D6 – Report on integration of technical system: Macedonia

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

FNSNF
Swiss National Science Foundation

Deliverable Lead:	FORS
Related Work package:	WP1
Author(s):	Aneta Cekikj (ISPJR) Klime Babunski (ISPJR) Bojana Tasic (FORS) Irena Vipavc Brvar (ADP) Maja Dolinar (ADP)
Dissemination level:	Public (PU)
Submission date:	30rd April 2017
Project Acronym:	SEEDS
Website:	http://www.seedsproject.ch
Call:	Scientific cooperation between Eastern Europe and Switzerland (SCOPES 2013-2016)
Start date of project:	1st May 2015
Duration:	24 months

Version History

Version	Date	Changes	Modified by
1.0	February 28, 2017	Released version	FORS
2.0	April 14, 2017	Draft version	ISPRJ
2.1	April 21, 2017	Revision	UL- ADP
3.0	April 28, 2017	Final	FORS

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
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
University of Zagreb, Faculty of Humanities and Social Sciences, Zagreb	FFZG	Croatia
Swiss Foundation for Research in Social Sciences, Lausanne	FORS	Switzerland
University of Ljubljana, Social Science Data Archive, Ljubljana	UL-ADP	Slovenia

Table of Contents

1 Introduction

The aim of WP1 of the SEEDS project is to implement the various features of the data service establishment plans. This includes organisational, policy, and technical developments, all geared up toward preparing for “day one” of the new data services in partner countries.

The last activity of WP1 is the integration of the archiving system (chosen in D9 - Report on technical improvements) into the technical infrastructure of the partner institutions. Besides creating a set of policy documents for the data services (see D5 - Policy and procedures document) and new individual websites (see D11), it involves the development of a technical prototype that will allow for the basic archiving functions, following the OAIS model: ingest, preservation, and dissemination. Thus, as a key result of the SEEDS project, project partner have now chosen the tools to provide the capacity to take in new data, properly document, store and distribute these data, all according to international standards.

This deliverable describes the technical prototype and its related processes. The purpose is to provide the tools and processes that will allow the new data services to begin building their data collections, to structure their data and metadata in ways to allow for discovery and reuse, to store and secure data for the long-term, and to provide the conditions and platforms for data access for their future users. In sum, the prototype supplies a basic archiving infrastructure, with all needed hardware and software.

As has been the case in all previous project outputs, the intention was to put through the whole process of conducting and maintain as much commonalities as possible across new data service, especially for establishing technical platform. Common and compatible tools have been chosen among partners and determined as tools that will allow for future data and information sharing, as well as for synergies across the national services.

The Macedonian social science data archive (MK DASS) is national research infrastructure and public service which provides long time preservation and distribution of research data in the social sciences in the R. Macedonia. Our data service serves the research community, including researchers, teachers and students, as well as the broad public interested in social science research outputs. 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.

1.1 OAIS Model

The rapid growth of digital material in both volume and complexity, the rising expectations of archives’ users for access services, and the emerging digital preservation strategies, have all contributed to the definition of digital archive functions. The functionalities and procedures of a digital archive have been collected into the OAIS reference model, which became an ISO standard in 2003 (ISO 14721:2003). The OAIS provides both a functional model – the specific tasks performed by

the archive, such as storage or access – and a corresponding information model, which includes a model for the creation of metadata to support long-term maintenance and access (see figure 1-1).

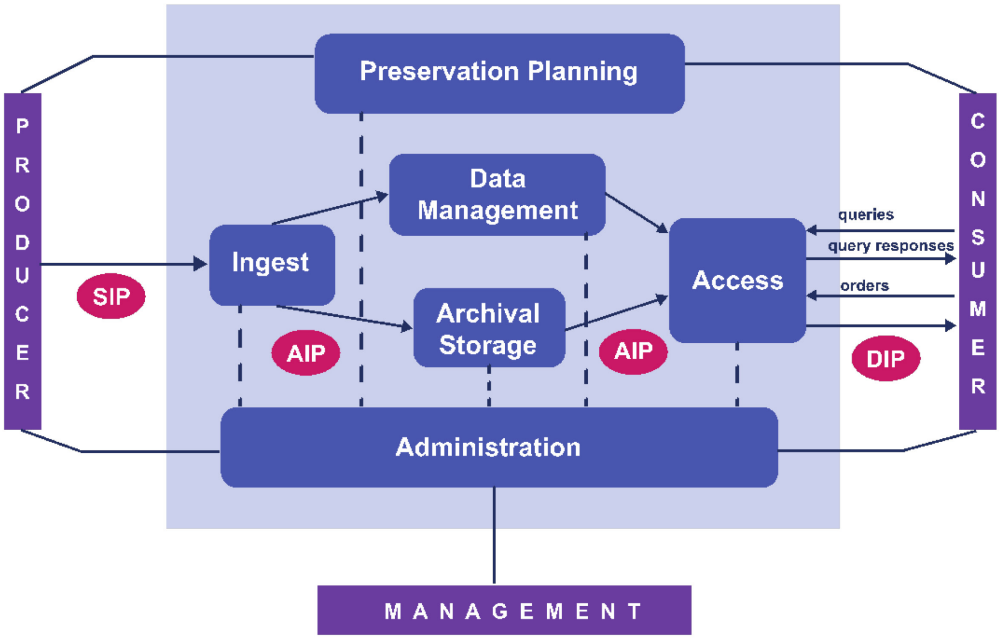

Figure 1-1: OAIS Functional Entities

The OAIS reference model is separated into six functional entities: Ingest, Data Management, Archival Storage, Preservation Planning, Administration, and Access. Outside the OAIS are the Producer (data producers, depositors, researchers), the Consumer (readers, researchers, academics, public, user community), and the Management (data managers, archivists, programmers, database managers, data centre managers). The data within the OAIS are represented as information packages (IPs). Each information package consists of metadata and physical files. There are three types of IPs: submission information package (SIP), archival information package (AIP), and dissemination information package (DIP).

2 Functional Specifications

2.1 Conceptual Model and Workflow

2.1.1 Ingest

Ingest provides the services and functions to accept SIPs from the Producer and prepare the content for Archival Storage and Data Management within the archive (see figure 2-2).

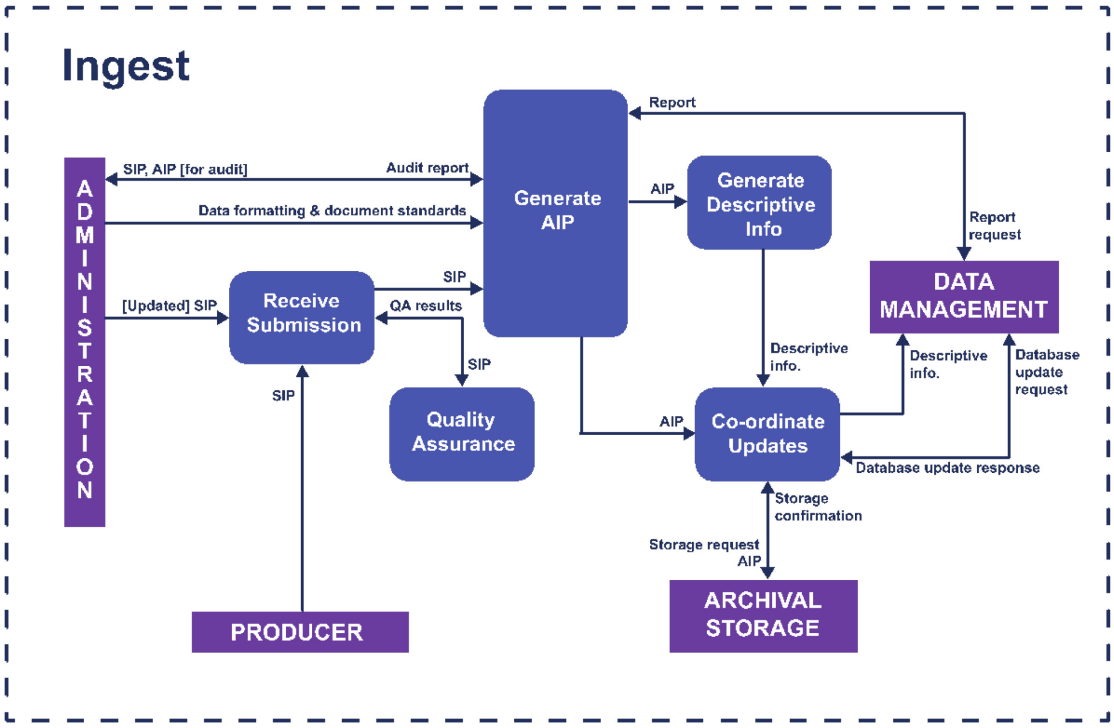

Figure 2-2: Functions of Ingest

Receive submission: The submission of the research data can be made via cloud or in person, on appropriate media (as defined in section 2.3 on Files and formats). In any case, the Macedonian Social Science Data Archive (MK DASS) should, in written through electronic communication or on paper, confirm the receipt of data and all accompanying materials (documentation, research instruments, publications, etc.). The precondition for receiving SIP will be a contract determining relations between the party submitting the data and the MK DASS.

Quality assurance: After the initial submission the MK DASS will have to clarify different issues in connection with the quality of the data sets. It will need to check if the data files and the data set are complete, if the scope of documentation is sufficient and in relation with the methods and instruments used in the research, it will have to compare the data set and data documentation and check the anonymization strategies. In the process of quality assurance other tasks should need to be carried out automatically by the system, like: checking for viruses, identifying file formats or generating unique identifiers. However, some more complex tasks, like checking if and how confidentiality of data is guaranteed and visual quality control will need to be done by the data experts at the MK DASS.

AIP will be generated as an end product of processing SIP at the Ingest and will be the base for DIP. In the process of generating AIP, on the side of data processing, two activities will be done:

rearranging of data and decisions about description level. On the side of documentation processing, the activities will be: reorganization of documentation, marking of material, extraction of descriptive metadata, introducing of the persistent identifier.

When processing data and documentation, additional activities will include: extraction of metadata (technical, administrational, structural), and preservation, conversion of files to archival format and separate AIP storage from SIP.

Metadata should provide all necessary information on the study level. MK DASS will follow metadata specifications in accordance with Data Documentation Initiative (DDI) metadata specification, version 2.5 or higher.

The following types of metadata will be created:

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;[1]

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;[2]

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.[3]

Deposit contract will be signed between MK DASS and the party that submits the data. The contract will regulate the mutual rights and obligations between the parties. Main issues handled in the contract will be: copyrights and/or author rights based on the ownership of the data, access conditions and any type of restrictions.

2.1.2 Archival Storage

Archival Storage provides the services and functions for the storage maintenance and retrieval of AIPs (see figure 2-3). Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive

[1] Dublin Core Metadata Initiative: http://dublincore.org/
[2] Metadata Encoding and Transmission Standard (METS) http://www.loc.gov/standards/mets/
[3] PREMIS Data Dictionary for Preservation Metadata: http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

holdings are stored, migrating files into the archival formats, performing routine and special error checking and providing disaster recovery capabilities.

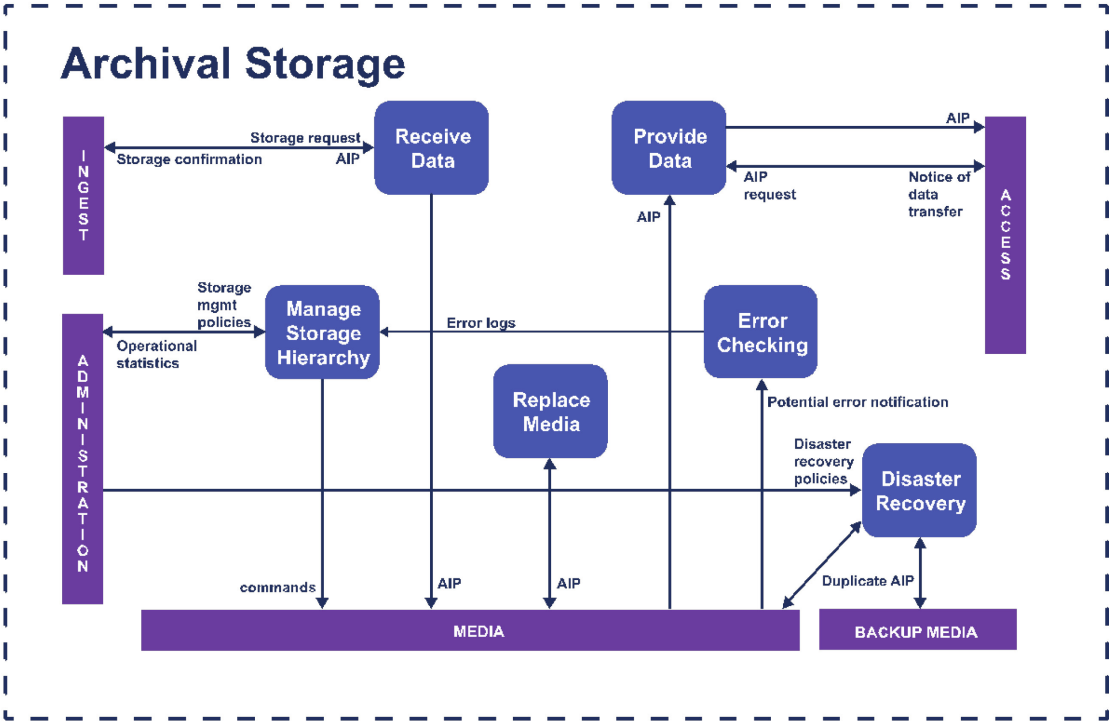

Figure 2-3: Functions of the Archival Storage

The main goal of the MK DASS is to ensure long term accessibility of the digital materials preserved in the Archive, ensuring the highest level of authenticity of formats. Media replacement, that is transferring of data into new data formats, will be done in accordance with the technological cycles of the data files. Regular error checking will be part of the operating activities of the MK DASS. The Archive will manage daily on-site and off-site backups of stored data. Backups of sensitive data hosted locally will be stored in two different locations. Disaster recovery plan will secure the archive holdings in a safe environment and on the long run.

2.1.3 Data Management

Data Management provides the services and functions for populating, maintaining and accessing both metadata, which identify and document repository holdings, and administrative data, used to manage the repository (see figure 2-4).

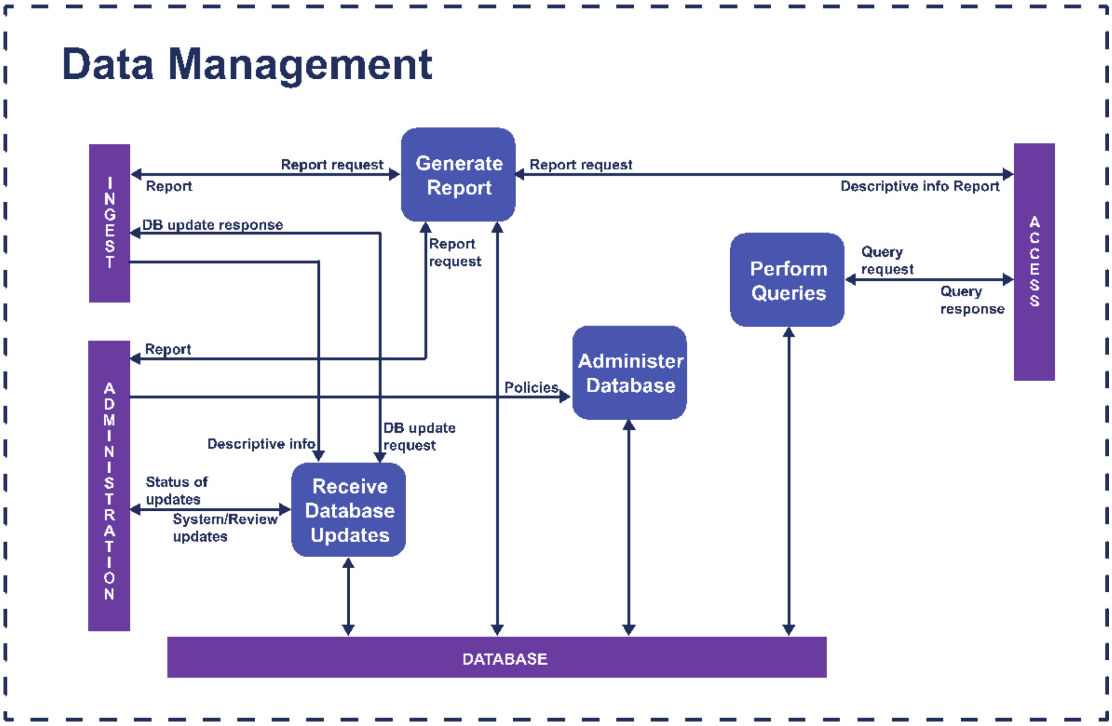

Figure 2-4: Functions of Data Management

The Data Management activities in the MK DASS will involve the maintenance of all the archived information and files, including the management of metadata, communication with the producers and users, and access statistics. The primary functions of Data Management will be carried out by the archive's 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).

2.1.4 Administration

Administration provides the services and functions for the overall operation of the archive system (see figure 2-5). Administration functions include soliciting and negotiating submission agreements with the Producer, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software. It is also responsible for establishing and maintaining archive standards and policies, providing user support, and activating stored requests.

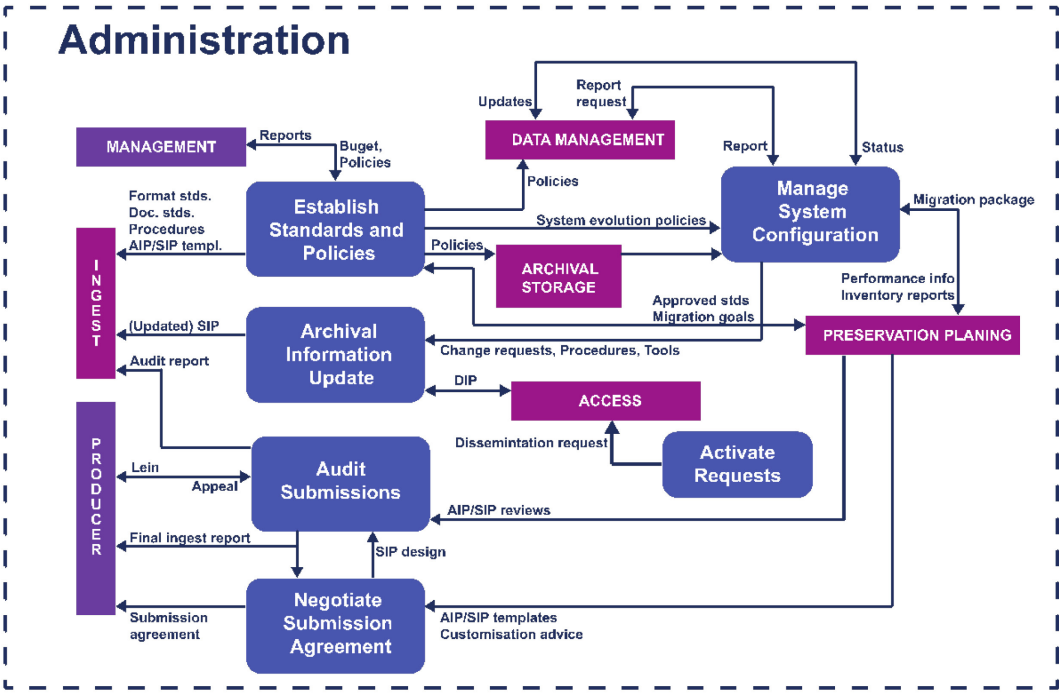

Figure 2-5: Functions of Administration

Submission agreement negotiations at MK DASS will be done by a data specialist who will be in communication with data depositors. Following these negotiations, the final agreement has to be approved by the Data manager. All legal issues will be resolved with support of the legal specialists from the hosting institution. Standards and policies will be clearly defined in appropriate documents, which will be publicly available (for example on the MK DASS web site), with clear rules, and also guidelines for future data depositors and users. Following these guidelines, researchers will be able to prepare their submission materials for the archive.

2.1.5 Preservation Planning

Preservation Planning provides the services and functions for monitoring the environment of the archive and making recommendations to ensure that the information stored in the archive remain accessible over a long-term, even if the original computing environment becomes obsolete (see figure 2-6). Preservation planning functions include evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for archive standards and policies, and monitoring changes in the technology environment and in the user's service requirements. Preservation Planning also develops detailed migration plans, software prototypes, and test plans to enable implementation of Administration migration goals.

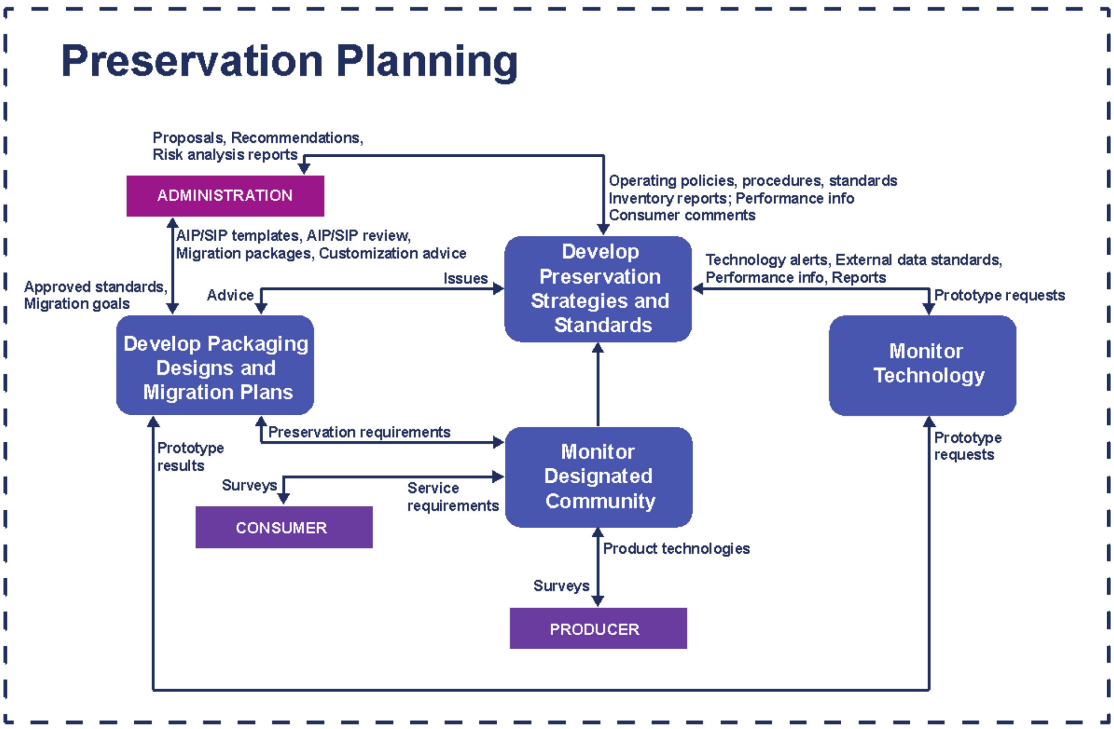

Figure 2-6: Functions of Preservation Planning

The Preservation Planning in MK DASS will adhere to the internationally accepted principles, standards and values, especially to the preservation practices outlined by the OAIS reference model, with the aim to provide long term maintenance and access to digital materials. To fulfil this aim, the MK DASS will have to ensure the following standards: (1) data are usable and accessible – the data can be located, retrieved, and presented; (2) data are reliable – the contents of data can be trusted as a full and accurate representation of the transactions, activities, or facts to which they attest, and can be depended upon in the course of subsequent transactions or activities; (3) data are authentic – the record of data can be proven to be what it purports to be, to have been created or sent by the person purported to have created it, and to have been created at the time purported; and (4) integrity and quality of data - to prove that a record is complete and unaltered. In doing this, the MK DASS should become a trustworthy organisation – to show the potential for long-term archive for the research community and also to act transparent - to demonstrate to the designed community and also to the funders that the MK DASS is an organisation with a long-term commitment of data preservation.

The tasks of Preservation planning will be: (1) development of preservation strategies and standards, (2) development of packaging designs and migration plans, and (3) monitoring of technology (innovations in storage and access technologies) and the designated community (shifts in scope or expectations). In addition to this, MK DASS will monitor the technical fitness of its archive, will do 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 reports will be regularly

gathered within the preservation activities of the MK DASS.

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 or unreadable at one time or another. 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, as well as all subsequently generated manifestations of the original files. In this case, we will adhere to the principle of reversibility: being able to revert to an earlier version of a digital file after migration. We will also fully document the migration process in the form of a detailed migration history as part of the metadata associated with the data file.

2.1.6 Access

Access provides the services and functions that support Consumers in determining the existence, description, location, and availability of information stored in the archive, and in allowing them to request and receive data (see figure 2-7). Access functions include communicating with Consumers to receive requests and applying controls to limit access to specially protected information. This includes coordinating the execution of requests until its successful completion, generating responses, and delivering the responses to Consumers.

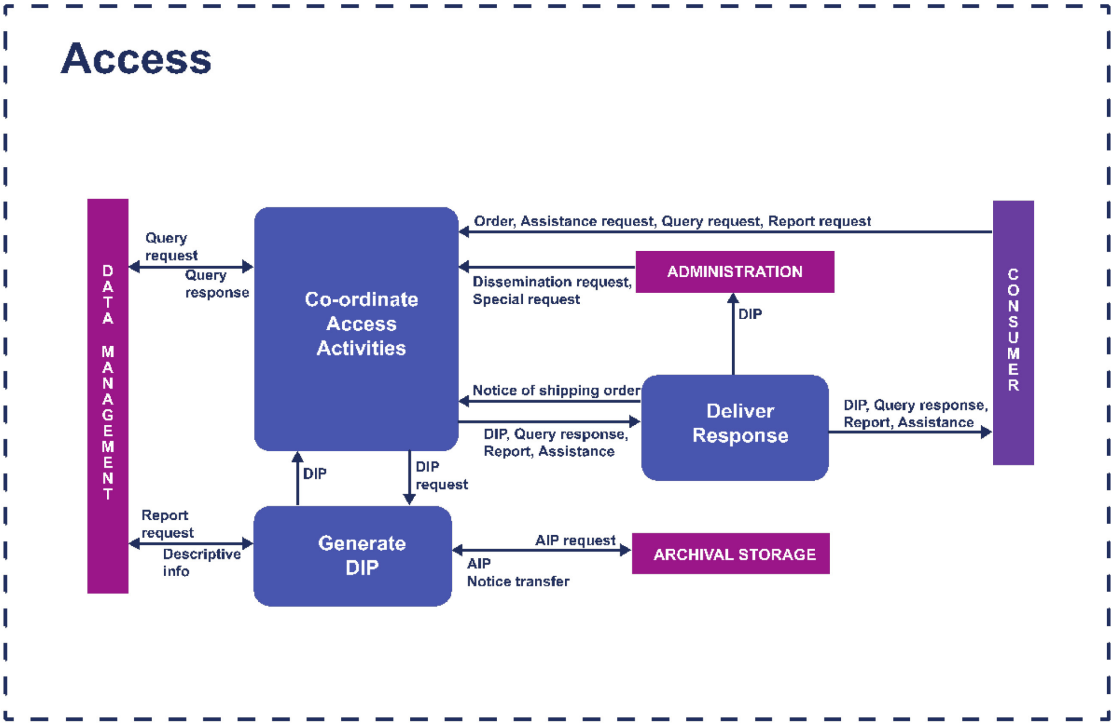

Figure 2-7: Functions of Access

The MK DASS currently uses the SEEDSBase platform for data archiving, discovery and dissemination. The platform is based on FORSBase software developed by the Swiss Centre for Expertise in the Social Sciences (FORS) at the University of Lausanne, Switzerland, and it supports policies and procedures defined by the Data service. SEEDSBase is hosted at FORS infrastructure and maintained by their staff. However, access to datasets will be managed by the MK DASS staff and individual dataset producers.

Datasets and documentation held by the MK DASS are available within certain limitations and controlled access depending on the sensitivity of data and the specific conditions determined by the depositor. The latter are set out in the User agreement, a version of which all users must accept before being given access to any dataset. The End User Licence describes the legal framework under which the material is distributed.

There are three levels of access:

Research data available without any restrictions (open access)

Research data available only to registered users

Research data available only to registered users with special permissions given by the Archive staff or depositor

Distribution of research data based on defined access rights is done:

Remotely via SSL secured web service

Local access – user has a physical access to dataset in a "safe room"

Registration of users:

Users will be able to send a Registration request via contact form on the MK DASS web site. Registered user needs to accept the "User agreement" and, if needed, a depositor's special conditions agreement for a particular dataset. In special cases there is also a possibility for requesting physical signature of the "User agreement" and/or the depositor's special conditions.

2.2 Metadata Specifications

Metadata of a study will be described in Data Documentation Initiative (DDI) metadata specification.

Due to the fact that MK DASS will use the SEEDSbase platform we will adhere to the DDI version 3.2. that FORS / FORSbase will use from December 2017.

Complete documentation on DDI is available on the DDI alliance web page[4].

The DDI is designed to be fully machine-readable and machine processable. It is defined in XML,

[4] http://www.ddialliance.org/Specification/DDI-Codebook/2.1

which facilitates easy Internet access. DDI Controlled Vocabularies[5] , CESSDA Controlled Vocabularies and CESSDA topic classification are planned to be used. The use of a predefined topic classification will make possible future inclusion in the CESSDA data catalogue[6] easier, since these are the topics that enable browsing in the catalogue.

The fields for ingest in the archive ingest tool are made using the CESSDA-recommended fields[7], relevant for study description.

2.3. Files and File Formats

There are several reasons why a data archive should be concerned with file formats: they exist in big numbers, are relevant during the whole workflow of the OAIS reference model, and are largely proprietary. File formats are subject to rapid obsolescence if they are not evaluated according to crucial criteria, such as open standards, ubiquity, interoperability, and metadata support. Therefore, file formats that are well-documented, non-proprietary and usable on different hardware and software platforms are much less at risk of not being usable anymore in the future. In addition, their frequency of migration and their costs of preservation are lower.

File formats are an important issue during the entire workflow of the archive (see chapter 2.1). In the functional entity Preservation Planning, the composition and attributes of the information package are defined. This includes the selection of file formats for the SIP, the AIP and the DIP. The decisions of the archive on which file formats are acceptable as archival and distribution formats are linked to the significant properties of the files (what aspects of the digital material we want to preserve). That is why it is important that file formats are controlled and validated, according with the available specific tools, already in the Ingest phase.

There are a number of tools on the market for migrating a file format into a more reliable and sustainable file format:

- *Native Java Image library* for most image formats;
- *Imagemagick* for most image formats, esp. Raster;
- *FFMPEG* for various AV formats;
- *readpst* for email;
- *Ghostscript* for PDF;
- *Libre Office* for Office Open XML, and word processor files – also shifts various office formats to PDF and PDF/A;
- *Inkscape* for Vector images.

When selecting target formats, the following criteria should be considered:

- Ubiquity;

[5] http://www.ddialliance.org/controlled-vocabularies

[6] http://www.cessda.net/catalogue/

[7] https://cessda.net/content/download/709/6350/file/CESSDA%20mandatory%20and%20recommended%20metadata%20fields.pdf

- Support;
- Disclosure;
- Documentation quality;
- Stability;
- Ease of identification and validation;
- Intellectual Property Rights;
- Metadata Support;
- Complexity;
- Interoperability;
- Viability;
- Re-usability.

The selected file formats represent a summary of different recommendations from CESSDA partners and internationally recognised institutions: [8]

File formats considered as appropriate for SIPs:

- 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 OpenOffice 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.

File formats considered as appropriate for the AIP:

- Tabular data: Microsoft Excel File Format (XLS) (.xls), ASCII, Comma Separated Values (CSV) (.csv; .txt);
- Text: Adobe Portable Document Format (PDF/A) (.pdf), XML (.xml), Standard Generalised Markup Language (SGML) (.sgml);
- Audio: Waveform Audio File Format (.wav);

[8] The formats highlighted in bold are preferred over the others of the same category.
FORS: Qualitative Data Archiving at FORS – Policy and Procedures: http://www2.unil.ch/daris/IMG/pdf/Donnees_qualitatives_archivees_chez_FORS_-_Politique_et_Procedures.pdf,
UK Data Archive: Formats table: http://www.data-archive.ac.uk/create-manage/format/formats-table, UK Data Archive: Assessment of UKDA and TNA Compliance with OAIS and METS Standards, p. 89 http://www.dptp.org/wp-content/uploads/2010/08/UKDAp90.pdf.

- Raster (bitmap) images: TIFF (.tif);
- Vector images: DFX (.dfx), SVG (.svg);
- Video: MPEG-2 (.mpg2).

File formats considered as appropriate for the DIP:

- Tabular data: SPSS portable format (.por), SPSS (.sav), Stata (.dta), R (.txt);
- Text: Adobe Portable Document Format (PDF) (.pdf), Rich Text Format (RTF) (.rtf);
- Audio: MP3 (.mp3);
- Raster (bitmap) images: JPEG (.jpg)
- Vector images: DFX (.dfx), SVG (.svg);
- Video: MPEG-4 (.mpg4).

In addition, file format registries are a way of helping to identify file formats and looking up format specifications.

3 Technical Specifications

3.1 Tools

3.1.1 SEEDSbase

MK DASS will use the SEEDSbase to cover all four segments of the OAIS model.

3.2 Communication

3.2.1 General Communication

According to the OAIS model there are several different possibilities for how the data archive can communicate with the actors, that is the data producers and consumers. More precisely, it is the functional entities Preservation Planning and Administration that are responsible for the communication task. They include for instance the development of preservation strategies and standards of monitoring the community and technology in order to meet the needs of the producers and consumers (see figure 3-1).

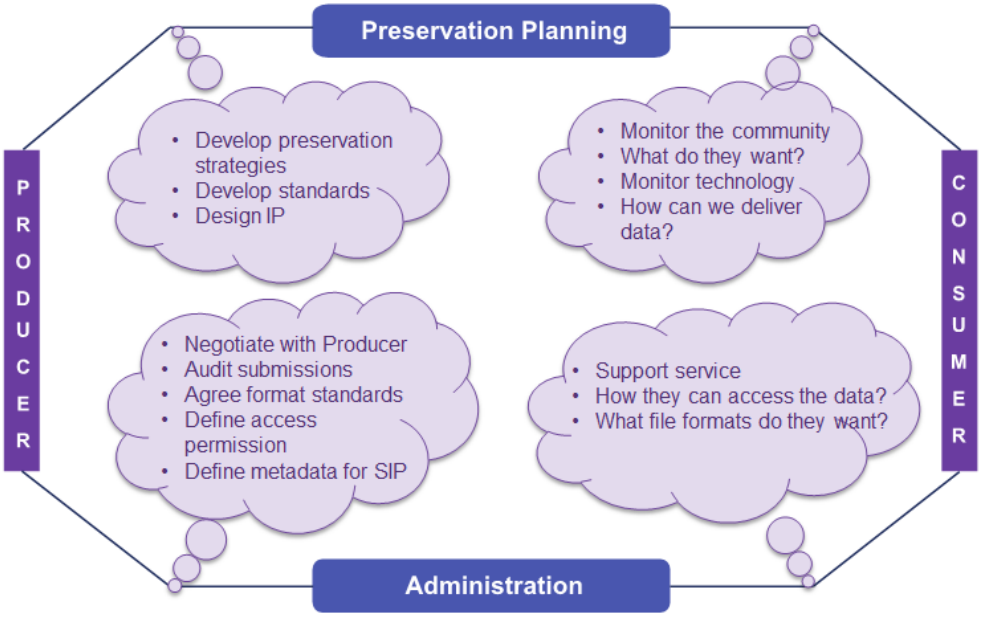

Figure 3-1: Communication

3.2.1.1 Website

The most common and wide-reaching channel to communicate with the community is by means of an institutional website. It is the showcase for interested consumers and producers of data to learn about how data can be obtained and submitted. It is the platform where the policy and procedures, reports and publications, guidelines for data preparation, description of data protection, and other training materials are made available.

The website of the MK DASS can be visited on:

https://mk.seedsproject.ffzg.hr/

3.2.1.2 Mailing Lists

A mailing list of potential users is established and will be regularly updated by the MK DASS in order to inform Producers and Consumers about the latest news and upcoming events, such as training and workshops.

3.2.1.3 Direct Contact

A third way of communication is direct contact of the Producers and the Consumers or potential users of the data service through sporadic interaction on an as-needed basis (e.g., for workshops, seminars, and conferences).

3.2.2 Specific Communication

All the specific communications with the users during user registration and Ingest will be recorded and maintained via the SEEDSbase but also via e-mail communication and direct meetings with data depositors (and possibly users). There are plans to decide on a project tracking software where all communication with depositors and other users will be stored.

3.3 Technical Infrastructure

Since MK DASS has not yet established its entire necessary technical infrastructure, the following chapter presents only an example of how a server architecture could be specified. In the future, MK DASS will determine in detail all the technical specification needed to fulfil its submission, archival administration and access policies.

3.3.1 Server Architecture (an example)

For the implementation of the SEEDS project 2 servers are needed:

- Virtual server 1 (currently in Croatia);
- Virtual server 2 (should be in each partner country);

The Virtual server 1 is used for the hosting of each national web portal. A single WordPress application with 6 website instances (one for each partner) is installed for the national web portals (see D11).

Here is the detailed Virtual server 1 configuration:

Configuration: 2 vCPU, 2GB RAM, 10GB HDD

OS: Debian GNU Linux 8.2 (Jessie)

HA: Ganneti cluster[9]

The Virtual server 2 should be used for the national catalogues and Ingest/Archival platforms.

Here is the detailed Virtual server 2 configuration example:

Configuration: 2 vCPU, 12.0 GB , 100.0 GB

OS: Debian GNU/Linux 8.2 (jessie)

[9] http://www.ganeti.org/

HA: Not Enabled

Both Virtual server 1 and Virtual server 2 should have redundant IT infrastructure, monitoring, and backup. In addition to local (on-site) file and database backup, there should be a daily automatic offsite backup solution as well. The local servers should be used for backup purposes.

The usage of virtual machines is valuable for prototype implementation and testing, but for the production system, the newly established archives should have more granular distribution of services. Sensitivity of data in the various components of the OAIS, requires us to think about different security levels of data and preservation requirements. To achieve this goal, the future architecture will be installed separately on different virtual machines, based on different platform deployment stacks:

- Centre's website
- Virtual research environment and self-archiving tool (SEEDSbase)
- Long-term preservation archival system (SEEDSbase)

Each of these components have different deployment requirements (database, web server, runtime language stack), so it makes sense to separate components on different VMs to enable easy maintenance (migration when changing components, deploying different components for new archives in the future, firmware upgrades).

Looking at the current state of development and support probability of chosen software of the established data archives, it seems that a future change in the components will be probable. This is one of the reasons why the easy maintainability of the system is important. The staff of the archive needs to be capable of testing other available software tools (in a state accessible to them), preferably under Free/Libre/Open Source licences, by using the process described in deliverable D9-Report on technical improvements.

Since each application is installed on a separate virtual machine (and each might have its own set of issues/bugs), security issues are addressed for each virtual machine individually. This means for example that in case of security problems on the web portal, there will be no effect on the security of the archival copy of the data or any other component of the archival infrastructure.

All virtual machines should have two copies stored on different physical machines locally. Machines should be located in different buildings to ensure continuous operation in case of environmental problems in one of the buildings (fire, flooding etc.).

During the process of developing an OAIS based data archive, two distinct types of data required for keeping in the archive were identified - SIP and AIP, which require long-term preservation together with an audit log. This also requires the ability to check whether data is correctly stored on the media that requires checksums on the level of the file system (scrubbing). For this requirement, ZFS[10] storage and snapshots using LVM could be implemented to provide a long-term archival copy of

[10] http://bit.ly/dc14-zfs

current prototype on different locations (e.g. in faculty building), which should be updated daily (from computing centre location). This would enable disaster recovery in case of one location failure. It is also possible to have multiple remote copies, if needed.

The management of applications and data could be done using Ganeti[11], an open source cloud solution that enables high availability for virtual machines and provides data storage requirements outlined above.

3.3.2 Network and Telecommunications

The network infrastructure and telecommunications are accessed using the host organisations' systems. In the case of the MK DASS, this is the infrastructure of the Ss. Cyril and Methodius University in Skopje's.

3.3.3 Hardware and Software for production systems

Based on best practices and international standards for social science data archives, the data services have determined the hardware and software they will use.

Workstation computers that will be used by future archive staff for Data Management should include the following software: office tools; conversion tools; software for statistical analysis (STATA, R, SPSS); tools for preparing metadata description of a study, etc.

If the archive wants to use a proprietary product, they will have to buy a licence or use the existing licences of their hosting institution, if available.

[11] http://bit.ly/dc14-ganeti

4 Conclusions and Future Development

In conclusion, the prototype described in this paper provides the technical basis for all key archiving functions, following the OAIS model. The new data services will be able on “day one” to integrate and manage new datasets, safely store and protect data, as well as disseminate data and documentation to users. Their technical systems will function according to international norms and best practices, even if some of the archiving workflow will need to be handled manually.

It should be noted, however, that while the prototype will enable certain basic services, it will not be as comprehensive or as flexible as the one used by mature social science data archives. Future work should expand the technical development to accommodate for a greater volume and variety of data, to automate more everyday practices, and to enhance communication potential and exchange with data producers and users. This work will continue for many years, and will build on experience, further training, and funding. Like any others, these new data services will have to adapt technically to the ever-changing research and policy environments.