South-Eastern European Data Services

D5 – Policy and procedures: Macedonia

Deliverable Lead: FORS, ISPJR

Related Work package: WP1

Author(s):   Aneta Cekikj (ISPJR)

Klime Babunski (ISPJR)

Vesna Zabijakin-Chatleska (ISPJR)

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	March 23, 2017	Draft Final Version	ISPJR

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.

A crucial perquisite 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 want 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 into various Capability Process Areas (CPA), each of which having its own purpose. Within each CPA there are several activities defined to achieve the objective(s) of that CPA.
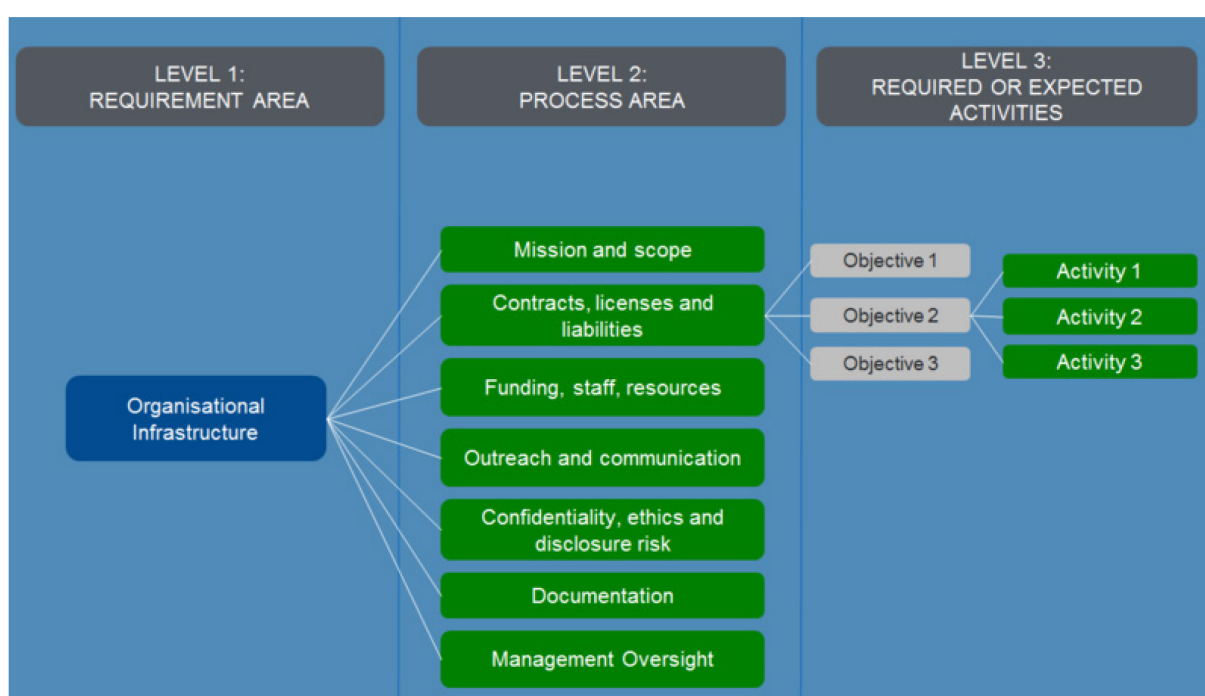
CESSDA SaW Capability Development Model (CESSDA-CDM), in: CESSDA SaW, Deliverable 3.1 Heuristic Maturity Development Model, 2016, p. 11 (http://cessdasaw.eu/content/uploads/2016/06/D3.1.pdf).

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 preserving and maintaining 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 related 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 the 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 AIPs 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]
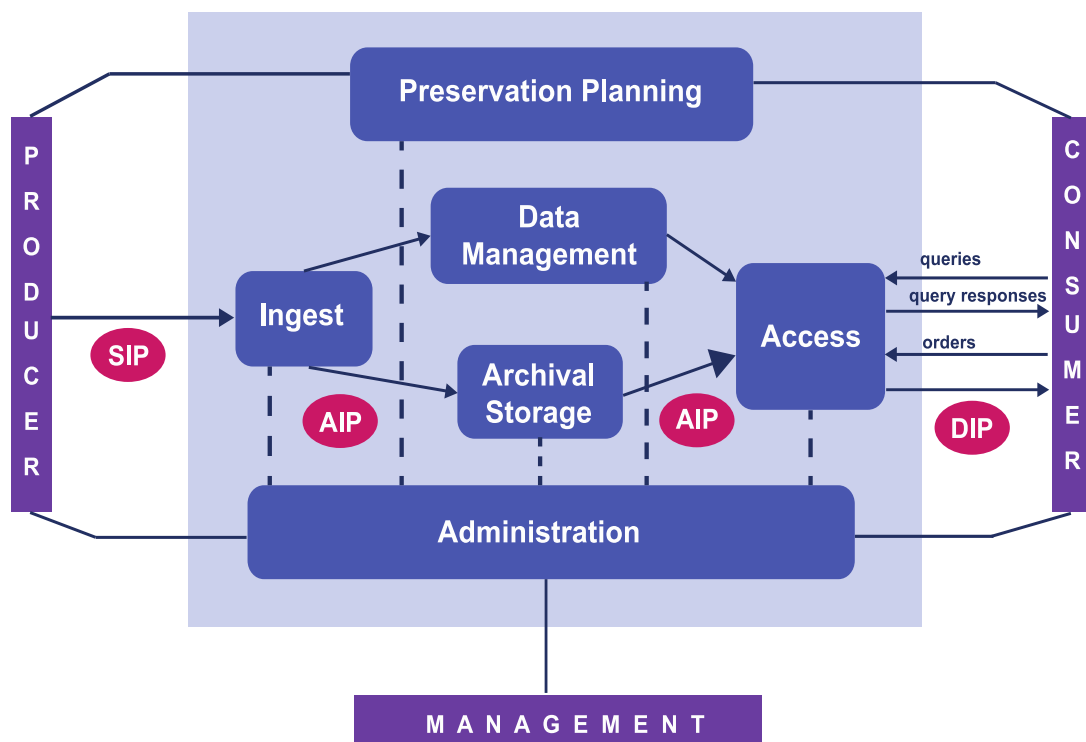
Reference Model for an Open Archival Information System (OAIS); CCSDS 650.0-M-2; Consultative Committee for Space Data Systems: Washington, DC, 2002, p. 4-1 (https://public.ccsds.org/pubs/650x0m2.pdf).

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, 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]

---

[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

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 the Data Seal of Approval, the Repository Audit and Certification Working Group of the Consultative 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 and activities

MK DASS is national research infrastructure and public service which provides long-term preservation and distribution of research data in the social sciences in the Republic of Macedonia.[13] Our data service serves the research community, including researchers, teachers and students, as well as the broad public interested in social science research outputs.

---

[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://mk.seedsproject.ffzg.hr/

We provide curation of research data produced by the research community in the country, and access to these data to researchers and the broader public.

Our **mission** is:

- To acquire, properly document, preserve and disseminate research data in the social sciences produced by researchers in Macedonia;
- To advance social science research by making the collected studies available for secondary analysis;
- To bring social science outputs to the public by providing a searchable inventory of projects, and related publications and datasets.

The **key activities** of MK DASS are:

- Preservation and dissemination of quantitative and qualitative research data;
- Provision of access and support for both users and data providers including mediation between the demands of data users and data providers;
- Permanent monitoring of compliance with international standards in the field of data management and preservation;
- Collaboration with the wider scientific community in the field of collecting and distributing data.

Our **additional services and activities** include:

- Training in data management and methodology of scientific research for researchers;
- Promotion of Open Access initiatives;
- International cooperation and collaborative projects.

3.2 User Orientation and Designated Community (definition and monitoring)

The Macedonian Social Science Data Archive (MK DASS)'s designated community consists of producers and users of research data. The producers of social science research data in Macedonia are the higher education and research institutions part of the public and private universities, other public and private research institutions, individual researchers, public institutions such as the State statistical office, as well as non-governmental organizations and think tanks. Our users are registered researchers from the country and abroad, and other interested parties such as public institutions, policy makers, journalists, students and others.

In addition to research institutions, our key national stakeholders are the Ministry of Education and Science of the R. Macedonia, the State Statistical Office of the R. Macedonia, the National Archive of the R. Macedonia, Macedonian Academy of Arts and Sciences, the National and University Library St. Clement of Ohrid- Skopje, the Inter-University conference of the R. Macedonia, NGOs, foundations, think tanks, chambers of commerce, as well as other relevant institutions. MK DASS will maintain close contacts and cooperation with academic research institutions in the country and the other stakeholders.

We are going to carry out a number of promotional activities, including an e-mail digest for researchers, newsletters, training courses for data management and methodology of social science research, social networking activities and others in order to attract researchers in the process of data acquisition. We are also going to organize conferences and promotional events in order to obtain the necessary support from the national institutions which develop and fund scientific activity in the country, as well as the other stakeholders relevant for the work of MK DASS.

3.3 Financial sustainability and Resources

In order for this type of national service to function properly, the funding is usually secured through the relevant Ministry, or existing funding bodies in the scientific sector of a given country. In the case of Macedonia this is the Ministry of Education and Science of the R. Macedonia. In addition to this, funding will be sought from donor organizations from the country and abroad, as well as through collaboration with (members of) the Consortium of European Social Science Data Archives (CESSDA) in the implementation of specific projects.

3.4 Staffing

3.4.1 Roles and responsibilities

The Macedonian Social Science Data Archive (MK DASS) will be organized as a separate unit at the Institute for Sociological, Political and Juridical Research (ISPJR) at Ss. Cyril and Methodius University in Skopje. The internal structure and the number of staff of the MK DASS are defined in accordance with the available resources and volume of work (workload). However, MK DASS will include at least the following job positions and formal tasks, assigned to individuals and units: MK DASS manager; Data Specialist (one position) and IT Specialist (one position). Through internal recruitment, all these job positions will be assigned to people who already are full-time employees of the ISPJR and the assigned tasks will be part of their overall job description. An exception is the IT specialist who needs to be recruited as a new employee or to be hired through outsourcing. A possible solution could also be to rely on staff from the University Computer Centre of the Ss. Cyril and Methodius University.

The ISPJR has four employees who are already trained to work for the Archive as Data Specialists within the SEEDS project and RRPP Data Rescue Project. In the first phase, two of them will be appointed: a MK DASS manager and one Data Specialist. If the workload increases, then one more Data Specialist will be recruited from the trained employees in the ISPJR.

Working hours per month will be distributed evenly among staff members, and according to current workload. The staff members' compensation will be determined according to job position and according to hours worked on a monthly basis.

3.4.2 Organizational structure

As a unit of ISPJR, MK DASS management structure will be organized as a flat structure, with 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 MK DASS manager. Also, employees will adopt a team-based approach and close collaboration between different units/tasks on a daily basis, and also, will collaborate with all other sectors, units, and centers in ISPJR. An organizational chart is developed that shows the division of tasks, responsibilities, reporting relationships, and hierarchy (see MK DASS Organizational chart).

MK DASS Organizational chart
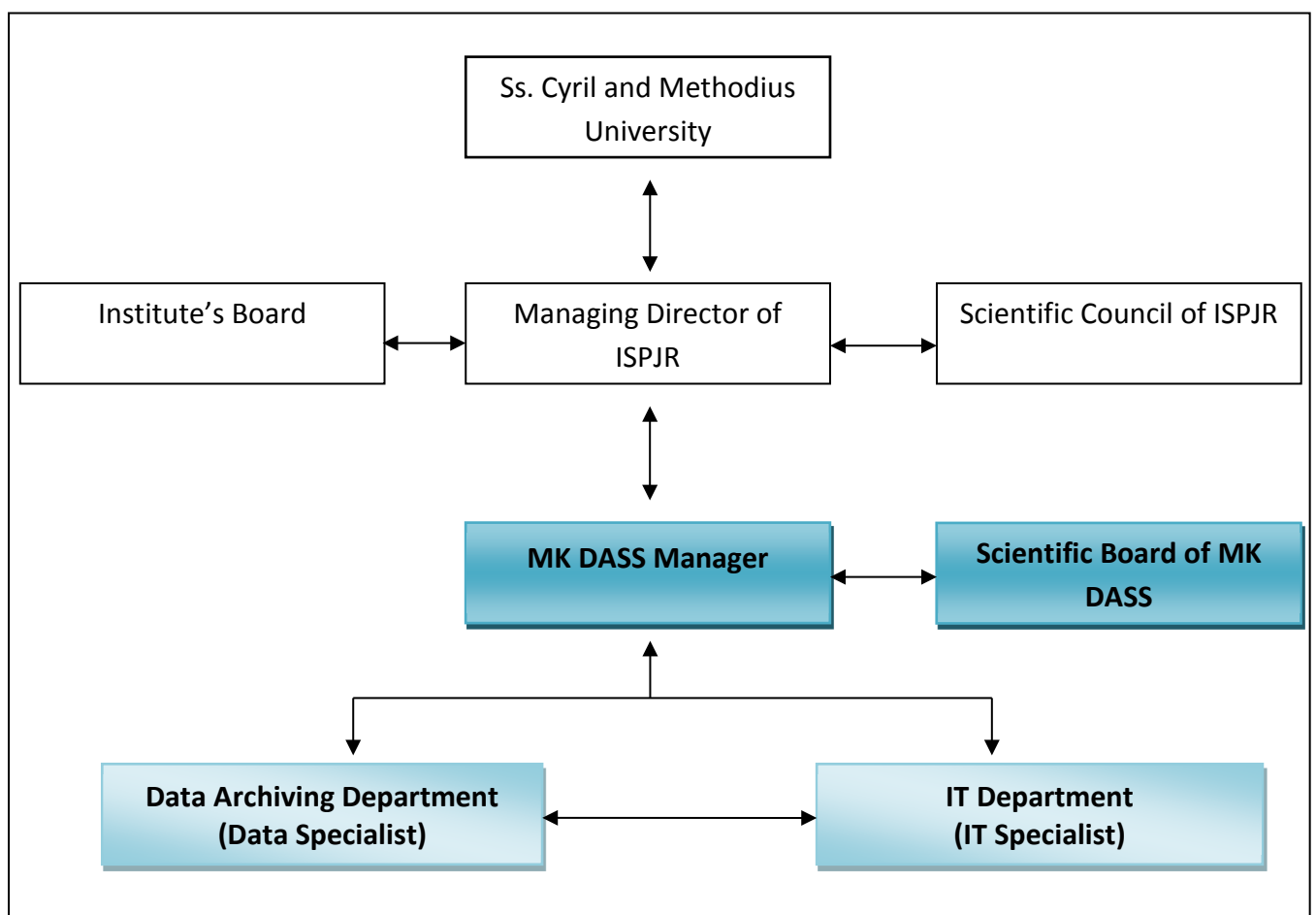
3.4.3 Roles and responsibilities

The **MK DASS Manager** is appointed by the Scientific Council of ISPJR, for a four-year period. He/She is responsible for **the i**mplementation of a reliable, adequate and quality data service, and an unimpeded functioning of the Archive. The MK DASS Manager should ensure the operation of the Archive to be focused on implementation/execution of its strategic goals, which are pre-determined and clearly defined. He/She needs to take care of the continuous development of skills and abilities of employees. The MK DASS Manager is responsible for establishing and maintaining good external

relationships with key stakeholders and cooperating institutions. He/she is responsible for a successful presentation of the Archive and its mission outside the organization.

The MK DASS Manager should be able to communicate in a positive manner in order to promote the need for open access to research data while emphasizing the benefits that this approach provides, to underline the importance of transparency in scientific research in the social sciences research community, and its importance in a wider social context.

The **Data specialist** has duties and responsibilities related to the entire process of data archiving - solicitation, curation and dissemination of research data.

The **IT Specialist** is responsible for the overall technical support of the system and ensures appropriate functioning of the Archive in its storage, preservation and dissemination activities. He/she is also responsible for maintaining the website of the MK DASS. In the initial operation, the Archive will use the services of an IT specialist (through outsourcing) who is currently full-time employed in the University Computer Centre of the Ss. Cyril and Methodius University.

3.4.4 Job descriptions

MK DASS manager
Responsibilities and duties:
- Achieves defined strategic mission and goals of the organization.
- Prepares strategic and operational plan for the next three years.
- Assures financial viability; Planning, allocating and controlling financial resources.
- Makes decisions; Organizing, coordinating and supervising daily activities according to international standards for data management and data preservation.
- Develops plan for continuous improvement of the quality of work; Ensures employee training and development through participation at conferences, workshops, seminars, and trainings on data management, metadata documentation, data preservation, and other fields relevant for the work of the MK DASS.
- Reports about MK DASS activities to the Managing Director of ISPJR and to the Scientific Council of ISPJR.
- Establishes relations and collaboration with designated communities, stakeholders and partner institutions/organizations.
- Initiates and conducts international cooperation.
- Manages the application processes for both national and international projects.

Required knowledge, skills and abilities:
- PhD in a field of the social sciences.
- Specific knowledge in data preservation and data management.
- English language proficiency.
- Experience and knowledge in international projects and project management.
- Reference in the field of data management and data archiving.
- Good management skills and communication skills, skills for team leader.
- Solid social networking skills.

Reports to: Managing director of ISPJR.

Supervises: Data specialist and IT specialist.

Works with: All Institute sectors, units, centers and labs.

Outside the ISPJR: Establishing and maintaining good external relations, which means close collaboration with stakeholders and clients, as well as partnership with other related national and international data services or corresponding regional and international organizations.

Data specialist

Responsibilities and duties:
- Responsiveness and communication with data providers (depositors) and data users.
- Metadata and data quality control and processing.
- Preparation of data for secondary use; data conversion in different formats if needed.
- Manage and control access to research data.
- Mediate communication between registered data users and data providers.
- Provision of data services to third parties.
- Provision of training to the registered data users.

Knowledge, skills and abilities:
- Minimum of bachelor degree in social sciences.
- Knowledge in data sciences.
- Knowledge of statistics and statistical software packages and tools.
- English language proficiency.
- Good communication skills.
- Ability to work in a team.

Reports to: MK DASS Manager

Works with: IT specialist and all Institute sectors, units, centers and labs.

Outside the ISPJR: communication and close collaboration with data depositors and data users.

IT specialist

Responsibilities and duties:
- Manages the servers, hardware and software used for storage and distribution of data.
- Performs regular backup and system recovery.
- Maintains and updates the website of MK DASS.

Knowledge, skills and abilities:

- Bachelor degree in computer science and engineering.
- Knowledge of server and network protocols.

- English language proficiency.
- Knowledge and skills in adaptation of new or different technologies.
- Good communication skills.
- Ability to work in a team.

Reports to: MK DASS Manager

Works with: Data specialist

3.5 Legal framework

3.5.1 Legal status and legal responsibility

Since MK DASS is a unit of ISPJR, the Archive's legal status and legal responsibilities are subject to the internal legal acts of ISPJR and Ss. Cyril and Methodius University, which in turn comply with the 2008 Law on Scientific and Research Activities, the 2008 Law on Higher Education and various bylaws.

3.5.2 National laws

The functioning of MK DASS is in line with the relevant legal framework which regulates the area of preservation, access and use of digital data.

MK DASS is operating and is subject to the following relevant legislations:

- **Law on copyrights and related rights** (Official gazette of the R. Macedonia No.115/10, 140/10 and 51/11). This law regulates the copyrights of authors over their work, among others, the rights of "(…) authors of data sets over their works, or related rights…", as well as the practicing and protection of copyrights and related rights (Article 1). According to this law, related rights can also be "data bases and their authors". Related rights are regulated with the General provisions on related rights, especially in Part 6- The rights of authors of data bases (Article 118- 122).
- **Law on archival material** (Official gazette of the R. Macedonia No.95 from 26.7.2012) regulates the protection, storage, processing and use of archival material. According to this law, archival material is defined as "documentary material of permanent value for R. Macedonia, the science, culture, its possessors, as well as for other needs." Research databases are defined as "nonconventional archival and documentary material", as "data kept/noted in data bases which can serve as a basis for creation of conventional documents (…)". This law also regulates the rights and obligations of the possessors of archival and documentary material, which can be institutions, legal or natural persons, and carriers of the right of possession over the archival and documentary material. The law defines the right of access and use of the data. The electronic archival and documentary material is regulated separately by the law where provisions with regard to "electronic archive" and "electronic data" exist, concerning their preservation, accessibility, security, confidentiality and authenticity.

- **Law on protection of personal data** (Official gazette of the R. Macedonia No.7/05, 103/08, 124/08, 124/10, 135/11, 43/14 and 153/15) regulates the protection of personal data through the discourse of protection of personal liberties and freedoms of individuals, and the right of privacy in connection with the usage of personal data. In particular, this law regulates secrecy and protection in processing the personal data of citizens of the Republic of Macedonia. Article 2 determines the meaning of certain terms used in this law. "Personal data" shall be any information about an individual whose identity may be determined directly or indirectly. "Personal data processing" is "every operation or a sum of operations performed on personal data, automatically or otherwise. The law applies to entirely or partly automated personal data processing. Personal data protection is guaranteed to every person without discrimination on any grounds. The policy for protection of the privacy, personal and family life of the personal data subject from their unauthorized use, shall be applied when personal data are used for historic, scientific or statistical research purposes, and in as short term as possible the data shall be made anonymous. Finally, there is the question of revealing data to users. Regarding the provision of personal data, the Law prescribes that "The personal data processed in scientific and research and statistical purposes may not be revealed to the user in a form which enables identification of the person to whom the personal data refer".
- **Law on free access to information of public character** (Official Gazette of the R. Macedonia N. 13/06) provide transparency in the work of the state and public bodies, and also to provide free access to information that is produced and/or possessed by such entities. According to the law, the *information holder* can be "… (a) public institution or service,[…](a) legal and physical person that have public duties and their activity is in public interest …" (Article 1- 3). *Information requester* is any legal or natural entity without any discrimination, which includes domestic and foreign entities (Article 3 and 4). The main logic of the law is that any information is under "free access" if it is in the public interest and especially if it is produced and held based on public money. On the other hand, the main reasons for denying free access are: protection of privacy, the data or information is classified, or they are part of criminal or court investigation and procedure, and also, when displaying of the data or information will bring harm to someone's commercial interests and/or to someone's intellectual or industrial property, and when the data/information is in very draft stage and will bring confusion to the public.

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 MK DASS will be flexible and respond to future developments and shifting requirements that will influence the archive's data collections (technology, scientific standards, etc.).

The concept of the collection policy aims to encompass quantitative and qualitative research data in the disciplines of anthropology, sociology, political science, communication science, psychology, law,

education science, economics, business and management, demographics and related social science disciplines.

But regardless of the discipline, in order **to be accepted the research has to provide:**
- greater evidence and understanding of Macedonian society or parts of it;
- comparison or continuous research (panel data, longitudinal surveys, and time series that allow examination of trends);
- good methodological quality or excellence;
- general interest for social science research (quality and exemplary studies with analytic potential).

In parallel to this there are three **additional reasons why certain data are accepted**. First, because they possess potential value for secondary use and analysis for research. Second, because they may serve teaching and learning purposes. And finally, for validation reasons - to make possible the replication of research for important studies.

MK DASS reserves the right to **refuse** to accept material under the following conditions:
- Data do not match the criteria of the collection development policy and would be better dealt with at another institution (also legal conditions);
- Insufficient or poor quality documentation;
- Depositor and MK DASS do not agree on access and dissemination conditions;
- If the data are of a nature or volume which may make it difficult or impossible to process given the resources, staff, facilities, or capacities of MK DASS.

Eligible depositors are researchers, members of research institutions, think-tanks, NGOs, government departments, State statistical office, and public and commercial sector sources (e.g., marketing research firms). Depositors may only deposit their own work and will be expected to sign a data deposit agreement.

In general, the Archive will prefer that the researcher depositing the data provides its anonymisation. This will be one condition in the data collection process. At the same time, MK DASS will be willing to provide guidelines and help to the research team for the purpose of data anonymization measures. An additional condition within the data collection of the Archive is the rule that all data have to be accompanied with appropriate metadata and documentation, as a necessary material that will provide secondary use of data.

As MK DASS will not only be in an early stage of its existence but also the first and unique institution of this type in Macedonia, the **acquisition strategy** will be highly proactive and persistent. This will include open calls, announcements at conferences, and direct solicitation. Currently, there is no national research inventory for social science research data in Macedonia where one can actively search for data or deposit data.

Being a completely new institution, special effort will need to be invested in the process of negotiation:

- Good **working relations** with stakeholders and potential depositors and acquire new sources of data;
- **Links** with Historical Archives and 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.

The final task of this strategy will be to establish and institutionalize a norm, (present in some other countries, like UKDA in the United Kingdom), within the appropriate legislation that will oblige research institutions and researchers to provide information about publicly funded projects and to deposit the research data in MK DASS.

4.1.2 Data deposit at pre-ingest

Due to the fact that there is no legal requirement from funders, research institutes or journals for researchers to deposit their data at MK DASS, it is very important to inform and to explain to the designated community which data formats will be accepted and which metadata standards will be applied.

When it comes to data and documentation file formats that are accepted (Submission Information Package - SIP), we prefer formats that are most likely to be accessible in the future. In other words, formats that are non-proprietary, openly documented, unencrypted and uncompressed that are commonly used by the research community.

The following formats are appropriate for the 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);
- Audio: Waveform Audio Format (WAV) (.wav) from Microsoft, Audio Interchange File Format (AIFF) (.aif) from Apple, FLAC (.flac); Bitstream (.ogg)
- Images: TIFF (.tif) ideally version 6 uncompressed, JPEG (.jpeg, .jpg) only when created in this format, Adobe Portable Document Format (PDF/A, PDF) (.pdf), RAW image format (.raw), Photoshop files (.psd); Scalable Vector Graphics (.svg)
- Video: 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.

MK DASS will follow the OAIS model (see chapter 2.2) with the aim to enable a sufficient amount of information on the study level (metadata) that enables final users to independently understand individual parts of the data file or additional materials.

In doing this the Archive will be oriented 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 MK DASS we will be compliant with the **Data Documentation Initiative (DDI)** metadata specification, version 3.2.[14] This means that MK DASS will create the following types of metadata on the basis of accompanying documentation provided from the depositor.[15]

- **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;[16]
- **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;[17]
- **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.[18]

Because the number of digital objects that require more reliable and stable references is constantly growing (regardless of their location and across system change), MK DASS will aim to assign a persistent identifier (DOI (Digital Object Identifier)/URN (Uniform Resource Name) to every study coming into the archive.

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.[19]15 Based on ISAD(G) is the **Encoded Archival Description (EAD)** standard;

---

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

[15] 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.

[16] Dublin Core Metadata Initiative: http://dublincore.org/

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

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

[19] General International Standard Archival Description, ISAD (G): http://www.ica.org/10207/standards/isadg-generalinternational-standard-archival-description-second-edition.html

- **ISAAR(CPF)**: is the International Standard Archival Authority Record For Corporate Bodies, Persons and Families standardised by the International Council on Archives in 2003.[20] 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)

4.2 Ingest

4.2.1 Data deposit at ingest

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.

Data documentation that will be submitted to MK DASS should explain how data were collected (context of data collection), what they mean, what is their content and structure, and specify any manipulations that may have taken place. It should contain all 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 is planned at the beginning of the project. To provide appropriate understanding within the research community and with the aim to improve the quality of data documentation, MK DASS will produce a Guide for Data Producers, organise appropriate trainings and will have to establish close cooperation with the researchers.

Some of the main issues that MK DASS will have to talk about with data providers will be: confidentiality, anonymization of research data and copyrights. This includes the rights of the archive and its successor to ingest data, to subject them to digital curation, to preserve and allow access to them under specified condition of access – for example possible embargo, special treatment of sensitive data and/or special restrictions of commercial use of research data.

On this basis, the Archive and data provider will sign an agreement where mutual rights and obligations will be given (the nature of curation and preservation of the data, and rights and conditions of access to data and duties in data dissemination).

---

[20] International Standard Archival Authority Record For Corporate Bodies, Persons and Families, ISAAR (CPF) http://www.ica.org/10203/standards/isaar-cpf-international-standard-archival-authority-record-forcorporate-bodies-persons-and-families-2nd-edition.html

<u>4.2.2 Data authenticity</u>

As already mentioned, Ingest includes a set of activities through which the information and data from research projects are processed and accepted. The following activities are most often included in the process of accepting the information and data submitted from researchers:

- Scan and check for viruses;
- Review of completeness of the study data, metadata, and documentation;
- Identification and validation of file formats of the SIP: making sure that the files are what they are declared to be;
- Checks if the scope of documentation is sufficient (e.g., variables are described and can easily be attributed to the questions in the questionnaire);
- Comparison of data sets and data documentation to ensure that there is a match between variables, variable labels, and value labels, and that the number of observations and variables match the stated numbers in the documentation;
- Checks for anonymisation strategies if data are sensitive (proper name, postal code, profession, etc.);
- 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;

<u>4.2.3 Data protection</u>

MK DASS 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 anonymization measures, specific user contract conditions (restricted access) and informed consent among study participants. Our practices are in accordance with national law (**Law on protection of personal data,** Official gazette of the R. Macedonia No.7/05, 103/08, 124/08, 124/10, 135/11, 43/14 and 153/15). The deposit contract should indicate 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 *(preservation policy)*

<u>4.3.1 Data management and Persistent Identifiers</u>

Data management and preservation is central element of data storage and maintaining of digital materials, which is an essential function of archiving an information package as a component of OAIS. The task here is to ensure that what is added to the archive from the ingest process remains identical and accessible, which actually means that the stored and preserved material remains complete and readable over the long-term.

Primary goals of MK DASS preservation policy will be 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.

MK DASS will follow the established best practices for managing datasets over the long-term.

The issue of ID attribution to the stored data in the Archive and the issue of persistent identifiers will be solved in accordance with existing national research practices, and in consultation with the appropriate national institutions. .

4.3.2 Archival storage

This is the second functional component of OAIS. The aim of archival storage is to ensure that the result of the ingest process (all files that have been added to the archive; data files with all accompanying metadata files (AIP – Archive Information Package)) will remain identical, accessible and readable over the long-term.

Hence, within its preservation policy MK DASS will ensure the longstanding accessibility of the digital material and the highest level of authenticity of any formats, in accordance with the constraints of cost and research value of the data. The AIP will be stored in a separate location from the dissemination information package (DIP).

4.3.3 Preservation Planning

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). MK DASS 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 and 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

This is another component of OAIS and it is related to providing 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. It means that within this function the Dissemination Information Packages (DIP) are generated through addition of descriptive metadata, and the dissemination request is processed

accordingly. When it comes to access, it is necessary to implement security or access control mechanisms associated with the archive's content.

The data will be discoverable through: (1) the newsletter of MK DASS that will be distributed to the designated community for announcing a new project in the Archive; (2) the catalog – searchable inventory; and (3) upon specific requests by users.

Due to the lack of appropriate resources at the beginning we will use SEEDSbase[21] developed by FORS, as a technical tool to provide access to the metadata and data.

4.4.2 Access control

MK DASS does everything possible to provide open and easy access to data, while at the same time protecting the confidentiality of research participants and the rights of the data depositors. Metadata and other similar information about archived projects will be open to all, including the public, but the access to data files will be managed depending on the conditions set by the depositor of the data. The control of data files will range from simple filing of the registration form up to fulfilling conditions like embargo or obligatory communication with depositor, from the side of the user.

The conditions that will form the basis for access control will be part of the user contract.

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. A data management plan (DMP) is a written document that describes the data a researcher expects to acquire or generate during the course of a research project, how they will manage, describe, analyse, and store those data, and what mechanisms they will use at the end of a project to share and preserve the data. It is also intended to meet funder requirements, and help others to use the data if shared. Our data service offers training courses regarding DMP and other workshops and help for data depositors and users equally.

MK DASS is going to organize a number of promotional activities, including e-mail digest for researchers, newsletters, training courses for data management and methodology of social science research, social networking activities and others.

5. Technical infrastructure and risk management

5.1 Technical infrastructure

Having in mind the limited resources (lack of human resources with appropriate IT knowledge and experience; lack of funding for regular daily activities and especially lack of funding for hardware and software equipment) MK DASS will use SEEDSbase as a tool for archiving.

In parallel to the solving of these establishing problems MK DASS will take decision(s) and find solutions regarding the broader technical infrastructure.

---

[21] https://seedsdata.unil.ch/

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-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