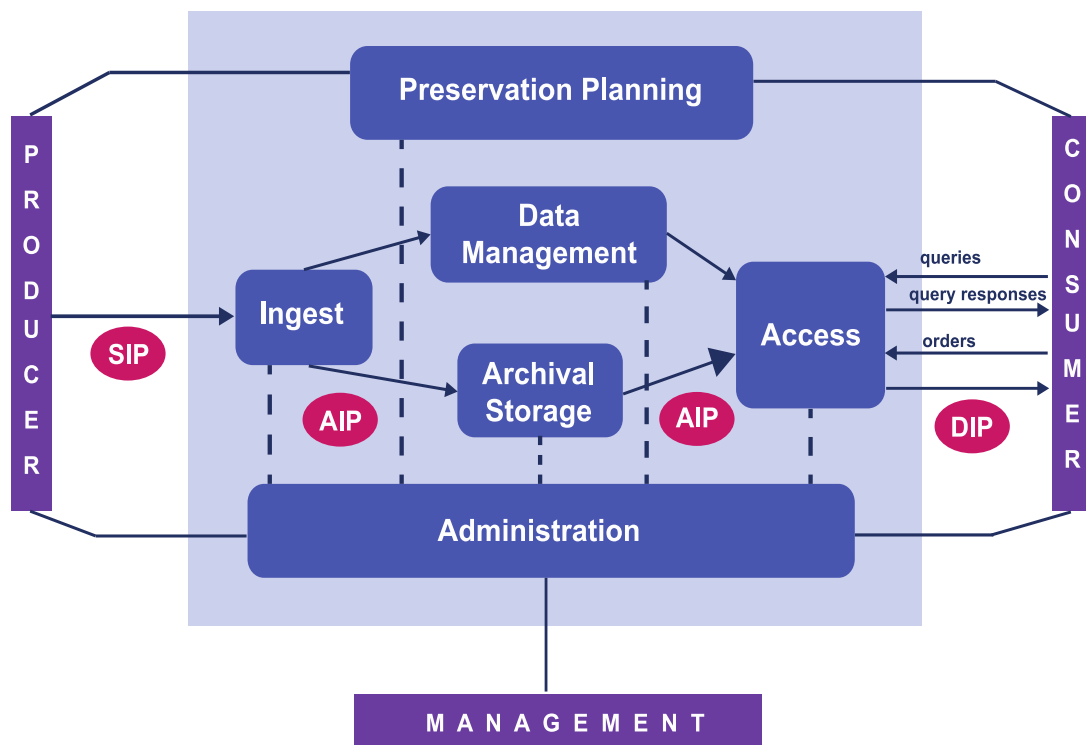
Reference Model for an Open Archival Information System (OAIS); CCSDS 650.0-M-2; Consultative Committee for Space Data Systems: Washington, DC, 2002, p. 4-1 (https://public.ccsds.org/pubs/650x0m2.pdf).

2.3 DSA and DSA-WDS

The Data Seal of Approval (DSA) was developed in 2008 by DANS (Data Archiving and Networked Services) in the Netherlands. It has been further developed, expanded and internationalized and handed over to an international board in 2009. The DSA involves 16 guidelines for applying and verifying quality aspects concerning the creation, storage, use and reuse of digital data. The guidelines serve as the basis for awarding the Data Seal of Approval by the DSA Board.[4]

The DSA is part of the European Framework for Audit and Certification (also known as Trusted Digital Repository EU). The European Framework for Audit and Certification is a collaboration between Data Seal of Approval, the Repository Audit and Certification Working Group of the Consultative

---

[3] In addition to the OAIS model that supposes that the ingest procedure starts with SIPs that after some handling become AIPs and can be ingested, we feel that there is a need for an extra stage in the ingest procedure: pre –ingest. In this pre-ingest stage the received material will be checked on various aspects, which are fundamental to make a decision whether the material should be accepted to enter the repository in the first place (see chapter 4.1.2).

[4] http://www.datasealofapproval.org/media/filer_public/2014/10/03/20141003_dsa_overview_defweb.pdf

SEEDS: D5 – Policy and procedures: Kosovo

Committee for Space Data Systems (CCSDS)[5] and the DIN Working Group "Trustworthy Archives – Certification"[6]. The framework consists of three trust and certification models: the DSA (Data Seal of Approval)[7], the DIN 31644 (the Nestor seal for trustworthy digital archives)[8] and the ISO 16363 (audit and certification for trustworthy digital repositories)[9]. The DSA guidelines can be seen as a minimum set of requirements and as a lightweight approach in this framework.

There are currently just over 60 data services, which obtained the Data Seal of Approval by the end of 2016. Through a self-assessment the data service should supply evidence that it meets the 16 DSA guidelines and the relevant level of compliance. After submission, the DSA Board appoints a peer reviewer evaluating the self-assessment.

In 2012, the DSA and the ICSU World Data System (WDS)[10] started a common working group under the umbrella of the Research Data Alliance (RDA)[11] with the objectives of realising efficiencies, simplifying assessment options, stimulating more certifications, and increasing the impact on the community. The DSA and WDS certifications both offer a basic certification standard for trusted digital repositories. Their catalogues of requirements and their review procedures are based on the same principles of openness and transparency. Up to this point, the two standards have evolved and operated independently. The primary focus of DSA has been on data services in the Humanities and Social Sciences. For historical reasons, the focus of WDS has been on Earth and Space Sciences.

At the End of 2016, the ICSU World Data System (WDS) and the Data Seal of Approval (DSA) Board presented a unified catalogue of requirements. The group built on inherent complementarity between the criteria previously established by the two organisations to harmonise unified and universal requirements reflecting the Core Characteristics of Trustworthy Data Repositories. The applicant must indicate a compliance level for each of the requirements, which are similar to the structure in the CESSDA-CDM model (see chapter 2.1): Organisational aspects, management of digital objects and technical aspects.[12]

3. Organisational infrastructure

3.1 Mission/Scope/Purpose/Mandate

The Kosovo Social Sciences Data Center (KSSDC) in a national e-infrastructure for social sciences and humanities, which provides long time preservation and distribution of social science and humanities research data.[13] Kosovo`s Data Service will be opened and dedicated to the broad research community including researchers, teachers, students as well as other stakeholders involved in scientific research in Kosovo and abroad.

---

[5] https://public.ccsds.org/default.aspx
[6] http://www.din.de/en/getting-involved/standards-committees/nid
[7] http://www.datasealofapproval.org/en/
[8] http://www.langzeitarchivierung.de/Subsites/nestor/EN/Siegel/siegel_node.html
[9] http://www.iso16363.org/standards/iso-16363/
[10] https://www.icsu-wds.org/
[11] https://www.rd-alliance.org
[12] https://drive.google.com/file/d/0B4qnUFYMgSc-eDRSTE53bDUwd28/view
[13] More information at our web site: https://ks.seedsproject.ffzg.hr/

The Kosovo Social Sciences Data Center (KSSDC) will be established by the Centre for Political Courage (CPC) within the Institute for Social Studies and Humanities (ISSH), at the Faculty of Philosophy, University of Prishtina. The ISSH will provide support in data management and will promote the use of the data. The holdings will include social science data in a broader sense. The primary focus, however, will be on quantitative data in sociology, psychology, educational science, political science and economics. The KSSDC will also collect and curate qualitative data, but with more careful selection and with consideration of available resources.

The services offered will include:

- Selecting and acquiring data;
- Processing and cataloguing data and documentation;
- Data dissemination;
- Promotion of open access;
- Enabling secure access to sensitive data;
- Digital preservation and (technical) systems development and maintenance;
- Technical support;
- Education and support for depositors (how to manage data, data management plan) and users (how and why to use data for secondary research);
- Cooperation with national (e.g. Ministry of Education, Science and Technology of the Republic of Kosovo, Kosovo Agency of Statistics), regional and international stakeholders.

3.2 User Orientation and Designated Community (definition and monitoring)

The designated community encompasses a wide range of institutions and individual researchers in Kosovo such as higher education institutions, both public and private, research institutes and institutions, NGO`s, IGO`s, public institutions such as the Kosovo Agency of Statistics, and think thanks. Basically a data producer can be anyone who proves the validity of their research data and who is able to share its data for secondary use with the broad research community. On the other hand, everyone who has access to the data preserved is considered a user. Users will be registered and include students, teachers, members of public and private institutions, policy makers, and journalists from Kosovo and abroad.

The KSSDC `s core stakeholders and partner institutions, inter alia, include:

- Ministry of Education, Science and Technology (MEST) as the main science policy institution, which also provides multi-annual institutional financing of research activities (distributed to universities and public institutes);
- All relevant public and private scientific institutions and domestic and international agencies operating in Kosovo that deal directly or indirectly with social science data. This includes faculties (constituents or universities); it also includes public institutes that have their own policies about data archiving and may also have their local information specialists/librarians who can promote and facilitate data depositing and data use;
- The Kosovo Agency of Statistics, which produces national statistics, valuable and very heavily used data in social science disciplines;
- The National Archive which can play a role in the long-term preservation of some kinds of data (e.g., publically available data as national heritage);

- The Agency of Information Society (AIS) / Ministry of Public Administration, which is the main institution responsible for developing, maintaining and preserving the IT infrastructure and administrative data for public institutions;
- The National Library;
- The University Library;
- Think tanks and NGOs, which are engaged in social science research activities and which promote the concept of open data. This should also include a study of the legal questions concerning formal relationships with stakeholders.

Formal communication will be established with the managers of relevant institutions. Formal letters will be sent to inform institutions about on-going activities. Also, the Data Service will identify one contact person in every relevant partner institution. In some institutions, they are librarians, and in others they are researchers and/or members of the management team. Valuable candidates for contact persons are researchers who are involved in international research projects. Annual meetings with representatives will be organized, and based on the discussions and needs this meetings can intensify. A mailing list will be used to disseminate information about on-going activities. All interested researchers, librarians and others can be added to the mailing list.

Promotional activities to raise visibility have already begun and should be on-going. These will include targeted messages to specific audiences by way of different means and platforms. The means could include meetings, workshops and conferences, email and letter campaigns and invited visits by key stakeholders. Another way of communication with our community is through the adequate use of social media such as Facebook, Twitter, LinkedIn groups etc.

3.3 Financial sustainability and Resources

3.3.1 Funding

Ideally, services will be financed by the government budget, e.g., within the regular budget of the University of Prishtina in order for the national service of data archive to function adequately. Beside the official financial resources which should secure the sustainability of the data archive continuously, we will ask for other financial resources from other relevant national stakeholders as well as international funding from different institutions and agencies e.g. European Commission, in order to provide additional promotion and capacity building activities.

The general costs for the creation and maintenance of the data archive will depend on different conditions such as staffing, equipment and economic adjustment at the time of creation/maintenance. A more detailed budget will be determined after a deep analysis of the possible costs at the time of creation.

3.3.2 Collaboration

Since Kosovo has not yet established its National Data Archive, one of the crucial parts for the future operation of this data service will be the cooperation with relevant national and international stakeholders. On a national level, it is very important that the data archive will be recognized as a valuable national service by all relevant science and research institutions, which are part of the

science community and higher education system in Kosovo. For specific stakeholders with whom we will try to build continuous partnerships see section 3.2 above.

On the other side, it is of the same importance to cooperate with regional and especially international institutions or data archives, since we can learn a lot from their experience in terms of research, data management, maintenance, staff training and capacity. Kosovo seeks to be a member of the Consortium of European Social Science Data Archives (CESSDA) one day, which is conditioned by the best practices and standards which can ensure the quality of data archive work and services as well as to ensure data availability and sharing throughout the international community.

3.4 Staffing

3.4.1 Roles and responsibilities

The precise number of staff as well as their definite roles and responsibilities will be determined at the time of the establishment of the KSSDC, based on the available resources and expected services. However for the data service the following dedicated staff is needed:

- Director/Head of the KSSDC (S/he should be responsible for financial and executive management, communication with stakeholders, and with regional and international partners);
- Data Expert with proven knowledge in social science methodology and data management;
- Experienced IT expert in data management, software application, documentation standards, data exchange protocols and programming and database skills;
- Office and Financial Officer to maintain day-to-day office duties and finances.

The KSSDC will be organized as a separate unit at the ISSH at the University of Prishtina. The staff in the aforementioned positions will be chosen through an open call for application and only if they fulfil the expected requirements and are qualified in accordance with the work descriptions.

3.4.2 Units and departments

The KSSDC management structure will be organized within two departments: Data archiving department and IT department, and it is characterized by an overall broad span of control with one hierarchical level. All department/task employees will report directly to the KSSDC manager. The KSSDC oversight Board composed by CPC and ISSH members, in close cooperation with the host institution, Faculty of Philosophy, University of Prishtina, will fulfil the function of a Joint Oversight Board of the KSSDC. An organizational chart is developed that shows the division of tasks, responsibilities, reporting relationships, and hierarchy as in the following:

Kosovo Social Sciences Data Center (KSSDC) Organizational Chart

| Head / Manager of the KSSDC | ⟷ | Joint Oversight Board of the KSSDC (CPC and ISSH) |

SEEDS: D5 – Policy and procedures: Kosovo
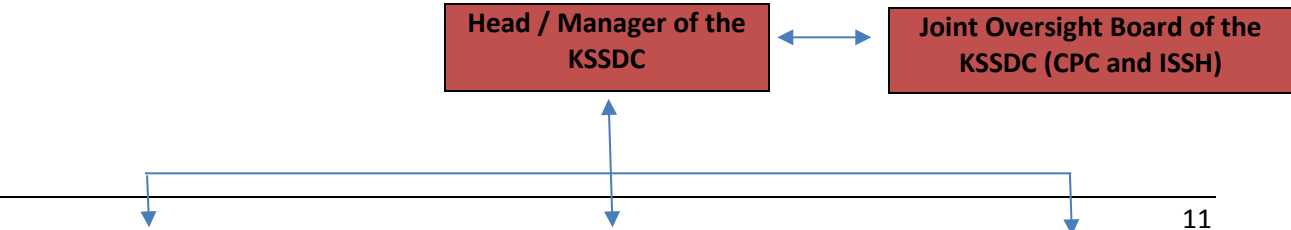
Data Archiving Department	Office and Financial Officer	IT Department IT specialists

3.4.3 Job descriptions

The Head / Manager of the data archive must be a person with a full understanding of the research data management workflow, as well as issues and challenges associated with these processes. He or she would preferably hold at least a Master's degree in a field of the social sciences and have experience with technical aspects of research data management. The Head of the KSSDC reports to and is accountable for the Joint Oversight Board of the KSSDC and supervises both data specialist as well as IT specialists.

Data expert roles can be divided into junior and senior positions, based on previous work experience. Senior data experts 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. Junior-level data experts should hold at least a BA in a social science field. This does not require previous extensive experience in research data management, but they should have sufficient knowledge of statistical operations and data processing.

For the IT experts, systems administration, software programming and database management skills are necessary for a position in the data archive. He or she has to have experience in software applications architecture, development, and maintenance. The position also requires an overall understanding of data exchange protocols in a networked environment and security issues associated with it. 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.

3.5 Legal framework

3.5.1 Legal status and legal responsibility

The KSSDC will be a separate unit of the ISSH and therefore its legal status and legal responsibility will be subject to the internal legal acts of the University of Prishtina, in compliance with the current law on Higher Education and the Law on Scientific and Research Activities among others.

3.5.2 National laws
The following laws regulate issues related with the functioning of the KSSDC in Kosovo.

The Law on Scientific Research Activities (Official Gazette, no. 04/L-135) regulates, defines and specifies all scientific research activities, rights, and obligations of scientists and researchers, as well as the financing sources of scientific research activity. In general, the law places great emphasis on creating and promoting an innovative and competitive research environment. The development of

advanced research technologies and infrastructures, enhancement of cooperation between different stakeholders regarding scientific research activities, as well as the promotion of an effective dissemination of scientific results are seen as essential preconditions for an overall economic and social prosperity.

Article 5 of this law defines the funding of scientific research activities. Specifically, section 1.4 states, among others, that the financing means provided by the government of the Republic of Kosovo include "activities in systematical preservation and distribution of activities outcomes of basic research and those applied, such as scientific conferences, exhibitions, publications and other forms of documentation".[14]

The Law on State Archives (Official Gazette, no. 04/L-088) regulates the organisation and operation of archival services in Kosovo's institutions. This law specifically defines archiving procedures and conditions of archived material and documentary material that are produced in public institutions. According to section 1.2 of article 3 of this law, archived material is defined as "all original and reproduced material by public institutions, legal and natural entities in their activities such as: manuscripts, printed records, drawings, photographs, stamps, audiovisual materials, electronic materials and all other supporting tools that enable a clear understanding of the content of information and facilitate the use of them".[15]

This law also discusses the organisation of the State Archive Agency, as the main body responsible for receiving, identifying, maintaining, and publishing the archival material. In this sense, a great importance is given to a free, open and informed access of the interested general public to the material possessed by the Agency. Article 8 specifies that the Agency, "in order to improve opportunities for the use of archive material shall create, maintain and update the public archive registry". Furthermore, paragraph 6 of this article states that the public archive registry should be managed by using advanced technology and should be available on the website of the Agency. And Article 12 of this law guarantees the right to access after the request is archived in the Agency without discrimination on any basis.[16]

The Law on Official Statistics (Official Gazette, no. 04/L-036) is one of the most important laws regarding statistical data production, storing and dissemination in the Republic of Kosovo. It mainly defines the core principles of official statistics, the organisation of the official statistics system, data collection, processing, and storing of statistical data, as well as the dissemination and use of official statistics. According to Article 4, paragraph 1.5 of this law, one of the criteria that guarantees the quality of official statistics is that of accessibility and clarity, defined as "the conditions and modalities by which users can obtain, use and interpret data".

The law also specifies the responsibilities and tasks of the Kosovo Agency of Statistics (KAS) as the main coordinator of the statistical system in Kosovo. In this sense, among other responsibilities, KAS has to prepare the Programme of Official Statistics, which shall provide a detailed framework regarding data production and dissemination. With respect to statistical data storing, Article 26, paragraph 3 of this law states that "statistical material shall be stored in such a way as to prevent its

---

[14] Law No. 04/L-145 On Information Society Government Bodies Article 6, available at:
http://www.kuvendikosoves.org/common/docs/ligjet/Law%20on%20information%20society%20government%20b odies.pdf (accessed on 27 April 2017)

[15] Law No. 04/L-088 On State Archives, available at:
http://www.kuvendikosoves.org/common/docs/ligjet/Law%20on%20State%20Archives.pdf (accessed on 27 April 2017)

[16] Ibid.

destruction, misuse, alienation, misinterpretation, and its disclosure".[17] The law also obliges KAS and other authorities to disseminate all the official statistics in such a way that all users have the necessary support for ensuring quality on equal access.

The Law on Copyright and Related Rights (Official Gazette, no. 04/L-065) is also relevant to the aims and objectives of the SEEDS project. In general, the law regulates copyright that is based on intellectual property, which belongs to authors with respect to their works in the literary, scientific, and artistic domains. Among others, the law mentions the rights of database producers.

Regarding the right of distribution, Article 24 of this law states that it "is an exclusive right of the author to authorize or to prohibit the putting into circulation of the original or copies of his work, by sale or any other form of transfer of ownership, including their offering to the public for such purpose". In Article 33, on the other hand, the right of making the work available to the public is defined as "an exclusive right of the author to authorise or prohibit that his work is made available to public through linear and nonlinear communication, in a way which enables access to individuals from place and time they choose".[18]

The Law on Access to Public Documents (Official Gazette, no. 03/L –215) guarantees the right of access to official documents maintained, drawn or received by public institutions to every natural and legal person without discrimination on any grounds. Furthermore, it obliges all public institutions to assign units or officers who will be responsible for receiving and conducting an initial review of applications for access to documents. According to this law, public institutions are also obliged to publish in electronic form and through the publication in the Official Gazette of Republic of Kosovo all that can be public. Any applicant of a document has the right of access without the obligation to specify the reasons.[19]

The Law on the Protection of Personal Data (Official Gazette, no. 03/L – 172) determines the rights, responsibilities, principles and measures with respect to the protection of personal data that have been processed by public and private bodies. The law regulates the legitimacy of data processing, rights of the data subject, as well as the transfer of personal data. Specifically, Article 9 of this law defines the processing of personal data for historical, statistical, and scientific-research purposes. It states that personal data may be further processed for historical, statistical, and scientific research purposes regardless of the initial purpose of collection. In this case, personal data should be anonymised unless otherwise provided by law, or if the data subject has given his or her prior written consent. Also, the publication of the results of processing should be in anonymised form, unless the data subject has given his or her prior written consent for the publication in a non-anonymised form,

---

[17] Law No. 04/L-036 On official statistics of Republic of Kosovo, available at:
http://www.kuvendikosoves.org/common/docs/ligjet/Law%20on%20officiial%20statistics.pdf (accessed on 2 May 2017).

[18] Law No . 04/L-065 On copyrights and related rights of the Republic of Kosovo, available at:
http://www.kuvendikosoves.org/common/docs/ligjet/Law%20on%20copyright%20and%20related%20rights.pdf (accessed on 3 May 2017)

[19] Law No .03/L –215 On access to public documents of the Republic of Kosovo, available at:
http://www.assembly-kosova.org/common/docs/ligjet/2010-215-eng.pdf (accessed on 3 May 2017)

SEEDS: D5 – Policy and procedures: Kosovo

or unless the prior written consent for such publication has been given by the legal heirs of the deceased individual.[20]

The Law on Information Society Government Bodies (Official Gazette, no. 04/L-145) aims to determine the relevant institutions as well as their functions and responsibilities in the development and implementation of information technology in the Republic of Kosovo. Regarding the functions of the Agency for Information Society, Article 6 of this law, among others, states that this Agency is responsible for "accumulation, administration, dissemination and storage of data by creating the State Data Electronic Centre, security and protection of electronic communication infrastructure and data, facilitation of the access to public information in electronic form".[21]

4. Digital Object Management

4.1 Pre-Ingest

4.1.1 Data collection *(collection policy)*

Data selection and appraisal play an important role in the acquisition of data in any archival setting. The collection policy indicates the principles and criteria by which the data service develops its data collection in order to serve the Designated Community. The collection development policy of KSSDC will be flexible and respond to future developments and shifting requirements that will influence the archive's data collections (technology, scientific standards, etc.).

The KSSDC will collect data and information from social sciences in a broader sense. This primary focus, however, will be on quantitative data in sociology, psychology, education science, political science and economics. The KSSDC will also collect and curate qualitative data, but with more careful selection and with consideration of available resources.

Nevertheless, in order for the research data provided to be ingested in the data archive, some criteria should be fulfilled, such as:

- Research that gives evidence and greater understanding of Kosovar society or parts of it (including international surveys);
- Research data where Kosovan researchers are involved;
- Comparative or continuous research (panel data, longitudinal surveys, and time series that allow examination of trends);
- Data produced with methodological excellence;
- Works of general interest for social science research (quality and exemplary studies with analytic potential).

---

[20]Law No.03/L – 172 On the protection of personal data of the Republic of Kosovo, available at: http://www.afapdp.org/wp-content/uploads/2012/01/Law-on-Personal-Data-Protection-Kosovo.pdf (Accessed on 3 May 2017)

[21] Law No. 04/L-145 On information society government bodies of the Republic of Kosovo, available at: http://www.kuvendikosoves.org/common/docs/ligjet/Law%20on%20information%20society%20government%20bodies.pdf (accessed on 4 May 2017)

In contrasts to the research data, which fulfil the above mentioned criteria, KSSDC will not accept research data with the following conditions:

- Data do not match the criteria of collection development policy and would be better dealt with at another institution (also legal conditions);
- Insufficient or poor quality documentation;
- Depositor and KSSDC do not agree on access and dissemination conditions;

There are currently no restrictions on file sizes, or number of files or studies. However, KSSDC reserves the right to refuse to accept materials which are of a volume which may make it difficult or impossible to process given the resources, staff, facilities, or capacities of KS DAS.

There are three main reasons why data are solicited and accepted: First, because they possess potential value for secondary use and analysis for research; Second, because they may serve teaching and learning purposes; And third, for validation reasons – to make possible the replication of research for important studies.

Eligible depositors as already mentioned will be students, researchers, teachers, researchers from other public institutions such as scientific institutes, government agencies, researchers from private institutions such as private research companies, NGOs etc. Each depositor will sign a contract with the KSSDC, which will stipulate the legal framework in which data is deposited as well as the rights and obligations of both signing parties such as copyright, ownership and access conditions.

All depositors should ensure that data meet the requirements of confidentiality and non-disclosure for data collected from human subjects. Datasets should not contain direct identifiers (names, addresses, telephone numbers, etc.) or indirect identifiers (which, when combined with other variables, may provide enough details to identify an individual). In-house confidentiality checks by KSSDC help to guarantee that confidentiality requirements are met.

The acquisition process will consist of three stages: 1) identifying potential datasets (localization); 2) negotiation with the data creators; and 3) receiving the relevant material. A very important source will be the list of potential data producers and consumers that has been established during the SEEDS project.

As KSSDC will be in the early stages of its existence, the **acquisition strategies** will be highly proactive and persistent. This will include open calls, announcements at conferences, and direct solicitation. No national research inventory for social science research data, where staff can actively search for and proactively seek data for deposit, is currently available for this purpose in Kosovo.
Similar as in some other countries, collaboration in the process of preparing and depositing data is entirely voluntary. This means that special effort will need to be invested into the process of negotiating and gaining access to the data sets:

- Good **working relations** with stakeholders and potential depositors (Kosovo`s Agency of Statistics (as well as entity agencies), research institutes and faculties in social sciences, private research companies and individual researchers) to identify and acquire new sources of data;
- **Links** with State/Entity/Historical Archives and Research Libraries;

- **Promotional activities** to raise awareness and encourage deposit of data collections (workshops, round tables, conferences, and training events);
- **Agreements** could be established with funding bodies for the disciplines in question in order to discuss deposit of material with grant-holders;
- **Data discovery** activities by staff members in publications (academic journals), news media, conferences, e-mail lists, and websites (e.g., universities' online catalogues and websites of research funding bodies).

The aim is to establish links and working relations with the stakeholders and partners mentioned above in order to establish long-term cooperation and a common framework for the deposit of data collections.

One of our efforts regarding the acquisition strategy was discussing with the MEST the creation of an amendment/norm which obliges research institutions and individual researchers whom are supported through public funds, to deposit their research data to the KSSDC after the end of the projects.

4.1.2 Data deposit at pre-ingest

There is still no legal requirement by law which obliges researchers from funders or research institutes to deposit their data to KSSDC and therefore, it is of a crucial importance to inform relevant stakeholders regarding the data formats and metadata standards which will be accepted and applied respectively.

In order to make the data and documentation accessible for other users and to be able to process and use them over time, they should be in a form that best ensures their longevity. The data and documentation **file formats** that are preferred at the ingest phase by the KSSDC are non-proprietary, openly documented, unencrypted, and uncompressed that are commonly used by the research community.

We therefore consider the following formats as appropriate for the Submission Information Package (SIP):
- Tabular data: **SPSS portable format (.por)**, SPSS (.sav), Stata (.dta), Excel or other spreadsheet format files, which can be converted to tab- or comma-delimited text, R (.txt);
- Text: Adobe Portable Document Format (PDF/A, PDF) (.pdf), plain text data, ASCII (.txt), Rich Text Format (RTF) (.rtf), Microsoft Office and Open Office documents;
- Audio: **Waveform Audio Format (WAV) (.wav)** from Microsoft, Audio Interchange File Format (AIFF) (.aif) from Apple, FLAC (.flac);
- Raster (bitmap) images: **TIFF (.tif)** ideally version 6 uncompressed, JPEG (.jpeg, .jpg), PNG (.png), GIF (.gif) and BMP (.bmp) only when created in this format, Adobe Portable Document Format (PDF/A, PDF) (.pdf);
- Vector images: DFX (.dfx), SVG (.svg);
- Video: **MPEG-2 (.mpg2)**, MPEG-4 (.mpg4), motion JPEG 2000 (.mj2).

Compressed files: are accepted as long as they can be uncompressed by using open and freely available software, such as **7-Zip** or **Winzip**.

The KSSDC orients itself towards international standards adopted by CESSDA AS and its service providers so that our data can be available and sharable throughout the international community. Regular reviews will be undertaken in order to guarantee best practice.

At KSSDC we will be compliant with the **Data Documentation Initiative (DDI)** metadata specification, version 2.5 or higher.[22] To be more precise, KSSDC will create the following types of metadata generated from accompanying documentation:

- **Descriptive metadata**: Describes a resource for purposes such as discovery and identification (contains elements such as title, abstract, author, and keywords). DDI complies with Dublin Core Metadata Initiative which describes a core set of 15 elements intended to facilitate discovery of electronic resources;[23]
- **Administrative metadata:** Contains information about the use, management, and encoding processes of digital objects over time (e.g., information about data creation);
- **Technical metadata**: Provides information about the overall system environment and provides the technical information needed to use data (e.g., file format, application used and operation system)
- **Structural metadata**: Describes the logical structure of a multidimensional object. The Metadata Encoding and Transmission Standard (METS) will be used where appropriate;[24]
- **Preservation metadata**: Provides information needed to archive and preserve a resource. The PREMIS Data Dictionary for Preservation Metadata defines a core set of preservation metadata elements and describes relationships between digital preservation entities: Intellectual entity, Object, Event, Agent, and Rights.[25]

The following two international archival standards could be used as well:

• **ISAD (G)**: International Standard Archival Description (General) or ISAD (G) is an international standard which provides guidelines for creating the content of an archival description (descriptive metadata). It promotes the creation of consistent and appropriate descriptions, facilitating the retrieval and exchange of information, and the integration of descriptions into a unified information system. There is a list of elements which are considered mandatory (six elements) when writing an archival description.[15] Based on ISAD (G) is the **Encoded Archival Description (EAD)** standard;

• **ISAAR (CPF)**: is the International Standard Archival Authority Record For Corporate Bodies, Persons and Families standardized by the International Council on Archives in 2003.[16] Whereas ISAD (G) describes the archival material itself, ISAAR (CPF) is constructed to give information about the creator (authority) of the archival material with additional information on history, status, functions, structure and relations – generally speaking about the provenance of the data. Again, there has been another international standard called **Encoded Archival Context (EAC),** which is derived from ISAAR (CPF).

The KSSDC is committed to developing its digital preservation policies and strategies in accordance with the (OAIS) Reference Model. This approach aligns with policies and practices of members of the

---

[22] Data Documentation Initiative (DDI): http://www.ddialliance.org/

[23] As a general remark, metadata standards should be machine readable (XML coded) in order to be processed automatically. XML is an open, well-supported, and widely adopted standard for encoding textual data, designed to be used regardless of the hardware and software environment.

[24] Metadata Encoding and Transmission Standard (METS) http://www.loc.gov/standards/mets/

[25] PREMIS Data Dictionary for Preservation Metadata: http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

Consortium of European Social Science Data Archives CESSDA, and with the SEEDS deliverables from prior Work Packages, as well as the common standards of the international infrastructure CESSDA. The OAIS functioning is explained in the first pages of this document.

4.2 Ingest

4.2.1 Data deposit at ingest

Data documentation submitted to KSSDC will explain how data were collected (context of data collection), what they mean, what are their content and structure, and specifies any manipulations that may have taken place. It consists of any useful information about the project and its results (e.g. research proposals, publications) or of any relevant information that may help to understand the data and their production, thus increasing the re-use potential (e.g. questionnaires, codebooks, methodology reports, user guides). The quality of the documentation can be significantly improved if its creation and collation is planned at the beginning of the data life cycle, during the project conception phase.

With each data deposited at the data service, a deposit contract is signed with the data producer. It is a legal agreement between the depositor and the archive that covers arrangements regarding usage rights, authenticity, data protection responsibilities, and disposal.

Ingest is the first functional component of the OAIS reference model. It includes the set of processes responsible for accepting the information submitted by data producers (Submission Information Package, SIP) and preparing it for the archival storage.

4.2.2 Data authenticity

Since ingest means the acceptance of the data from research project, there are some data authenticity check which will be integrated in our ingest tool such as:

- Scan and check for viruses;
- Identify and validate file format of the SIP: make sure that the files are what they pretend to be;
- Create checksums for each data and document file in order to guarantee data integrity during the transfer process;
- Generate Unique Identifiers (UI): Make sure that each study, dataset and file is assigned a reference number that is permanent and unique;
- Quality assurance routine checks should be carried out for completeness, integrity and validation of the data files, the submitted documentation and metadata:
  - Labels in the data file should match the question text in the questionnaire;
  - In the data file there can only be more variables than in the questionnaire, not less;
  - Run frequencies on the data and match it with publication;
  - Be careful with weighting.

4.2.3 Data protection

The KSSDC is aware of the conflicting gap between the tendency to provide open and easy access to research data and at the same time to protect the confidentiality of research participants and the rights of the data depositors. In order to ensure confidentiality, we rely on a combination of anonymisation measures, specific user contract conditions (restricted access) and informed consent among study participants. Our practices are in accordance with national law (**The Law on the Protection of Personal Data,** Official Gazette, no. 03/L – 172). The deposit contract indicates that the depositor has collected the data in conformity with existing national legislation on data protection and confirms that the data has been anonymised. In any case, the deposited data is screened by staff for disclosure risk.

4.3 Data preservation

4.3.1 Data management and Persistent Identifiers

Data Management is the third functional component of an OAIS. The Data Management function involves the maintenance of all the archived information and files including the management of metadata, communications with producers and users, and access statistics. The primary functions of Data Management are carried out by archive staff and include maintaining the databases of metadata for which it is responsible, performing queries on these databases, and generating reports in response to requests from other functional components within the OAIS (Ingest, Administration, Access). The data that is stored within the data service's archive should be attributed an ID and its metadata safely stored in the corresponding data-base.

Persistent identifiers are long-lasting references to a digital resource and objects. It has typically two components: a unique identifier; and a service that locates the resource over time even when it's location changes. The first helps to ensure the provenance of a digital resource, whilst the second will ensure that the identifier resolves to the correct current location. There are several persistent identifier schemes for research materials currently in use.[26]

4.3.2 Archival storage

Archival storage is the second component of OAIS. It aims to ensure that all the data (files) that have been accepted and added to the archive through the ingest process will remain identical and accessible. More specifically, this component manages the storage and maintenance of digital materials (AIP) so that the preserved information remains complete and readable over the long-term.

The primary goal of the KSSDC's preservation policy is therefore to ensure the longstanding accessibility of the digital material, ensuring the highest level of authenticity of any formats. However, preservation decisions at KSSDC will always be made within the context of its collections development policy, balancing the constraints of cost, research and historical value, and accessibility, alongside the requirements of levels of authenticity and legal admissibility.

---

[26] Digital Object Identifier (DOI), Handle, Persistent Uniform Resource Locator (PURL), Universal Resource Name (URN). For more information see: http://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers

<u>4.3.3 Preservation Planning</u>

The fourth functional component of an OAIS is Preservation Planning. This component provides the services and functions for monitoring the environment of the OAIS, providing recommendations and preservation plans to ensure that the information stored in the OAIS remains accessible to, and understandable by, the Designated Community over the long-term, even if the original computing environment becomes obsolete. Digital data are always linked to its support which itself is subject to rapid technical change. In other words, all file formats and physical storage media will become obsolete one time or another.

Preservation Planning functions include evaluating the contents of the archive and periodically recommending archival information updates, recommending the migration of current particular archive holdings, developing recommendations for the archive's standards and policies, providing periodic risk analysis reports, and monitoring changes in the technological environment and within the community.

The aims of the preservation policy of KSSDC are to:

• Become a trusted digital repository in order to serve a designated user community;
• Retain authenticity, integrity and reliability of preserved datasets;
• Ensure that all data collections are protected and safely stored;
• Ensure that the relevant level of information security is applied to each data collection;
• Apply good practice in preservation management.

The KSSDC will follow established best practices for managing datasets over the long-term. It assumes responsibility for the preservation and accessibility of the digital material in its collection. Preservation in a digital archive usually means preserving the significant properties or the essential characteristics of data for these to maintain their meaning over time. Any changes (by machine/application or an individual) in the preservation process should be properly documented (either as a special document or within the archiving application).

The KSSDC will try to ensure that it is at the leading edge of technical advances by taking a strategic approach to long-term digital preservation, and by monitoring hardware and software developments and migrating its collections accordingly. The technical development of storage technology must be under constant observation and review: monitoring technological changes that will affect preservation and migration decisions, files suitable for archiving, the structure and design of the information packages, and hardware (server, storage media) and software (applications).

The functional entity "Preservation Planning" encompasses tasks such as development of preservation strategies and standards, development of packaging designs and migration plans, and monitoring of technology (innovations in storage and access technologies) and the designated community (shifts in scope or expectations). The data service monitors the technical fitness of its archive, does regular risk assessments of the stored digital objects (which includes technology monitoring for the different object types), and plans for preservation actions.

Migration planning, archive standards and policies as well as technology watch report are usually gathered in the preservation policy of a data service. Digital objects may become unreadable or obsolete after a certain number of years. The need might arise to migrate file formats that have come close to obsolescence to new file formats that are more sustainable and guarantee future usability. After migration the original manifestation of the data file will be maintained and all subsequently generated manifestations of the original files. In this case, we adhere to the principle of reversibility: being able to revert to an earlier version of a digital file after migration. We also fully document the migration process in the form of a detailed migration history as part of the metadata associated with the data file.

4.4 Access and data provision

4.4.1 Data discoverability and accessibility

The data producer can, during the negotiation phase, place the data under an **embargo** or special conditions, which means that data are not accessible to data users for a pre-defined period or only with certain restrictions. However, in the spirit of Open Access we highly recommended that data are made available to users as soon as possible after a research project has come to an end. The data depositor can also impose certain conditions on the use of the data (e.g., to be informed beforehand about who is going to use the data and for what purposes).

The fifth functional component of OAIS relates to access. This component provides the services and functions that support data users in determining the existence, description, location, and availability of information stored in the data archive, allowing them to request and receive data and documentation. To be more precise, as part of this functional entity, the Dissemination Information Packages (DIP) are generated through addition of descriptive metadata, and the dissemination request is processed accordingly. Access is also responsible for implementing any security or access control mechanisms associated with the archives content.

4.4.2 Access control

Access to data at KSSDC will be controlled depending on the level of sensitivity of the data and the specific conditions determined by the depositor. The latter are set out in the **end user license** a version of which all users must sign before being given access to any dataset. The end user licence describes the legal framework under which the material is distributed. It specifies the conditions under which a user is granted access to data and how they may use those data (citation in scientific fashion and use following typical scientific ethical norms of conduct). According to the sensitivity of data or due to additional access conditions negotiated with the data producer special conditions and licences might apply.

4.5 Outreach

Data management is a set of skills needed for handling data throughout the life cycle of a research project. Good data management practices mean more efficient research and a higher quality research product.

Promotional activities to raise visibility have already begun and should be on-going. These will include targeted messages to specific audiences by way of different means and platforms. The means could include meetings, workshops and conferences, email and letter campaigns and invited visits by key stakeholders. Another way of communication with our community is through the adequate use of the social media such as Facebook, Twitter, LinkedIn groups etc.

5. Technical infrastructure and risk management

5.1 Technical infrastructure

The Dataverse[27] software is being developed at Harvard's Institute for Quantitative Social Science (IQSS) with the help of many contributors and collaborators all over the world. This is a successor of the previous VDC - Virtual Data Center project (1999-2006). It is suitable for archiving the research data from any discipline and it covers basic elements of the OAIS reference model. It has API, which enables fairly easy integration with other systems. The software is built using Java programming language and technologies around it and it uses PostgreSQL as database server. There are 22 Dataverse installations available online with over 2000 dataverses and almost 50.000 datasets published. This is far less when comparing to EPrints or DSpace but Dataverse is a special type of repository, much more research data oriented. It's also worth to mention that 50% of data providers come from the Social Sciences. The Dataverse application code is licensed under the Apache License, Version 2.0 which implies distribution of the source code and the right to reuse, change it and make derivative works as long as the source code is further distributed and required notices are included. This permissive license contains a patent license from the contributors of the code.

After the testing done by our data specialist as well as the positive aspects of the Dataverse, we have come into conclusion that the Dataverse platform would be the best solution whenever the real deal with establishing a data archive comes to life. We are aware of the possible challenges, but we believe that having our own installation as well as using help from SEEDS partners in the region, who are going to use or who already use Dataverse, will be very helpful and appreciated.[28]

---

[27] http://dataverse.org/
[28] D9 report on integration on tool evaluation and integration, http://seedsproject.ch/wp-content/uploads/2015/06/D9_FINAL.pdf , p.16

SEEDS: D5 – Policy and procedures: Kosovo

6. Resources

Audit and certification of trustworthy digital repositories (ISO 16363) - http://www.iso16363.org/standards/iso-16363/

CESSDA – http://cessda.net/

CESSDA SaW – http://cessdasaw.eu/

- D3.1 – Heuristic Maturity Development Model (CESSDA-CDM) http://cessdasaw.eu/content/uploads/2016/06/D3.1.pdf
- Guide for Developing National Data Service Plans - https://cessda.net/eng/CESSDA-Services/Projects/CESSDA-SaW/Work-Packages/WP3/Guide-for-Developing-National-Data-Service-Plans

Consultative Committee for Space Data Systems (CCSDS) - https://public.ccsds.org/default.aspx

Data Seal of Approval (DSA) - http://www.datasealofapproval.org/en/

DIN 31644 – Nestor seal for trustworthy digital archives - http://www.langzeitarchivierung.de/Subsites/nestor/EN/Siegel/siegel_node.html

DIN Working Group "Trustworthy Archives – Certification" - http://www.din.de/en/getting-involved/standards-committees/nid

FORS - http://forscenter.ch/en/

- Deposit contract FORS: https://forsbase.unil.ch/media/general_documentation/en/deposit_contract_FORS_en.pdf
- User contract FORS: https://forsbase.unil.ch/media/general_documentation/en/download_contract_en.pdf
- Collections Policy FORS: http://forscenter.ch/wp-content/uploads/2015/09/Collections-Policy_E_v2.pdf
- Preservation Policy FORS: http://forscenter.ch/wp-content/uploads/2015/05/Preservation-Policy1.pdf

ISO 16363 - http://www.iso16363.org/standards/iso-16363/

KRDS (Keeping Research Data Save) Activity Model – User Guide: http://www.beagrie.com/static/resource/KeepingResearchDataSafe_UserGuide_v2.pdf

OAIS – Reference Model for an Open Archival Information System, CCSDS 650.0-M-2, Consultative Committee for Space Data Systems: Washington, DC, 2002 https://public.ccsds.org/pubs/650x0m2.pdf

Research Data Alliance (RDA) - https://www.rd-alliance.org

SEEDS - http://seedsproject.ch/

- D4 – Establishment plan: http://seedsproject.ch/wp-content/uploads/2015/06/Establishment-Plan_Montenegro.pdf; http://seedsproject.ch/wp-

content/uploads/2015/06/Establishment-Plan_Kosovo.pdf; http://seedsproject.ch/wp-content/uploads/2015/06/Establishment-Plan-ALBANIA_FINAL.pdf; http://seedsproject.ch/wp-content/uploads/2015/06/Establishment-plan-Macedonia.pdf

- D9 – Report on technical improvements: (unpublished document, at the time being)

SERSCIDA – http://serscida.eu/

- D4.2 – Data Service Training Manual: http://www.serscida.eu/images/deliverables/SERSCIDA_D_4_2_Training_Materials_V1_2.pdf
- D5.1 – Documents and Materials for Social Science Digital Data Archives: http://www.serscida.eu/images/deliverables/D5.1_FINAL.pdf
- D5.3 – Report on Prototype Database: http://www.serscida.eu/images/deliverables/D5.3_FINAL.pdf

World Data System (ICSU/WDS) - https://www.icsu-wds.org