Support for Establishment of National/Regional Social Sciences Data Archives

D5.3—Report on prototype database

Deliverable Lead:	FORS
Related Work package:	WP5
Author(s):	Bojana Tasić (FORS) Irena Vipavc Brvar (UL) Renate Kunz (FORS) Brian Kleiner (FORS) Marijana Glavica (FFZG) Dobrica Pavlinušić (FFZG) Alen Vodopijevec (Rudjer Boskovic Institute)
Dissemination level:	Public
Submission date:	30th June 2014
Project Acronym:	SERSCIDA
Website:	http://www.serscida.eu
Call:	FP7-SCIENCE-IN-SOCIETY-2011-1
Project Number	288985
Start date of Project:	01 January 2012
Duration:	30 months

Acknowledgements

This report has been developed within the project “Support for Establishment of National/Regional Social Sciences Data Archives” (SERSCIDA).
The SERSCIDA participants are:

Participant organisation name	Short Name	Country
University of Sarajevo, Human Rights Centre	HRC Sarajevo	BA
Swiss Foundation for Research in Social Sciences	FORS	CH
University of Gothenburg, Swedish National Data Service	UGOT	SE
University of Essex, United Kingdom Data Archive	UESSEX	UK
University of Ljubljana, Social Science Data Archive	UL	SI
Institute of Economic Sciences, Belgrade	IEN	RS
University of Zagreb, Faculty of Humanities and Social Sciences	FFZG	HR

Table of Contents

ACKNOWLEDGEMENTS **2**

TABLE OF CONTENTS **3**

FIGURES **5**

1 INTRODUCTION **6**

1.1 Work Package 5 (D5.1, D5.2, D5.3) **6**

1.2 Purpose of Prototype **6**

1.3 OAIS Model **6**

FUNCTIONAL SPECIFICATIONS **8**

2.1 Conceptual Model and Workflow **8**
2.1.1 Ingest 8
2.1.2 Archival Storage 10
2.1.3 Data Management 11
2.1.4 Administration 12
2.1.5 Preservation Planning 14
2.1.6 Access 16

2.3 Metadata Specifications **17**

2.4 Files and File Formats **17**

3 TECHNICAL SPECIFICATIONS **20**

3.1 Technology Considerations **20**

3.2 Tools **20**
3.2.1 EPrints 20
3.2.2 Request Tracker 21
3.2.3 Nesstar Publisher 21
3.2.4 Nesstar Server 22
3.2.5 Nesstar WebView 22
3.2.6 CESSDA Catalogue 23

3.3 Users **24**
3.3.1 Database of Users 24
3.3.2 Login, User Roles and Permissions 25

3.4 Ingest **26**

3.5 Data Management and Archival Storage **27**

3.6 Access **27**

3.7 Communication **28**
3.7.1 General Communication 28
3.7.2 Specific Communication 29

3.8 Technical Infrastructure **29**
3.8.1 Server Architecture for the prototype 29
3.8.2 Network and Telecommunications 30
3.8.3 Hardware and Software for production systems 30
3.8.4 Server architecture plan for production systems 31

4 USE CASES AND SCREENSHOTS **33**

4.1 Receiving data from the depositor (SIP), creating AIP and DIP in EPrints **33**
4.1.1 Receive data 33
4.1.2 Review submitted data and create SIP 35
4.1.3 Create AIP 36
4.1.4 Create DIP 36

4.2 User registration (WP), creating user accounts (EPrints and Nesstar), and communication with users (RT) **37**
4.2.1 User registers using the registration form in WP 37
4.2.2 WP registration form sends email to RT and creates RT ticket (automatic) 38
4.2.3 RT sends confirmation e-mail to the user and notification to staff (automatic) 39
4.2.4 User receives confirmation message from RT and archive staff gets notification by email about ticket creation (automatic) 39
4.2.5 Staff member logs to the RT and take the ticket - become owner 39
4.2.6 Staff member (owner) replies to the user using RT - sends user agreement to the user 40
4.2.7 The user (requestor) receives reply as email message (automatic) 41
4.2.8 User sends signed agreement 41
4.2.9 RT receives users email and adds it to the appropriate ticket (automatic) 41
4.2.10 All staff is notified about new correspondence (automatic) 41
4.2.11 Staff member checks if the user have right to access data 41
4.2.12 Management creates user in EPrints and Nesstar 42
4.2.13 Management sends username and password to the user via RT 43

4.3 Creating metadata and publishing data to the web catalog using Nesstar Publisher **43**

4.4 Accessing data in Nesstar catalog **46**

5 CONCLUSIONS AND FUTURE DEVELOPMENT **50**

6 ANNEX **51**

6.1 Criteria for choosing archiving tool **51**

6.2 SERSCIDA WP4 - Standards and Tools Table **52**

6.3 End user Licences **57**

6.4 Hardware **60**
6.4.1 CRODAS 60
6.4.2 SR-SSDC 60

6.5 Software **61**
6.5.1 List of software used at SND 61
6.5.2 List of software used at ADP 61

6.5.3 List of software used at FORS....62

7 REFERENCES....63

Figures

Figure 1: OAIS Functional Entities 1
Figure 2: Functions of Ingest 8
Figure 3: Functions of Archival Storage 10
Figure 4: Functions of Data Management 11
Figure 5: Functions of Administration 13
Figure 6: Functions of Preservation Planning 15
Figure 7: Functions of Access 16
Figure 8: Communication 28
Figure 9: Server Architecture 29
Figure 10: Server Architecture plan for production 32
Figure 11: EPrints 33
Figure 12: EPrints 34
Figure 13: EPrints 34
Figure 14: EPrints 35
Figure 15: EPrints 36
Figure 16: EPrints 37
Figure 17: User registration form 38
Figure 18: RT 39
Figure 19: RT 40
Figure 20: Create user in EPrints 42
Figure 21: Create user in Nesstar WebView 43
Figure 22: Nesstar Publisher 44
Figure 23: Nesstar Publisher 45
Figure 24: Nesstar Publisher 46
Figure 25: Nesstar WebView 47
Figure 26: Nesstar WebView 48
Figure 27: Nesstar WebView 49

1 Introduction

1.1 Work Package 5 (D5.1, D5.2, D5.3)

The aim of WP5 of the SERSCIDA project was to implement the various features of the data service establishment plans, designed as part of WP4 (described in deliverable D4.1). This included organisational, policy, and technical developments, all geared toward preparing for "day one" of the new data services in Croatia, Serbia, and Bosnia and Herzegovina. WP5 brought to bear all that was learned and conceived during the first four WPs, and the implementation at various levels reflects this.

Besides creating a set of policy documents for the data services (see SERSCIDA D5.1) and new individual websites (see SERSCIDA D5.2), WP5 involved 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 SERSCIDA, the data services now have the tools and capacity to take in social science data, and then to properly document, store, and distribute these data, all according to international standards. This final deliverable of WP5 describes this technical prototype and related processes.

1.2 Purpose of Prototype

The technical prototype developed for each new data service as part of WP5 follows the OAIS model (described below), and is compliant with best practice and international standards for social science data archives. The purpose is to provide the tools and processes that will allow the data services to begin building their data collections, to structure data and metadata in ways that allow for discovery and reuse of these, to store and secure data for the long-term, and to provide the conditions and platforms for accessing the data by users. In sum, the prototype supplies a basic archiving infrastructure, with all needed hardware and software.

As has been the case for all previous SERSCIDA project outputs, the intention was to maintain as much commonality as possible across the three data services, and this is especially true for the technical platform. Common and interoperable tools will allow for future data and information sharing, as well as synergies across the regional services.

1.3 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).

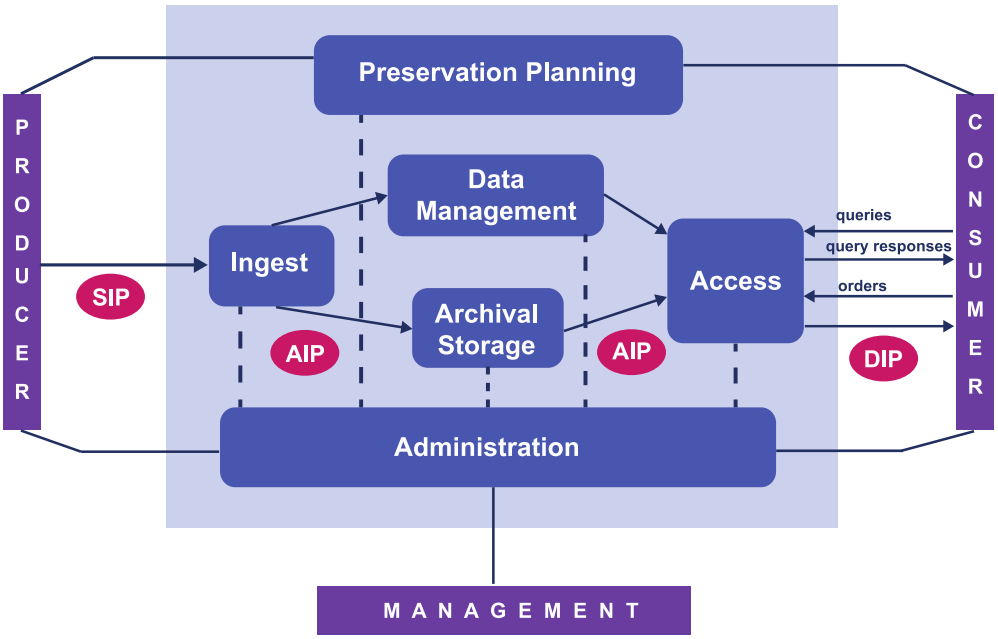

Figure 1: OAIS Functional Entities

The SERSCIDA Prototype Database (SERSCIDA-PDB) is fully OAIS-compliant; this means that its implementation supports the OAIS functional and information model and fulfils the OAIS responsibilities listed in chapter 2.1 of the SERSCIDA D5.1. During the implementation of the SERSCIDA-PDB, the OAIS reference model was used as a guide to identify and organise services and contents.

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). (For more information about the OAIS reference model, see SERSCIDA D5.1 chapter 2.1).

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 contents for Archival Storage and Data Management within the archive (see figure 2). For the SERSCIDA-PDB, the process of Ingest is composed of several different functions that include receiving SIPs, performing quality assurance on SIPs, generating an AIP that complies with the archive's data formatting and documentation standards, extracting Descriptive Information from the AIPs for inclusion in the discovery catalogue, and coordinating updates to Archival Storage and Data Management.

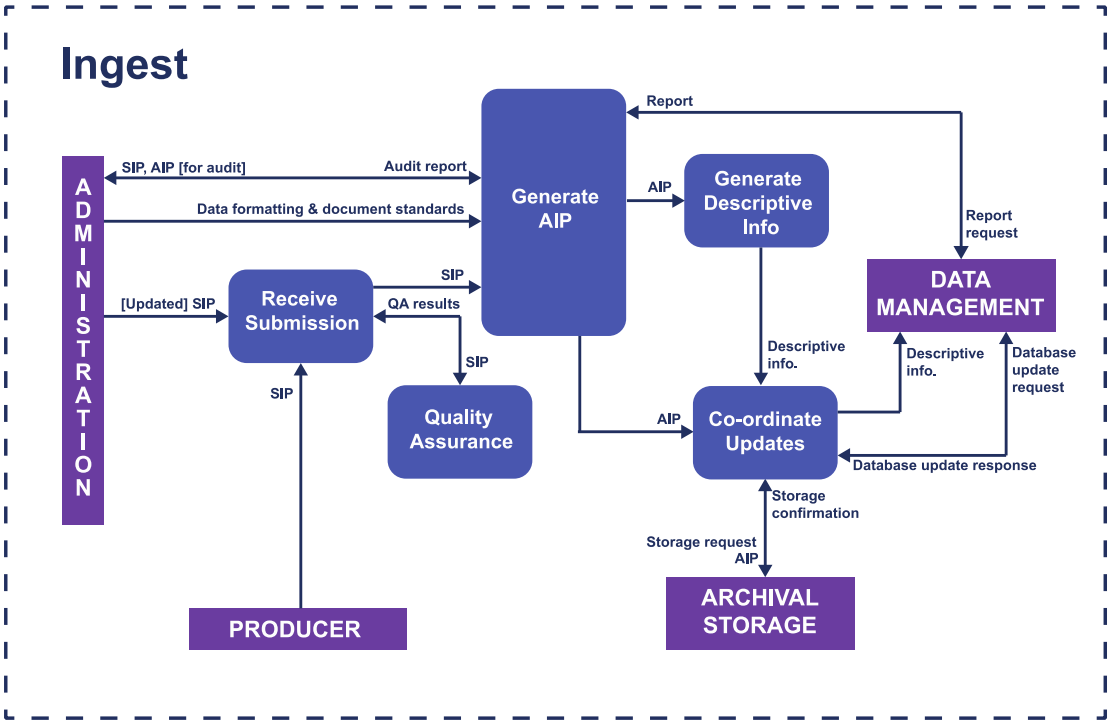

Figure 2: Functions of Ingest

2.1.1.1 Receive Submission (SIP)

The Receive Submission function provides the appropriate storage capability and tools to receive a SIP from the Producer (or from Management via the Administration function).

The SERSCIDA-PDB receives SIPs via an ingest tool (see chapter 3.4). If the SIP is submitted to the archive by email or from another device, the Management needs to enter the SIP into the ingest tool. The archive staff will proactively work with the Producer to agree on the content, quality, and format of the digital material. Also, the archive staff will specify preferred formats for the receipt of digital material (see chapter 2.4). Once material has arrived, it must undergo several reviews, including virus checking, format compliance and anticipated content and file type. The review and approval step will also include approval of the submitted content. Some metadata checks and data conversion may be necessary to allow initial materials to be accepted. The SERSCIDA-PDB uses a ticketing system (see chapter 3.2.2) to record all actions and decisions made concerning the submitted material, as well as all communication with

users concerning the submitted material. When the submitted material is approved, it will be prepared for ingestion into the repository. This procedure will include the generation of the AIP, which ensures that the material conforms to the repository data formatting and documentation standards (see SERSCIDA D5.1 chapter 3.2 and 4.1). Metadata will be generated, and the components of the AIP will be stored in the system.

2.1.1.2 Quality Assurance

The Quality Assurance function validates the successful transfer of the SIP to the system. It includes virus scan, checksums associated with each data file, and use the system log files to record and identify any file transfer or media read/write errors.
To ensure quality, the SERSCIDA-PDB implements the following functionalities:

- Virus checking on SIP;
- Automatic validation of the successful transfer of the SIP into the ingest tool;
- Checksums associated with each data file;
- Generation of statistical and error reports; and
- Audit trails of all actions.

In addition, the prototype will allow the archive's staff to perform manual/visual quality control assurance, using a user-friendly GUI interface of the ingest tool.

2.1.1.3 Generate Archival Information Package (AIP)

The Generate AIP function transforms the SIP into an AIP that conforms to the archive's data formatting and documentation standards. This may involve file format conversions or reorganization of the content information in the SIP. Generating the AIP deals with transforming or repackaging of data to the current standards of the archive (see SERSCIDA D5.1 chapter 4.1).

2.1.1.4 Generate Descriptive Information/Metadata

The Generate Descriptive Information function extracts Descriptive Information from the AIPs and collects Descriptive Information from other sources. This includes metadata to support searching and retrieving data. Archives will need several tools to prepare data. Statistical software is necessary to check the content of the file, clean, and anonymise it if needed. SPSS, R or STATA will most likely be used. In this phase the SERSCIDA-PDB will also collect the metadata needed for the access catalogue (see chapter 3.6).

2.1.1.5 Coordinate Updates

The SERSCIDA-PDB stores AIPs in the same location in the repository, but once the archive is established and fully operational it may be the case that AIPs are stored in different locations within the repository. The Coordinate Updates function will transfer components of the AIP to the relevant storage locations, and will maintain a record of the location of the components.

All the updates of the AIP in the SERSCIDA-PDB are implemented using the Data Management tool (see chapter 3.5)

2.1.2 Archival Storage

Archival Storage provides the services and functions for the storage maintenance and retrieval of AIPs (see figure 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 holdings are stored, migrating files into the archival formats, performing routine and special error checking and providing disaster recovery capabilities. The Archival Storage of the SERSCIDA-PDB is on the same server as the Ingest. Once the system is in production, the dedicated server will be prepared for this purpose.

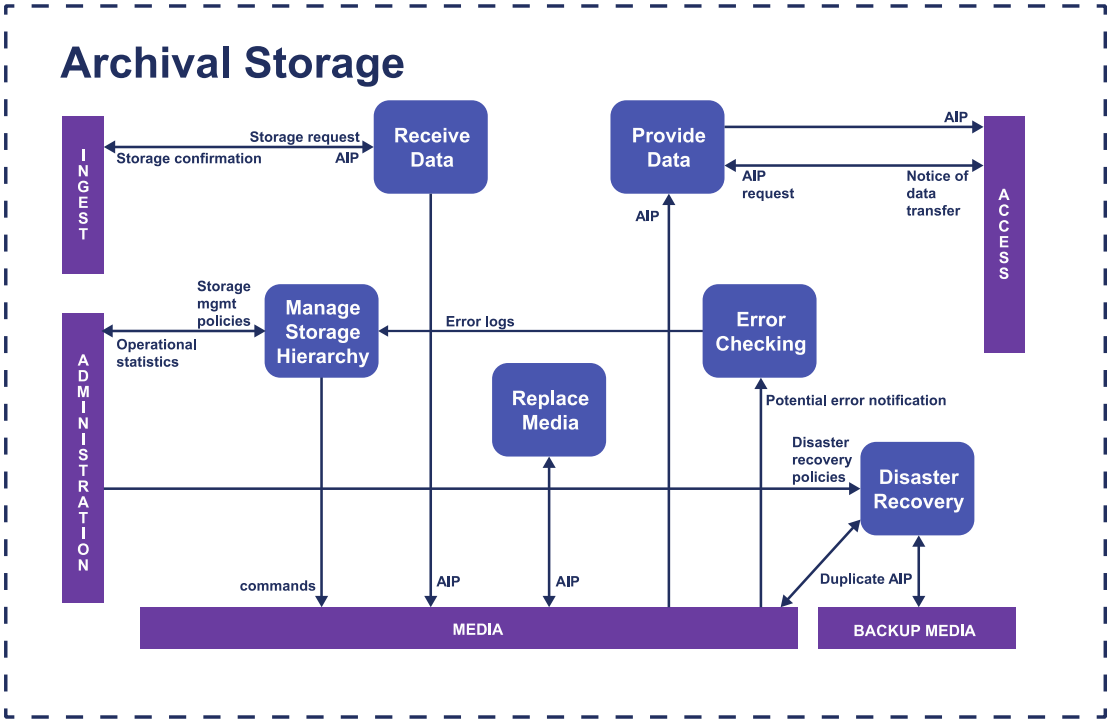

Figure 3: Functions of Archival Storage

2.1.2.1 Receive Data

Since Ingest and Archival storage are in the same system, and the AIP is not physically moved, the SERSCIDA-PDB doesn't implement any specific functionality for the Receive Data function.

Once the proper Archive Server is in place, the Receive Data function will receive a storage request and an AIP from Ingest and it will move the AIP to permanent storage within the archive. Upon completion of the transfer, this function will send a storage confirmation message to Ingest (Coordinate Updates), including the storage identification of the AIPs.

2.1.2.2 Replace Media

The Replace Media function provides the capability to reproduce the AIPs over time. Within the Replace Media function the Content Information and Preservation

Description Information (PDI) must not be altered. The migration strategy must select a storage medium, taking into consideration the expected and actual rates of errors encountered in various media types, their performance, and their costs of ownership.

2.1.2.3 Error Checking and Disaster Recovery

Error Checking provides statistically accepted assurance to ensure that none of an AIP's components are corrupted via any internal storage function.

The SERSCIDA-PDB performs routine and special data integrity checking such as checksums for each individual file and generates error reports. It also provides disaster recovery capabilities including data backup, off-site data storage and data recovery.

2.1.2.4 Provide Data

The Provide Data function provides copies of stored AIPs to Access. This function receives an AIP request that identifies the requested DIP(s), and then provides them on the requested media type or transfers them to a staging area.

2.1.3 Data Management

Data Management provides the services and functions for populating, maintaining and accessing both metadata, which identifies and documents repository holdings, and administrative data used to manage the repository (see figure 4). In the SERSCIDA-PDB the main functionalities of Data Management are integrated into ingest and metadata tools.

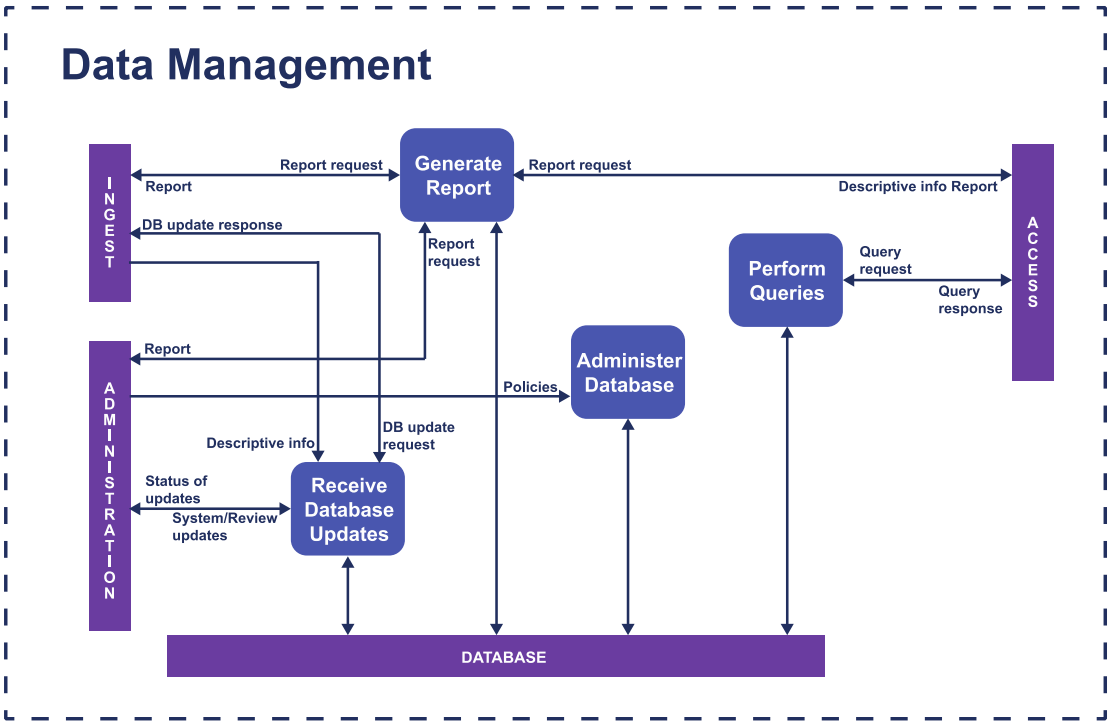

Figure 4: Functions of Data Management

2.1.3.1 Administer Database

The Administer Database function is responsible for creating any schema or table definitions required to support Data Management functions; for providing the capability to create, maintain and access customized user views of the contents of this storage; and for providing internal validation (e.g., referential integrity) of the contents of the database. In the SERSCIDA-PDB the Administer Database function is carried out in accordance with policies received from Administration.

2.1.3.2 Perform Queries

The Perform Queries function receives a query request from Access and executes the query to generate a result set that is transmitted to the requester. All queries will be verified against access rights restrictions.

2.1.3.3 Generate Report

The Generate Report function receives a report request from Ingest, Access or Administration and executes any queries or other processes necessary to generate the report that it supplies to the requester. Typical reports might include summaries of archive holdings by category, or usage statistics for numbers of access to archive holdings.

2.1.3.4 Receive Database Updates

The Receive Database Updates function adds, modifies, or deletes information in the Data Management persistent storage. The main sources of updates are Ingest, which provides Descriptive Information for the new AIPs, and Administration, which provides system updates and review updates. Ingest transactions consist of Descriptive Information which identifies new AIPs stored in the archive. System updates include all system-related information (operational statistics, Consumer information, and requests status).

Review updates are generated by periodic reviewing and updating of information values (e.g., contact names and addresses), and in the SERSCIDA-PDB this will be done via ingest and ticketing tools.

2.1.4 Administration

Administration provides the services and functions for the overall operation of the archive system (see figure 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.

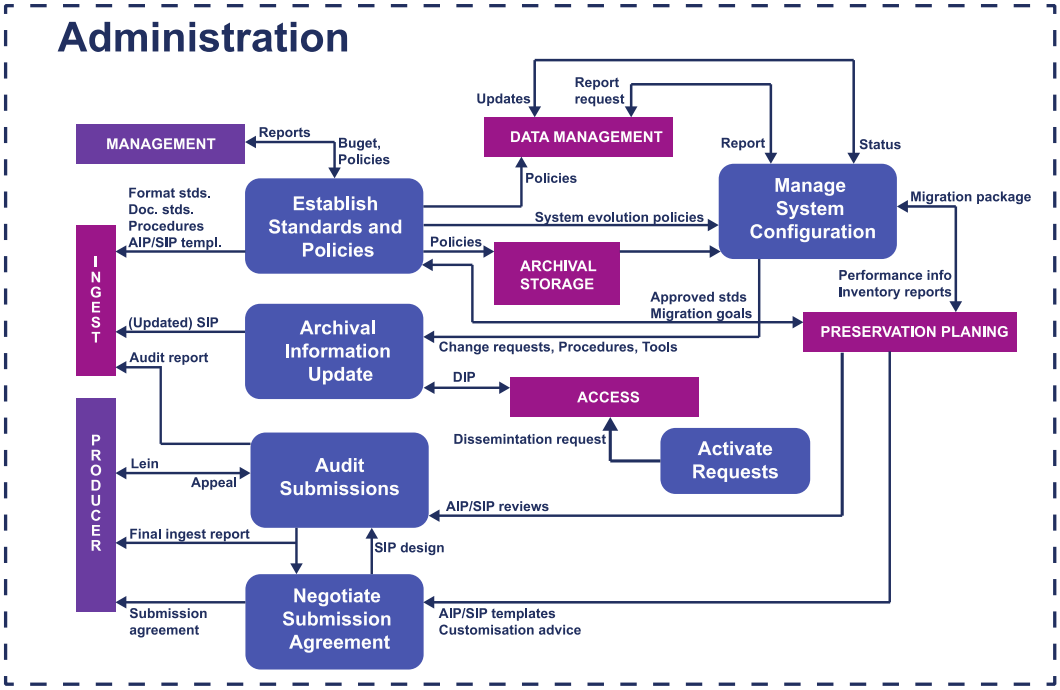

Figure 5: Functions of Administration

2.1.4.1 Negotiate Submission Agreement

The Negotiate Submission Agreement function negotiates submission agreements with the Producer. The submission agreement is a formal written agreement between the Producer and the archive defining the terms of the content, standards, metadata creation of the SIP, and future use by third parties. The data submission formats and procedures are clearly documented in the archive's data submission policies (see SERSCIDA D5.1 chapter 3.2).

2.1.4.2 Manage System Configuration

The Manage System Configuration function maintains integrity and tractability of the configuration during all phases of the system life cycle. It also audits system operations, system performance, and system usage. It sends report requests for system information to Data Management and receives reports; it receives operational statistics from Archival Storage.

2.1.4.3 Archival Information Update

The Archival Information Update provides a mechanism for updating Digital Objects (files and metadata) within the repository. This type of action may require the removal or update of the digital object and/or its associated metadata.

2.1.4.4 Establish Standards and Policies/Procedures

The Establish Standards and Policies function is responsible for establishing and maintaining the archive system's standards and policies. It receives recommendations for archive system enhancement, as well as proposals for new archive data standards

from Preservation Planning. It also receives performance information and archive holding inventories from the Manage System Configuration. Based on these inputs, archive standards and policies are established and sent to other Administration functions and the other Functional Entities for implementation. The standards include format standards, documentation standards and the procedures to be followed during the Ingest process. It provides approved standards and migration goals to Preservation Planning. This function will also develop storage management policies (for the Archival Storage hierarchy), including migration policies to ensure that archive storage formats do not become obsolete, and database administration policies. It will develop disaster recovery policies. It will also determine security policies for the contents of the archive, including those affecting Physical Access Control and the application of error control techniques throughout the archive (see SERSCIDA D5.1 chapter 5.1).

2.1.4.5 Audit Submissions

This function receives AIP/SIP reviews from Preservation Planning and may also involve an outside committee (e.g., science and technical review). The audit process must verify that the quality of the data meets the requirements of the archive and the review committee. The Audit process may determine that some portions of the SIP are not appropriate for inclusion in the archive and must be resubmitted or excluded. If material is not judged as suitable for collection and preservation, then the Producer will be notified. In some cases the Producer may be requested to modify the material, or provide alternate material. All the communication and decision-making will be recorded in the ticketing system.

2.1.4.6 Activate Requests

The Activate Requests function maintains a record of event-driven requests and periodically compares it to the contents of the archive to determine if all needed data are available. In the SERSCIDA-PDB this function will be performed through a regular checksum procedure.

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 the long-term, even if the original computing environment becomes obsolete (see figure 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 users' service requirements. Preservation Planning also develops detailed migration plans, software prototypes, and test plans to enable implementation of Administration migration goals.

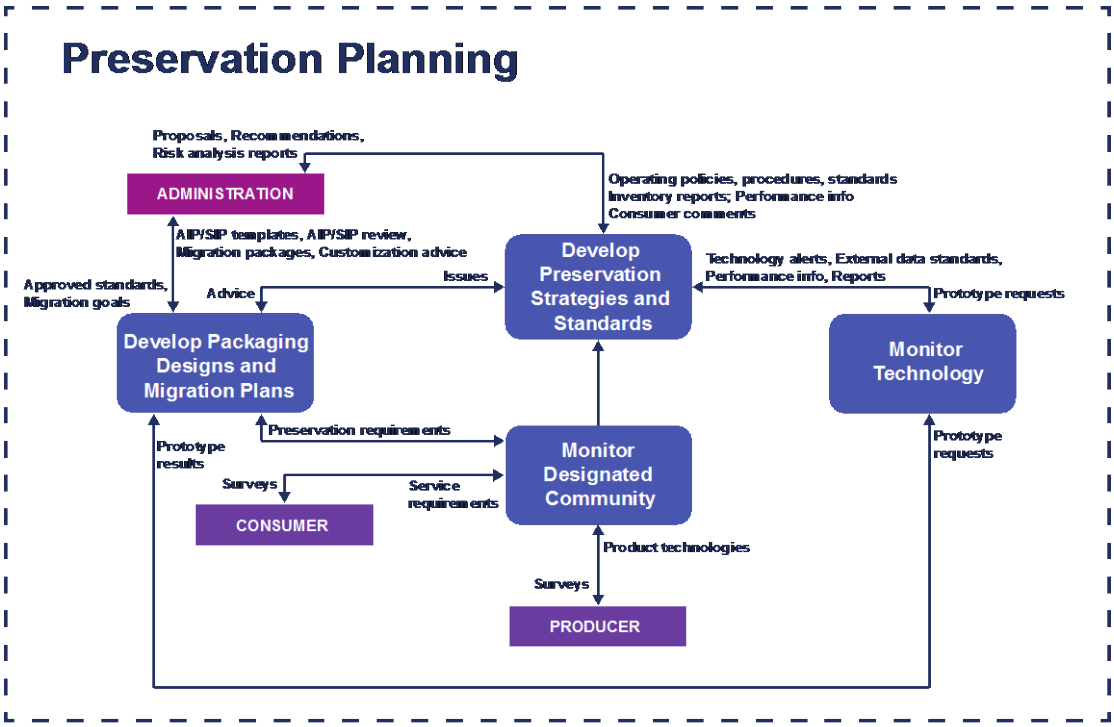

Figure 6: Functions of Preservation Planning

2.1.5.1 Monitor Designated Community

The Monitor Designated Community function interacts with the archive Consumer and Producer to track changes in their service requirements and available product technologies. Such requirements might include data formats, media choices, preferences for software packages, new computing platforms, and mechanisms for communicating with the archive. This function may be accomplished via surveys, a periodic formal review process, and community workshops, or by individual interactions.

2.1.5.2 Monitor Technology

The Monitor Technology function is responsible for tracking emerging digital technologies, information standards and computing platforms (e.g., hardware and software) to identify technologies that could cause obsolescence in the archive's computing environment and prevent access to some of the archives current holdings. This function may contain a prototyping capability for better evaluation of emerging technologies.

2.1.5.3 Develop Preservation Strategies and Standards

The Develop Preservation Strategies and Standards function is responsible for developing and recommending strategies and standards to enable the archive to better anticipate future changes in the user service requirements or technology trends that would require migration of some current archive holdings or new submissions.

2.1.5.4 Develop Packaging Designs and Migration Plans

The Develop Packaging Designs and Migration Plans function develops new IP designs and detailed migration plans and prototypes, and implements Administration policies and directives. This function also receives archive approved standards and migration goals from Administration. The standards include format standards, metadata standards, and documentation standards. It applies these standards to preservation requirements and provides AIP and SIP template designs to Administration.

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 7). Access functions include communicating with Customers to receive requests and applying controls to limit access to specially protected information. This includes coordination of the execution of requests to successful completion, generating responses, and delivering the responses to Consumers.

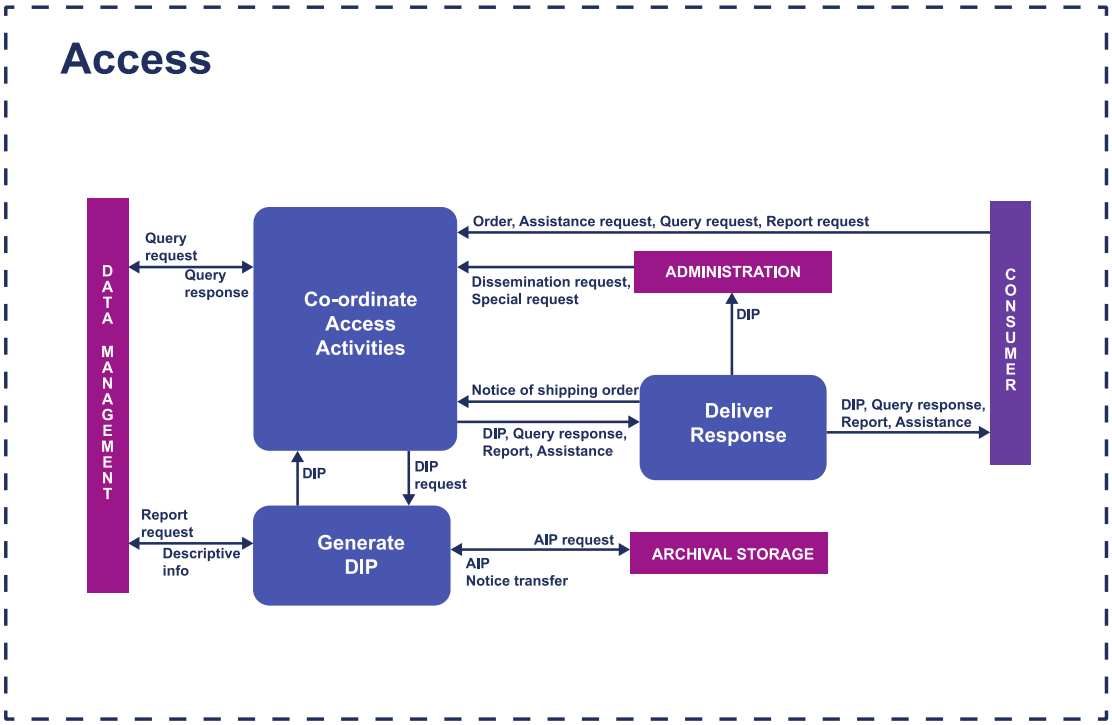

Figure 7: Functions of Access

2.1.6.1 Coordinate Access Activities

The Coordinate Access Activities function provides an interface layer that will support access functions. Rights restrictions may limit the level of access that can be provided to different types of users (see chapter 3.3.2).

2.1.6.2 Generate DIP

The Generate DIP function accepts a dissemination request, and retrieves the information required to form a Dissemination Information Package. The types of operations that may be carried out include statistical functions, conversions between different data types or output formats, and other specialized processing. This function places the completed DIP response in the staging area and notifies the Coordinate Access Activities function that the DIP is ready for delivery.

2.1.6.3 Deliver Response

The Deliver Response function identifies the intended recipient, determines the transmission procedure requested, places the response in the staging area to be transmitted, and supports the on-line transmission of the response.

2.3 Metadata Specifications

Metadata of a study will be described in Data Documentation Initiative (DDI) metadata specification, version 2.5 or higher, named DDI-Codebook, used for documenting simple survey data. This specification is used in European and North American data archives, libraries and official statistics agencies all around the world[1].

The main elements of the DDI-Codebook are:

- Document Description;
- Study Description (general study and methodological information);
- File Description;
- Data Description;
- Other Material.

Complete documentation is available on the DDI alliance web page[2].

The DDI is designed to be fully machine-readable and machine processable. It is defined in XML, which facilitates easy Internet access. Special Specification, based on CESSDA recommended fields[3], was defined for the SERSCIDA data services, and implemented in the SERSCIDA-PDB access tool (see chapter 3.2.5). DDI Controlled Vocabularies[4] and CESSDA topic classification are also used. Using a predefined topic classification will make inclusion in the CESSDA data catalogue[5] easier. These are topics by which you can browse in the catalogue.

Fields for ingesting in SERSCIDA-PDB ingest tool are made on CESSDA-recommended fields relevant for study.

2.4 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

[1] http://www.ddialliance.org/community
[2] http://www.ddialliance.org/Specification/DDI-Codebook/2.1
[3] http://www.ddialliance.org/sites/default/files/cessda-rec.pdf
[4] http://www.ddialliance.org/controlled-vocabularies
[5] http://www.cessda.net/catalogue/

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 any more in the future, and the frequency of migration and the costs of preservation are lower.

File formats matter during the whole 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 about which file formats are acceptable as archival and access 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 already at Ingest phase file formats are controlled and validated for which specific tools are available. 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;
- Support;
- Disclosure;
- Documentation quality;
- Stability;
- Ease of identification and validation;
- Intellectual Property Rights;
- Metadata Support;
- Complexity;
- Interoperability;
- Viability;
- Re-usability.

The selected file formats for the SERSCIDA-PDB represent a summary of different recommendations from CESSDA partners and internationally recognised institutions: [6]

[6] 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.

File formats considered as appropriate for SIPs:

- Tabular data: **SPSS portable format (.por)**, SPSS (.sav), Stata (.dta), Excel or other spread sheet 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.

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);
- 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.[7]

[7] One commonly used registry is the online registry about data file formats (PRONOM) developed by the National Archives: http://www.nationalarchives.gov.uk/PRONOM/ (see SERSCIDA D5.1 chapter 3.1 for more information).

3 Technical Specifications

3.1 Technology Considerations

The objective of the SERSCIDA data services is to become a national service provider and to provide international CESSDA-level services in their countries. CESSDA -- The Consortium of European Social Science Data Archives -- is a pan-European distributed research infrastructure with member benefits and obligations. To be included as new members, archives should to adjust to CESSDA requirements already at the time of establishment. For this reason, many tools, specifications, and vocabularies that are recommended by CESSDA will be used in new services.

The "old" dream of CESSDA members is still relevant today, and many of the goals that they wanted to achieve in the past have not changed much: all existing empirical data on-line; an integrated resource discovery gateway to identify and locate these data; extensive amounts of available metadata; the ability to carry out simple on-line browsing, visualization, and analysis; the ability to seamlessly subset and download data and metadata to specialized analysis or visualization tools; and efficient hyperlinks from data sources to relevant papers and notes.

To reach those goals the CESSDA archives developed the Nesstar software system. Nesstar is widely used by CESSDA members for data publishing (Access) and online analysis. It was originally developed by and for the archives, and it is designed to fit many important documentation and dissemination use cases of data archives. Nesstar was also the first tool to support DDI, which is still a highly relevant specification for data documentation. It is implemented in a way that the user does not need to have any knowledge of XML to generate DDI compliant data documentation[8]. Nesstar is currently used by most data archives in Europe, and by a reasonable number of data libraries in the US and Canada. There are currently more than 130 instances of the Nesstar Server worldwide. For all these reasons, Nesstar has been selected for use for the SERSCIDA-PDB Acces tool.

The selected platform for Ingest, Data Management, and Archival Storage was a collaborative process. For this purpose a special technical group was formed in late 2013. Group members prioritized together the requirements, and generated a weighted "criteria checklist" (see Annex 6.1). Besides these criteria, the tools that are used currently within CESSDA were evaluated as well (see Annex 6.2). After careful analysis and hands-on examination of several tools (e.g. EPrints, DSpace, Invenio and CKAN), EPrints was chosen as the tool for Ingest, Data Management, and Archival Storage. The selected tools for the prototype are described in the following sections.

3.2 Tools

3.2.1 EPrints

[8] http://www.nesstar.com/software/publisher.html

EPrints[9] is free and open source software, which can be used to build open access repositories. It fulfils all the criteria listed below:

- Easily archives submitted material;
- Has a robust database for storing data;
- Exposes the digital content to search engines;
- Easy to install and maintain;
- Secure;
- Compliant with open archives initiative protocol for metadata harvesting;
- Free and open source.

It was created by University of Southampton School of Electronics and Computer Science in 2000. The configuration files of EPrints are written in Perl / XML and can successfully run under multiple operating systems like Linux, Windows, Mac, and Solaris. It is easy to edit the configuration files according to the requirement. For example, the user interface of SERSCIDA-PDB is in English. Once it is no longer a prototype the user interface will very easily be translated into the local languages by changing an XML file in the EPrints configuration

SERSCIDA-PDB will use EPrints for the Ingest, Metadata management and Archival storage. Besides that, EPrints will be used for maintaining the user database. For all these purposes EPrints was installed with 3 instances (one for each archive):

- https://deposit.ba.serscida.eu
- https://deposit.hr.serscida.eu
- https://deposit.rs.serscida.eu

EPrints was customized by the SERSCIDA team in order to support special study metadata profiles (see chapter 2.3), SERSCIDA-PDB Ingest and Archival workflow (see chapters 3.4 and 3.5) and the user management facility (see chapter 3.3.1).

3.2.2 Request Tracker

Request Tracker (RT)[10] is an enterprise-grade ticketing system that enables a group of people to intelligently and efficiently manage tasks, issues, and requests submitted by a community of users. It is written in Perl, and it is very easy to extend the RT interface using plug-ins written in Perl.

All three data services will use the same instance of the RT, but each of them will have one queue: https://www.serscida.eu/rt

The RT will be used for all communication processes and will support the user registration process (see chapter 3.3.1).

3.2.3 Nesstar Publisher

Nesstar Publisher is described as an advanced data management program. It consists of data and metadata conversion and editing tools, enabling the user to prepare these materials for publication to a Nesstar Server. However, it can also be used as a stand-

[9] http://www.eprints.org/software/
[10] https://www.bestpractical.com/rt/

alone tool for the preparation of data and metadata, and then later published through other platforms. The Publisher enables users to enhance datasets by combining a wide range of catalogue and contextual information, which can then be viewed within the Nesstar web client, Nesstar WebView[11].

Nesstar Publisher is a desktop standalone application available free of charge, which can publish data from a variety of sources (such as SPSS, STATA, SAS, Excel, etc.). The tool includes a specialised metadata editor, data and metadata validation routines, and metadata templates that provide standardisation and control (Ward, 2006[12]). On the basis of a metadata template for Nesstar Publisher used in ADP[13], a SERSCIDA template was developed. It includes the CESSDA topic classification and DDI controlled vocabulary described in chapter 2.3. In the current version of Nesstar (4.0), Publisher is the only possible tool for publishing data and documentation on a Nesstar Server.

In SERSCIDA-PDB the Nesstar Publisher is used for preparing for preparing of metadata description for the SIP as well as for converting AIPs into DIPs and preparing the data for Access.

3.2.4 Nesstar Server

The Nesstar Server is built as an extension to a normal web server[14]. It includes an SQL-based metadata management system, a data storage system, a powerful statistical engine as well as a flexible access control system[15]. Information needed for its installation and maintenance are provided[16] by the Norwegian Social Science Data Archives (NSD). An Administrator can edit most of the server configuration properties and can perform basic operations like starting/stopping the server by the Configuration Tool[17].

The Nesstar Server will save distribution materials (parts of DIP) of SERSCIDA archives. The Database of users who will download the data and access regulations is saved in its SQL-base (see chapter 3.3.1). With respect to Access, the SERSCIDA-PDB relies on the Nesstar Server, in conjunction with the Nesstar WebView.

3.2.5 Nesstar WebView

The Nesstar component used for presenting data and metadata of a study and for on-line analysis and download is Nesstar WebView. It is a web-based system optimised to operate with all commonly used web browsers (e.g., Internet Explorer, Firefox, Google Chrome, Opera, etc.).

Key features of Nesstar WebView are[18]:

- Searching and Browsing;
  - Simple and Advanced search facilities;
  - Ability to browse metadata, data and associated documentation;

[11] http://www.nesstar.com/software/publisher.html
[12] www.esds.ac.uk/news/eventsdocs/bestmwmar06.ppt
[13] http://www.adp.fdv.uni-lj.si/serscida_wp4_lj2013_workshop/presentations.html
[14] http://www.nesstar.com/software/server.html
[15] http://www.powershow.com/view1/1a6170-ZDc1Z/NESSTAR_powerpoint_ppt_presentation
[16] http://www.nesstar.com/help/4.0/server/index.html
[17] http://www.nesstar.com/help/4.0/server/knowing_server/components.html
[18] http://nesstar.com/help/4.0/webview/index.html

    - Ability to view metadata in a multiple languages;
- Analytical Tools;
    - Graphical representations of data in various forms, including maps;
    - Cross tabulations;
    - Correlation and regression analyses;
    - Ability to apply survey weights;
- Data Access;
    - The ability to download survey datasets in a variety of statistical formats;
    - A subset function to enable downloads to be customised by the user;
- Sharing;
    - Linking to any resource on the server;
    - Embedding views (tables, graphs and maps) in external webpages.

An extensive user guide[19] is available. This will in the future be partly translated in the local languages.

3.2.6 CESSDA Catalogue

All SERSCIDA archives, once established, will publish their data in the CESSDA Catalogue. The CESSDA Catalogue started 20 years ago when the first CESSDA website was launched in the spring of 1994. The website consisted of a map with links to eleven CESSDA archives and nine other data archives outside of Europe. When the CESSDA Expert Seminar was held in Grenoble in September 1994 the topic was 'Networks and Social Science Data Archives'. The central session of this expert seminar was the one discussing future prospects for harmonizing and integrating CESSDA Data Archives. The results of this discussion were shown already in November of the same year, when the first integrated data catalogue (IDC) was presented as part of the CESSDA Catalogue. The IDC made it possible to search data among the collections of 10 data archives. It was decided that at least a minimum of the information should be translated into English.

After the first years there followed a decade with hardly any changes to the portal. In April 2005 the CESSDA General Assembly (GA) decided it was time to form a Website Planning group. The group started to work in August the same year, and in October a website session was held at the Expert Seminar in Madrid. The first stage of renewal, including only the most urgent minor modifications, was completed at the end of 2005. At the same time the planning group made a detailed proposal to the CESSDA General Assembly to renew the site.

Just over a year later, in March 2007, the new CESSDA Catalogue could be browsed by topic, keyword, or data publisher, as well as searched using a free-text search. The CESSDA Catalogue enabled users to locate datasets, as well as questions or variables within datasets, stored at CESSDA archives. The number of participating data archives is presently fifteen and the data collections include sociological surveys, election studies, longitudinal studies, opinion polls, and census data.

To overcome the language barrier, the European Language Social Science Thesaurus (ELSST) is used to aid retrieval. ELSST is a broad-based multilingual thesaurus for the social sciences. It is originally based on the monolingual thesaurus Humanities and Social Science Electronic Thesaurus (HASSET) of the UK Data Archive and the source

[19] http://www.nesstar.org/Release-free/WebView_Userguide_v3.pdf

version of ELSST is the English thesaurus. Over the years ELSST has been developed by the members of CESSDA, and the thesaurus is currently translated into eight languages: Danish, Finnish, French, German, Greek, Norwegian and Spanish and Swedish. The thesaurus is updated annually and translation into further languages is planned.

Currently CESSDA (in the form of the new "CESSDA-AS") has given top priority for 2014-15 to upgrading and improving the CESSDA Catalogue. Including data from the three new SERSCIDA data services will increase their international visibility and will help move the services closer to eventual membership in this European infrastructure.

3.3 Users

3.3.1 Database of Users

Initial data will be collected from the user using Word Press (WP) web form (Ninja forms plugin). Once the user fills in the form and submits it for the registration, the WP will send the data to Request Tracker (RT) tool that will be used for further communication with the user. New ticket in RT will be automatically created, and the user will receive automatically generated message from RT with further instructions. Management will be notified about that transaction, and will continue communication with the user from RT. User will be asked to sign and send End User License Agreement. Management will check if the user has rights to access data according the data access policies and will open the same user account in EPrints (for accessing documentation and related documents) and Nesstar (for accessing data). User account data are than sent to user together with basic instructions on how to download data. All the registration process together with the signed contracts will be captured into Request Tracker.

The database of users in SERSCIDA-PDB will be maintained and managed using Word Press, which allows updates and exporting of user submitted data. EPrints and Nesstar will hold username, password, email and user rights. RT will store just user email, which can be used to identify and connect correspondence stored in RT with other user actions performed in other systems.

This solution needs to be improved in the future when the archives will start with production phase. Main advantage of this solution for SERSCIDA-PDB is that it was easy to implement without additional development. It was made functional and ready for testing using just the existing functionalities of the software tools already used for other components. Part of the process will still need manual work in SERSCIDA-PDB, but this will allow archive staff to understand the communication with user better. This experience should serve as information for software developers to better support the whole process in the future.

Future improvements could be implemented by connecting EPrints and Nesstar user databases where EPrints could take the role of master database. One-way automatic synchronization of user accounts from EPrints to Nesstar should be possible with customizations at both ends: extending EPrints user metadata, developing user creation trigger for pushing data to Nesstar and eventually extending Nesstar functionalities such as supporting "@" character in username field.

3.3.2 Login, User Roles and Permissions

The users of the SERSCIDA-PDB include both end-users (Producer and Consumer) and the archive staff (Management) who will be working with and managing the content and repository system. A crucial feature of SERSCIDA-PDB will be providing the various access levels required by the different types of users. The archive staff and system administrators will require access to the repository in order to ingest, administer, manage, preserve and access objects and their supporting information and structures. This will require multiple levels of access to the materials.

A role is a collection of permissions defined for the whole system that can be assigned to specific users in specific contexts. The combination of roles and contexts define a specific user's ability to do something within the system.

In SERSCIDA-PDB EPrints there are 3 types of roles defined:

1. Accredited user – has general read and write access; this role is assigned to the user manually, after the user signs the End user Licence Agreement;
2. Editor – archive staff; has all content related rights;
3. Admin – IT group; has all content and system related rights.

End-users will access metadata documentation, questionnaires and other related materials from surveys via different platforms. For the prototype we will offer these materials via Eprints and Nesstar. Metadata will be openly accessible.

Data from surveys will be available via Nesstar only. Most of the surveys will be offered as Open access surveys, meaning they are available free of charge, while others might have specific access conditions. However, all users should comply with access regulations set by the data repository and fill in and sign the End user Licence as explained in previous section. Nesstar offers the possibility to define access conditions for specific surveys. This way archives will be able to define their own rules for access.

The Nesstar Server provides an Access Control Unit (ACU) [20] that controls access to data and operations made available by the server.

The ACU controls functions such as the following:

- Entrance of authorized and unauthorized users.
- Access to operations for authorised users.
- Challenging of users (when necessary) by requesting they perform specific actions to gain access to a particular resource.

More information on how to work with the ACU is defined in guides on the Nesstar web page. See Understanding the Nesstar ACU Policies[21] and the introduction to Access Control Unit[22].

Three different ACU files were created for the SERSCIDA data services. A hierarchy of users was defined. For instance, researchers can perform more operations than a student. We also defined some detailed actions, such as rules for download and on-line analysis of data.

[20] http://www.nesstar.com/help/4.0/server/access-control-unit/acu-introduction.html
[21] http://www.nesstar.com/help/4.0/server/access-control-unit/acu-preloaded-policies
[22] See footnote 10.

For more information on the different components of Nesstar, see section 3.2.

3.4 Ingest

The whole Ingest process of SERSCIDA-PDB based on the OAIS reference model, is built using the EPrints and RT software.

After the login into the EPrints platform, the Producer can create a study with the specific metadata and upload files into the EPrints system. Once all the data and metadata are prepared, the Producer submits the study to the system. The Producer will not have any more rights to work on this study. Alternatively, the Producer can send the data to Management using other modes of transfer, which can be agreed between the Producer and Management. In that case, Management will submit data into EPrints, instead of the Producer.

After the submission, Management will receive notification that data were submitted and after the login into EPrints will see all the submitted content and will be able to work on it. If some information are missing or the Producer needs to rework the content, Management will send the study for rework and the Producer will be able to continue working on that study (both metadata and files part). Alternatively, if initial data were submitted to EPrints by Management, the need for improvements will be communicated with the Producer via email communication and RT. In that case, Management will make all changes in EPrints.

In general, the whole process of the communication between Producer and Management will be recorded into Request Tracker and all the changes performed in the EPrints will be captured into EPrints system. Management will be able to see and compare all the changes made on one study. Once the content is ready for the archiving, Management will create the final version of a SIP.

For alternative workflow, where Management receives the data from the Producer using other modes of transfer, the following tools could be used for that kind of transfer:

- Filesender (big files transfer over https) https://www.assembla.com/spaces/file_sender/wiki;
- Owncloud (also web interface but richer) http://www.owncloud.org.

Neither requires end users to have account on those systems. Owncloud enables some kind of collaboration platform for Archive staff. Sending data by non-encrypted email is not recommended.

For the future production system EPrints software needs to be further customized. Primarily, more metadata needs to be added to support collection of all useful information from the producer in a structured way. Also, EPrints should be able to export metadata in DDI format. The interface will need some improvements as well to visually connect the service to the particular archive. While making these changes, experiences from development of ReShare and ReCollect plugins for EPrints developed by UKDA staff for self-archiving will be used to guide the future development.

3.5 Data Management and Archival Storage

The Data Management and Archival Storage processes of SERSCIDA-PDB are built using EPrints, Request Tracker, and Nesstar Publisher.

In this phase, Management will work on creation and preparation of an AIP. Creation of all Information Packages in SERSCIDA-PDB is done by using the EPrints versioning facility and by selecting "SIP", "AIP" or "DIP" as the Information Package for specific study. By creating a new version of a study in EPrints and selecting for an Information Package "AIP", Management will start with the creation of an AIP. This new version of a study will have all the metadata and files from the SIP. To preserve and distribute a study properly, several management steps need to be made. Data archives will use several different tools and techniques to prepare all materials related to a study (that is, questionnaire, codebook, reports, and other related documentation), to clean and label data files (statistical programmes like STATA, R, SPSS), and to prepare a metadata description.

The tool for metadata management needs to meet certain criteria. Most notably, it needs to be DDI compliant (the specification that we chose for metadata description). From the tools that are available (see list of tools on DDI Alliance web page[23]), we decided to use Nesstar Publisher. Since we decided to use Nesstar as a distribution tool, it was logical to us to use Publisher, which is closely connected to other Nesstar components. After all data are prepared and uploaded into the AIP version of a study, Management will finalise the AIP by locking this version of a study.

3.6 Access

The Access process of SERSCIDA-PDB, again based on the OAIS reference model, is built using EPrints and Nesstar WebView. The DIP will be created from the AIP by adding a new version to a study, selecting the DIP for Information Package, and changing files and metadata in order to prepare DIP for dissemination. The users will access metadata documentation, questionnaires, and other related materials from surveys via Nesstar WebView and some special documents via EPrints. Metadata will be openly accessible.

There will be 3 levels of access, depending on the nature of the data:

- Data and documents that can be accessed by anyone – they will be accessible via the Nesstar WebView application to all registered and unregistered users;
- Data and documents that are only available for scientific analysis carried out in academic research and teaching - they will be accessible via the Nesstar WebView application to all registered users that sign the End user Licence Agreement; and
- Data that are released for academic research and teaching, where the user may gain access to data subject to written approval from the depositor. For this purpose the user must supply the archive with their details and specify their analytic intention. The data service will request written permission from the depositor and will only permit access on receipt of this. Those data will be stored in EPrints and will only be sent to the Consumer after he/she signs a specific contract related to those specific data. This process and the communication between the data service, Producer, and Consumer will be done via email.

[23] http://www.ddialliance.org/resources/tools

3.7 Communication

3.7.1 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 by monitoring the community and technology in order to meet the needs of the producers and consumers (see figure 8).

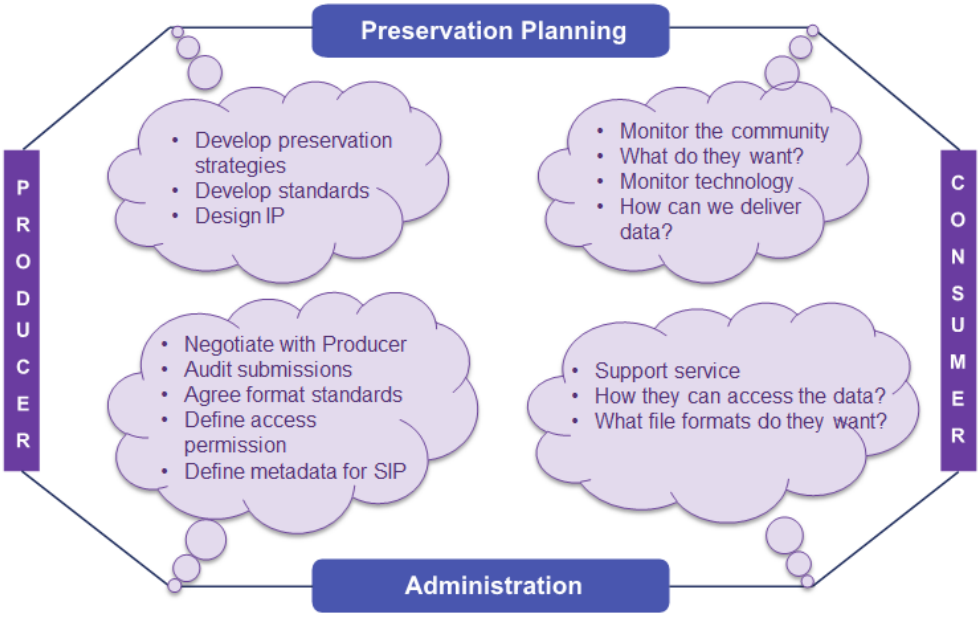

Figure 8: Communication

3.7.1.1 Website

The most common and wide-reaching channel to communicate with the community is by means of the 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 (see SERSCIDA D5.2).

3.7.1.2 Mailing Lists

A mailing list of potential users is established by the data archive in order to inform Producers and Consumers about the latest news and upcoming events, such as training and workshops. Each data service will have its own mailing list for those purposes:

crodas-users@serscida.eu
usdas-users@serscida.eu

sr-ssdc-users@serscida.eu

All SERSCIDA partners will keep as a communication channel the general SERSCIDA mailing list serscida@mjesec.ffzg.hr.

3.7.1.3 Direct Contact

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

3.7.2 Specific Communication

All the specific communications with the users during user registration and Ingest will be recorded and maintained via Request Tracker (see chapter 3.2.2).

3.8 Technical Infrastructure

3.8.1 Server Architecture for the prototype

For the implementation of the SERSCIDA project several servers where used (see figure 9):

- Virtual server 1 (in Croatia);
- Virtual server 2 (in Croatia);
- Local servers (in each WB country);
- CESSDA server.

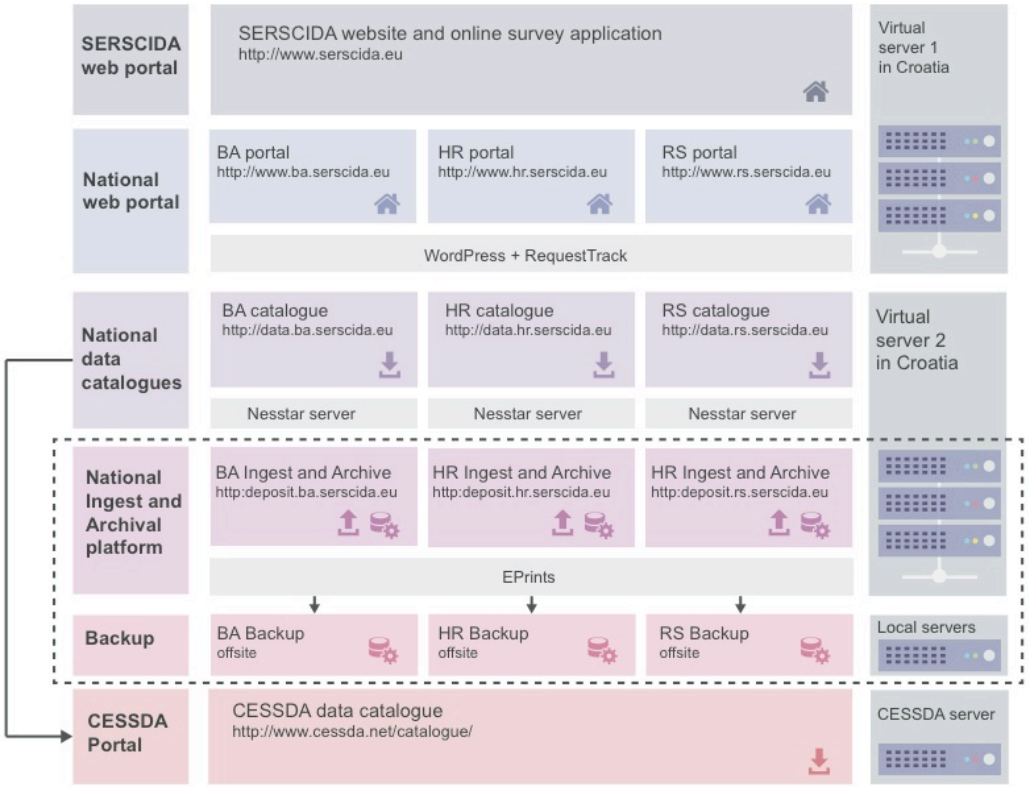

Figure 9: Server Architecture

The Virtual server 1 is used for the hosting of the SERSCIDA web portal, each national web portal, and the collaboration tool. A single WordPress application with three website instances (one for each WB partner) is installed for the national web portals (see SERSCIDA D5.2). Here is the detailed Virtual server 1 configuration:

Configuration: 2 vCPU, 8.0 GB, 100.0 GB
OS: Debian GNU/Linux 7.4 (wheezy)
Tools: Optimized (version 6.0 installed)
HA: Not Enabled

The Virtual server 2 is used for the national catalogues and Ingest/Archival platforms. A single instance of a Nesstar server for each national catalogue, and single EPrints installation with 3 instances for the Ingest/Archival platforms, are installed. Here is the detailed Virtual server 2 configuration:

Configuration: 2 vCPU, 12.0 GB, 100.0 GB
OS: Debian GNU/Linux 7.4 (wheezy)
Tools: Optimized (version 6.0 installed)
HA: Not Enabled

Both Virtual server 1 and Virtual server 2 are established using the Virtual Private Server (VPS) service, provided by the University of Zagreb University Computing Centre (Srce). This service offers redundant IT infrastructure, monitoring, and backup. In addition to local (on-site) file and database backup, there will be a daily automatic off-site backup solution as well. The local servers will also be used for backup purposes.

The archive will provide metadata to the CESSDA Catalogue, which will give access to the archive's portion of the harvested/indexed metadata holdings, and to a set of tools that can help improve quality and consistency in the metadata. The CESSDA server is configured and maintained by the UKDA.

3.8.2 Network and Telecommunications

The network infrastructure and telecommunications are accessed using the host organisations' systems: University of Sarajevo, Human Rights Centre for USDAS; University of Zagreb, Faculty of Humanities and Social Sciences for CRODAS, and Institute of Economic Sciences Belgrade for SR-SSDC.

3.8.3 Hardware and Software for production systems

Based on best practice and international standards for social science data archives the data services have determined the hardware and software they will use (see Annex 6.4 and Annex 6.5).

In order to enable proper functioning of the SERSCIDA-PDB even after the end of SERSCIDA project several software components were acquired during the SERSCIDA project, as well as servers and workstation (desktop) computers. The SERSCIDA-PDB will be used for promotional activities by WB partners involved in this project, and indeed to facilitate policy dialog and to open new opportunities for future funding. Domain name serscida.eu and related security certificates for it will last until 2017. Servers acquired during the SERSCIDA project are able to support the new archives for the 5 years lifetime, in spite of disk failures that are expected in this time frame or

occasional hardware faults. The system should be also easy to maintain. In order for the system to function in this way, hardware should be configured according to the server architecture plan for production systems described in the next chapter (see chapter 3.8.4). In this plain it is also mentioned which software tools will be used on the server side for the first phases of production deployment.

Workstation computers will be used by future archive staff for Data Management. Several tools for preparing materials related to a study will be installed on these desktop computers. This includes: office tools; conversion tools; software for statistical analysis (STATA, R, SPSS); tools for preparing metadata description of a study, etc.

The licence for using Nesstar Server was acquired by special offer under very low price. This price was kindly offered by the CESSDA, as CESSDA supports establishment of new data archives. The licence is valid for the next 3 years.

The licence for 4 concurrent users of the STATA 13 software, network version, as well as Stat/Transfer for 5 concurrent users were acquired. STATA will be installed in Croatia, but all other archives can use the software within the limits of the licence. For office tools, LibreOffice can be used for free, and if the archive wants to use proprietary product they will have to buy the licence or use the existing licence of their hosting institution.

3.8.4 Server architecture plan for production systems

Current prototype implementation is set up using two virtual machines located at the University of Zagreb University Computing Centre. The usage of virtual machines was valuable for prototype implementation, but in the production system archives should have more granular distribution of services. Sensitivity of data in various components, require us to think about different security levels of data and preservation requirements. To achieve this goal, the future architecture will be separated on different virtual machines, based on different platform deployment stacks:

- Mailing lists, virus checker (communication);
- WordPress, user form (communication);
- Request tracker (administration, communication);
- EPrints (ingest, SIP);
- EPrints (archive, AIP);
- Nesstar (dissemination, DIP).

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

In current state of data archive software support tools development in general, it seem that changing of components is highly probable in the future. Currently available tools all have certain advantages and disadvantages in supporting the whole process of data archiving. There is no integrated tool available, to support the whole process of data archiving, but rather there is a combination of different tools, connected together to support different stages of the data archiving process. This is one of the reasons why the easy maintainability of the system is important. Archive staff needs to be able to test other available software tools, as they will appear, preferably under Free/Libre/Open

Source licences.

To address security of applications (which might come with their own set of bugs) each virtual machine is secured by itself, so, for example, security problems in web portal won't affect archival copy of data in archive or any other component.

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

Two distinct types of data required for keeping in the archive were identified - SIP and AIP, which require long-term preservation together with audit log. This also requires ability to check whether data is correct on media that requires checksums on level of file system (scrubbing). For this requirement, ZFS[24] storage and snapshots using LVM will be implemented to provide long-term archival copy of current prototype on different locations (in faculty building) which is updated daily (from computing centre location). This enables disaster recovery in case of one location failure. It's also possible to have multiple remote copies if needed.

Management of applications and data will be done using Ganeti[25], open source cloud solution that enables high availability for virtual machines and provides data storage requirements outlines above. Hardware configuration as selected according to those requirements.

This architecture will require continuous maintenance, partly because software stacks changes, partly because new tools enable archives to offer new and better services.

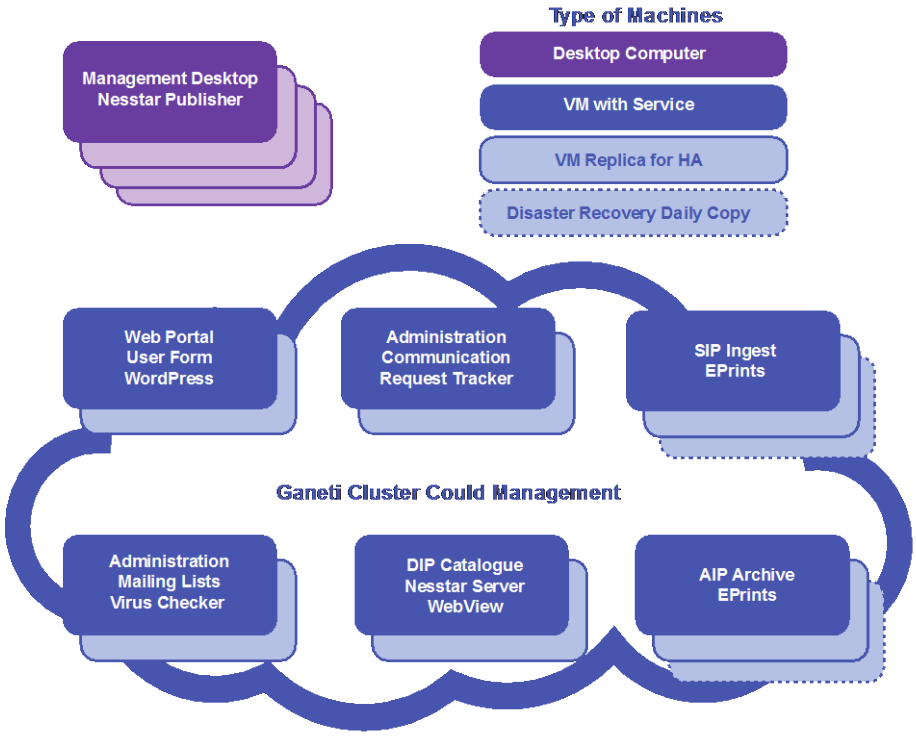

Figure 10: Server Architecture plan for production

[24] http://bit.ly/dc14-zfs
[25] http://bit.ly/dc14-ganeti

4 Use Cases and Screenshots

4.1 Receiving data from the depositor (SIP), creating AIP and DIP in EPrints

4.1.1 Receive data

Producer makes this action, but Management can do it as well.

1. Log in to EPrints, go to Manage Deposits

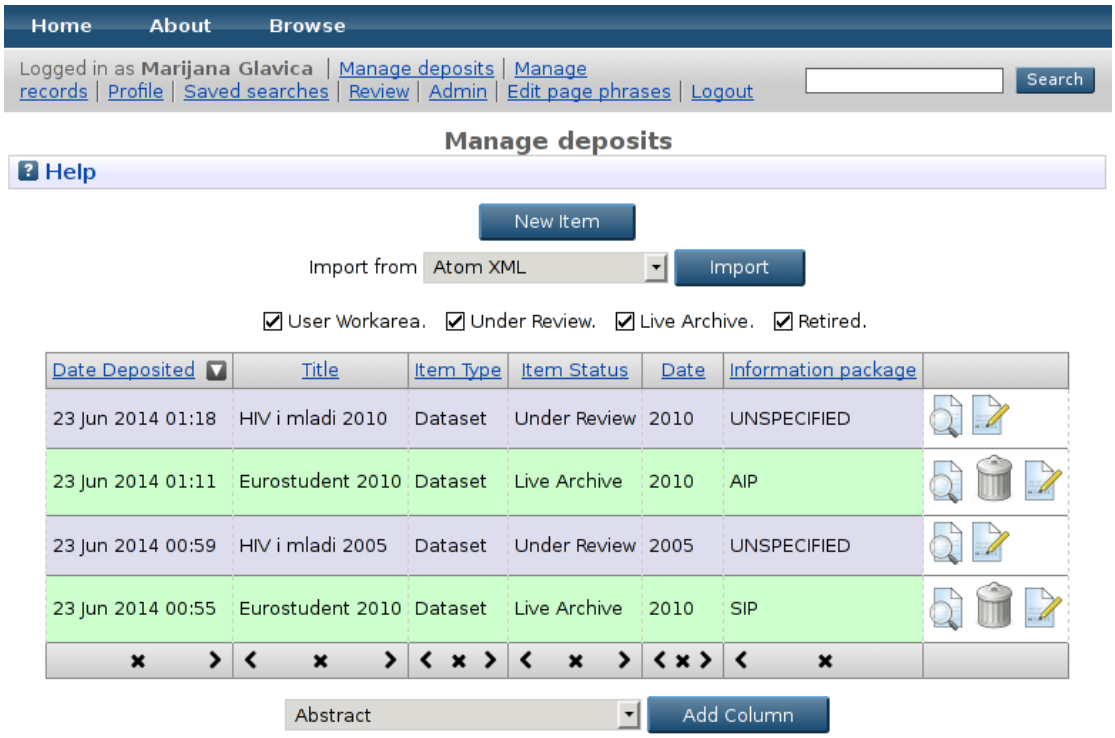

Figure 11: EPrints

2. Choose New Item and follow the workflow offered by the system:
   - Upload files, change file options (content version, permissions and language, and optionally for type, licence, and embargo permissions),
   - Enter metadata and subject information
   - Deposit item (sent to review)

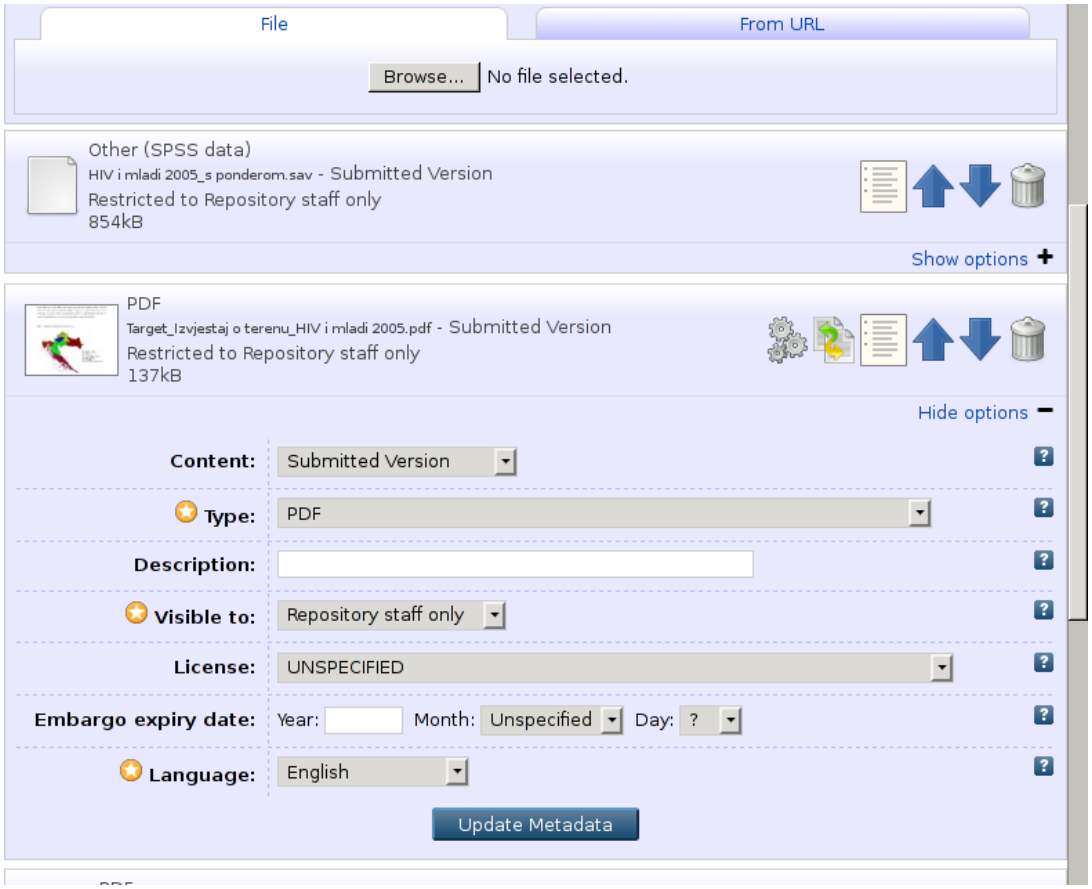

Figure 12: EPrints

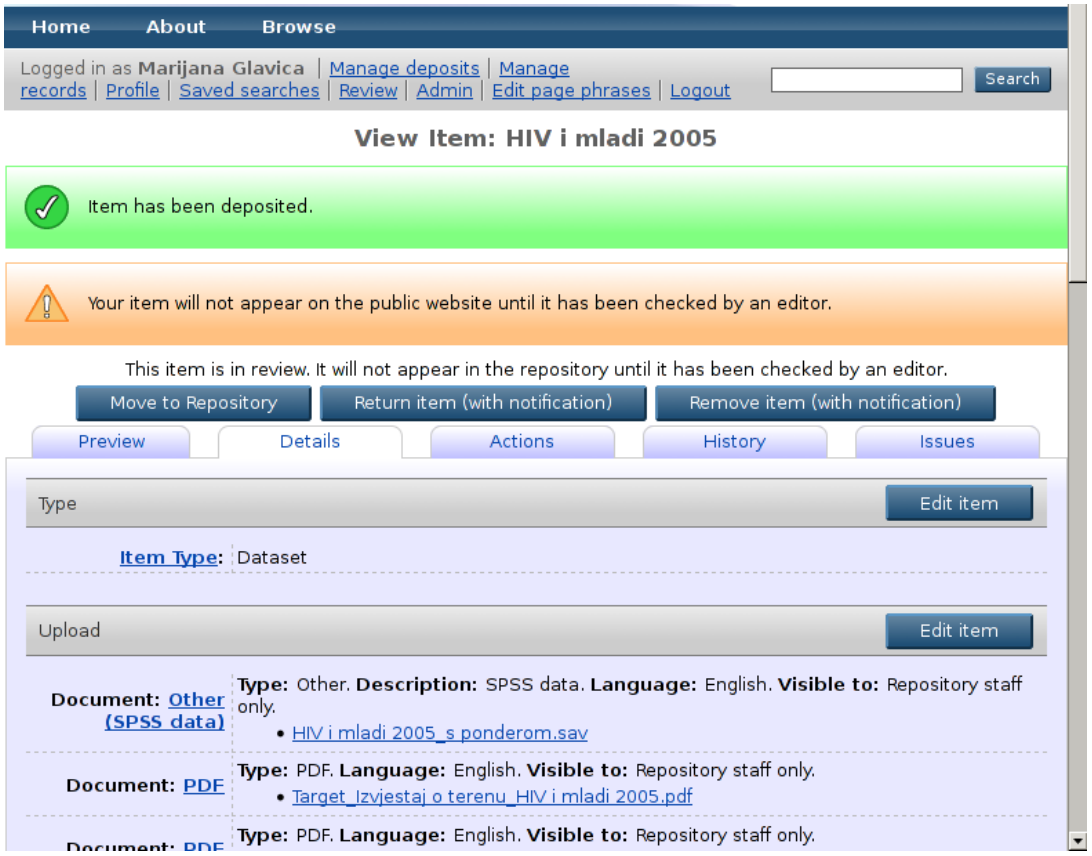

Figure 13: EPrints

In this phase the item is in review and it needs to be accepted by the Management. The Producer will not be able to change files until the Management allows and requests changes.

4.1.2 Review submitted data and create SIP

1. To review and accept submitted files choose "Review" option from admin menu.

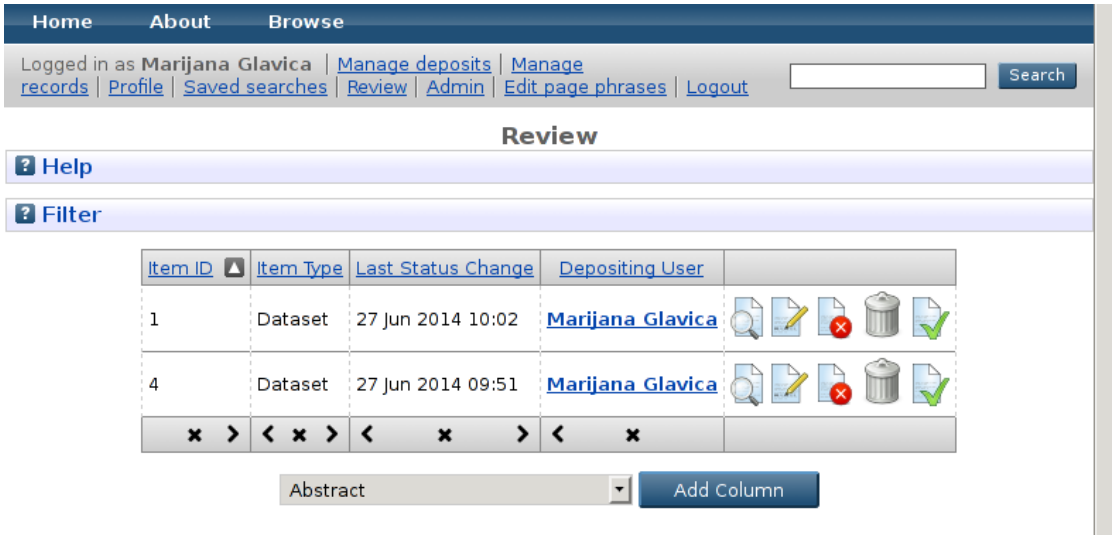

Figure 14: EPrints

2. Management will:

- Inspect submitted files and metadata;
- Add additional files and edit metadata;
- If necessary, make a request to Producer to upload additional files and enter more information.

When all material is received and checked, Management will mark the submission as SIP, and move (publish) the item to repository.

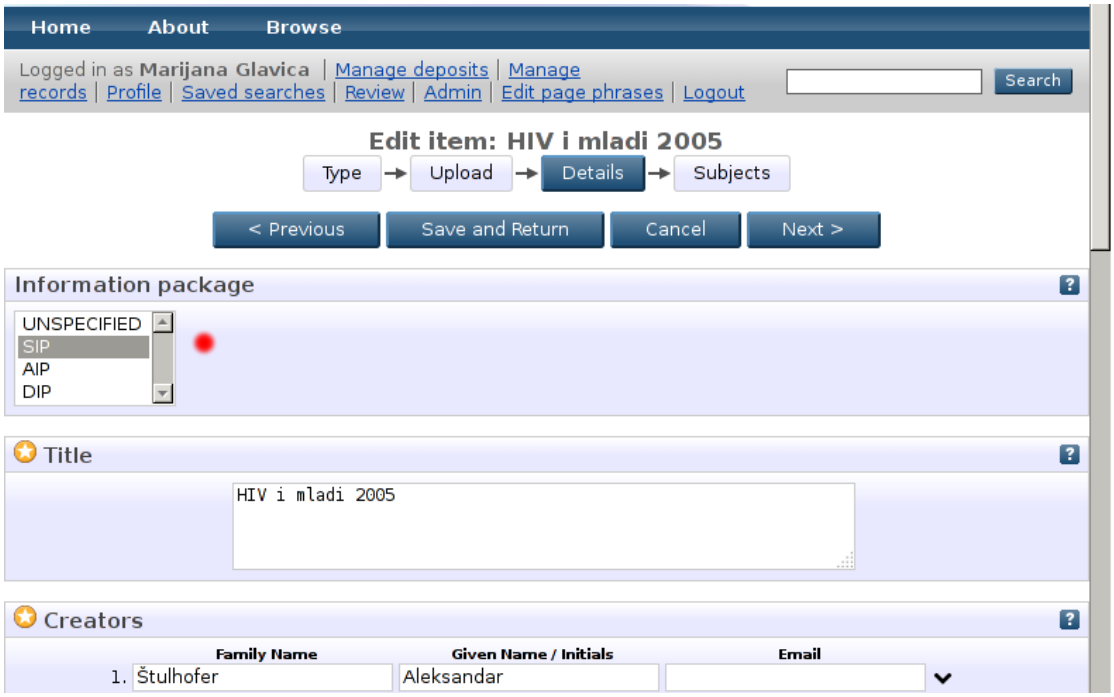

Figure 15: EPrints

4.1.3 Create AIP

1. From "View item" screen for the SIP version, select "Actions" and then "New version".

2. Exact copy will be created by the system and relationship with previous version will be automatically recorded.

3. Mark this version as AIP. Change metadata and files. Convert files to other formats.

4. Publish the new version.

4.1.4 Create DIP

1. From "View item" screen for the AIP version, select "Actions" and then "New version".

2. Exact copy will be created by the system and relationship with previous version will be automatically recorded.

3. Mark this version as AIP. Change metadata and files. Publish the new version.

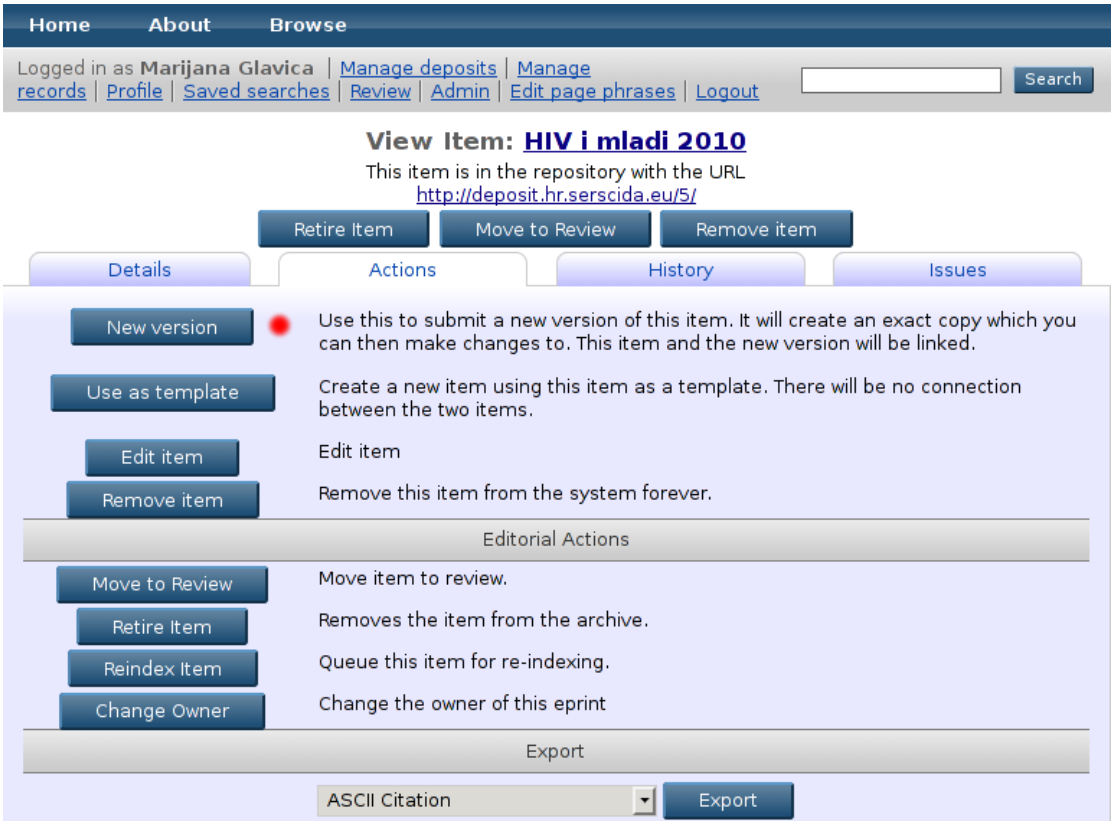

Figure 16: EPrints

4.2 User registration (WP), creating user accounts (EPrints and Nesstar), and communication with users (RT)

4.2.1 User registers using the registration form in WP

User registration form for CRODAS:
http://www.hr.serscida/eu/registration
User registration form for USDAS:
http://www.ba.serscida.eu/registration
User registration form for SR-SSDC:
http://www.rs.serscida.eu/registration

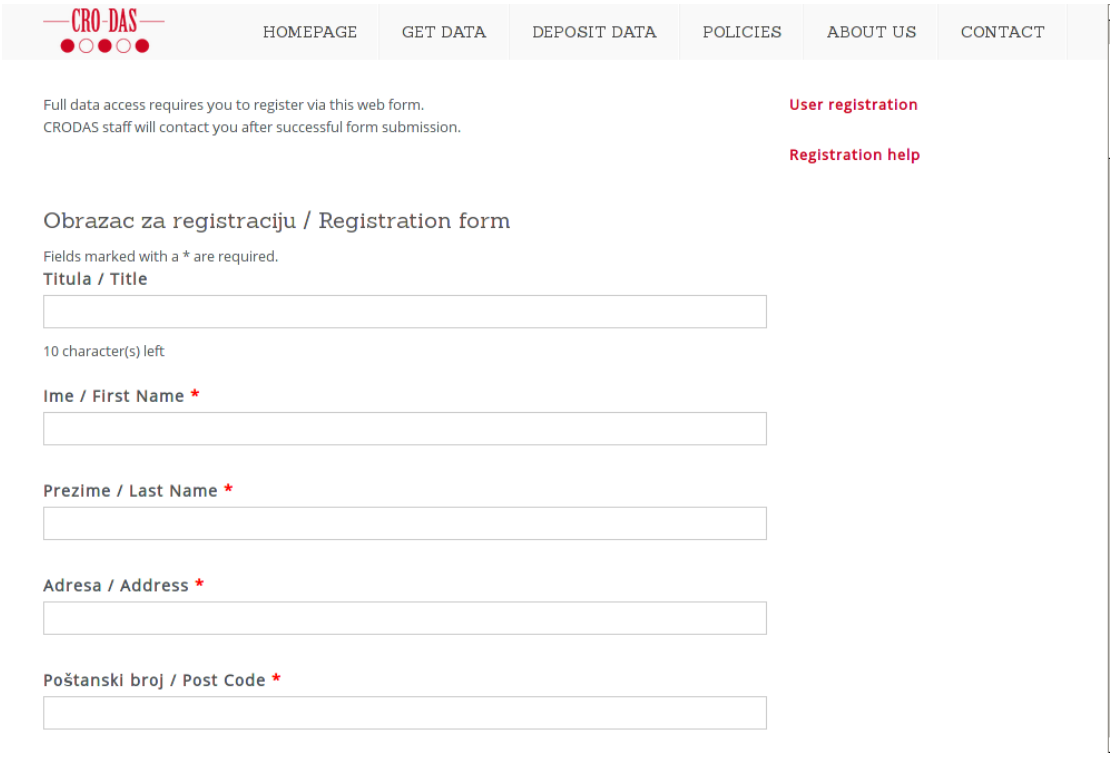

Figure 17: User registration form

4.2.2 WP registration form sends email to RT and creates RT ticket (automatic)

Email is sent to:

- hr-users@serscida.eu for CRODAS;
- ba-users@serscida.eu for USDAS;
- rs-users@serscida.eu for SR-SSDC.

RT ticket is created in appropriate queue:

- hr-users queue for CRODAS;
- ba-users queue for USDAS;
- rs-usersc queue for SR-SSDC.

Requestor field is filled in with users email.

STEPS:
1. Go to https://serscida.eu/rt/ and log in
2. Notice created ticket under Ten Newest Unowned tickets tab
3. Go to dashboard (under Dashboards tab or mouse over Home from the top left menu)
4. Notice created ticket in appropriate queue (left is for users, right is for depositors)

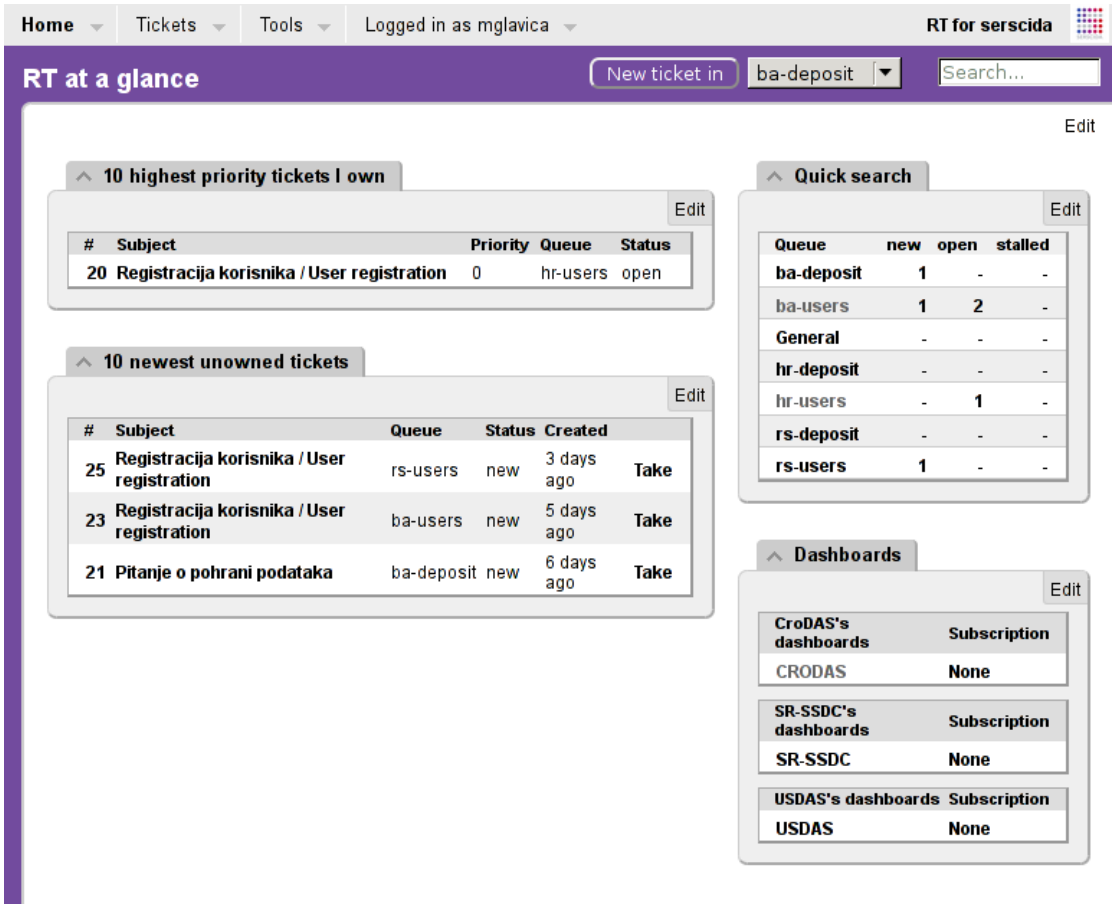

Figure 18: RT

4.2.3 RT sends confirmation e-mail to the user and notification to staff (automatic)

STEPS:
1. Click on ticket subject - go to ticket display screen
2. See the content under "History" tab
3. Go to end of the screen
4. Click on "Show" near the last two messages and see what system has sent to admins and user

4.2.4 User receives confirmation message from RT and archive staff gets notification by email about ticket creation (automatic)

STEPS:
1. Check your email boxes (both admin address and alternative address).
2. In your staff mailbox you should find a message sent from the user via RT together with ticket creation information at the beginning of the message
3. In users mailbox you should find a message sent from the user via RT

The Autoreply message is customized for every queue.

4.2.5 Staff member logs to the RT and take the ticket - become owner

STEPS:

1. Go to RT using ticket link received by email, and the ticket display screen will show up
2. Study the screen - check "People" and "More about the requestors" tab to see requestor and owners
3. If the ticket is not owned by anybody, take ownership (Basics tab, or top right menu Actions -> Take)
4. Don't confuse the owner of the ticket (under Basics tab) and the owner of the remainder (under Reminders tab)

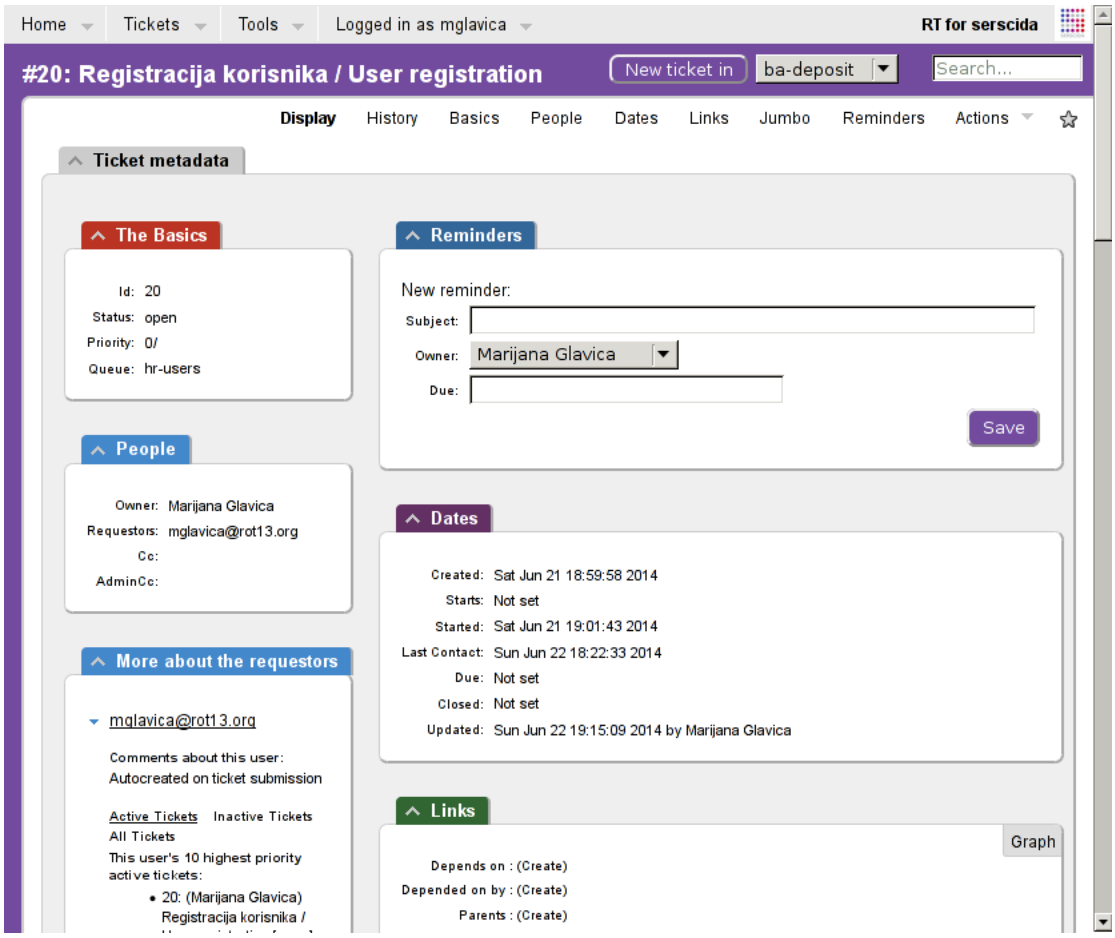

Figure 19: RT

4.2.6 Staff member (owner) replies to the user using RT - sends user agreement to the user

STEPS:
1. Go to ticket display screen using the link in your staff email
2. Choose Actions -> Reply from the top right menu, message editor will show up
3. Include article as template and change its content if necessary, or manually write a message
4. Optionally change the Subject of the message to "End user license agreement"
5. Optionally add addresses in cc and bcc (have in mind that staff receives every message, so no need to put them in cc or bcc)
6. Explore the rest of the screen (Scripts and Recipients and Ticket and Transaction)
7. Press "Update ticket"
8. Notice your reply message at the end of the screen and two messages below

4.2.7 The user (requestor) receives reply as email message (automatic)

STEPS:
1. Go to users mailbox and see the message

4.2.8 User sends signed agreement

STEPS:
1. From the user mailbox send reply to the message from the archive staff (received via RT), include user agreement in attachment
2. On ticket display screen you can see new tab "Attachments"

4.2.9 RT receives users email and adds it to the appropriate ticket (automatic)

STEPS:
1. Go to ticket display screen and look at the bottom of the message, notice the message

4.2.10 All staff is notified about new correspondence (automatic)

STEPS:
1. Go to your staff mailbox and see the notification about user correspondence

4.2.11 Staff member checks if the user have right to access data

STEPS:
1. Go to ticket display (follow the link from e-mail or find it directly on RT interface)
2. Open the agreement in attachment and check if everything is right

4.2.11.1 optional: if something isn't right, reply to user using RT

STEPS:
1. Go to ticket display screen
2. Choose Reply at the bottom of the screen right of user last message (to include users message in reply)
3. Write a message
4. Optionally change subject
5. Click "Update ticket"
6. Check history at the bottom of the screen

4.2.11.2 optional: staff member writes a comment related to ticket

Comments are not visible to users and they are useful for staff internal communication about the case.

STEPS:
1. Go to ticket display screen
2. Choose Actions -> Comment from the top right menu
3. Write a message (notice that editor is not red and that at the bottom of the screen the
4. Users address in not listed as receiver)

5. Click "Update ticket"
6. Check history at the bottom of the screen
7. Notice that staff member who wrote the comment didn't receive the notification (because he just sent the email, so he knows that some correspondence just happened), but other staff members did receive the notification by email

4.2.12 Management creates user in EPrints and Nesstar

4.2.12.1 Management creates the new user (Producer) in EPrints

STEPS:
1. Go to http://deposit.hr.serscida.eu/
2. Login with your username and password
3. Go to Admin -> System tools -> Create user
4. Enter new username and press create user
5. Choose user type (user) and click NEXT
6. Enter account details (email address and password), personal details (copy 7. information from registration form), click SAVE and RETURN
7. Summary data of registered user will appear
8. (You can always edit your user information)

Figure 20: Create user in EPrints

4.2.12.2. Management creates the same user (Consumer) in Nesstar

STEPS:
1. Login as admin: http://data.hr.serscida.eu:8080/admin/
2. Enter:
   - name,
   - username (@ sign is translated into *),
   - password (same as in EPrints),
   - role - authorizedUser for now
3. Click CREATE

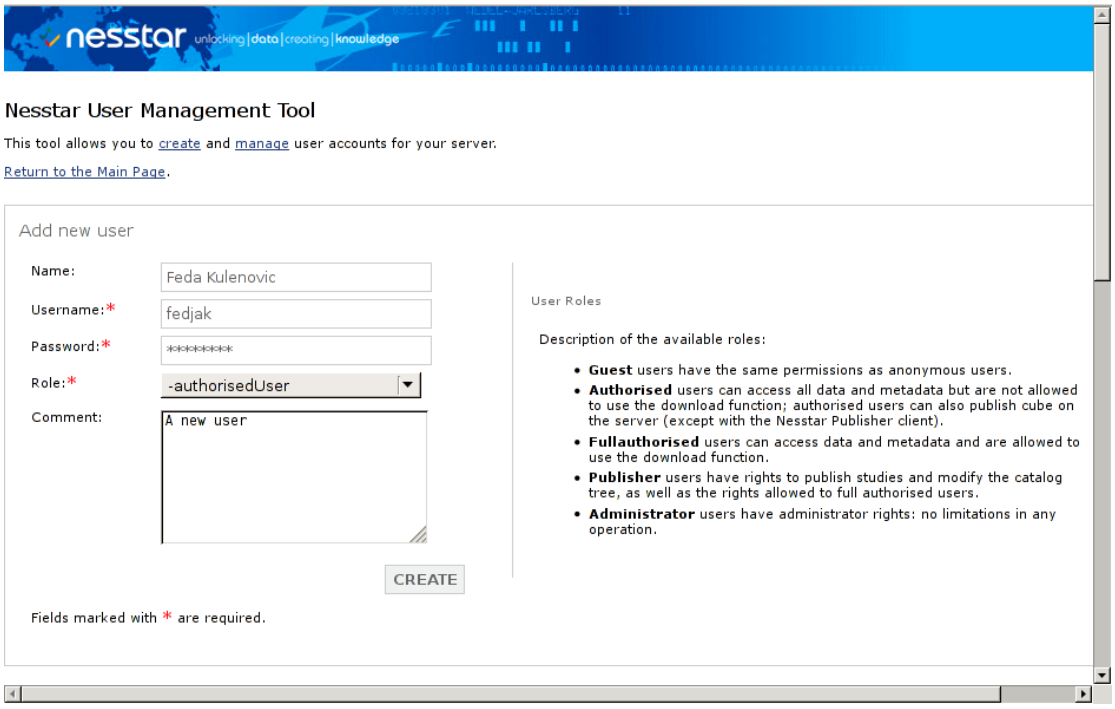

Figure 21: Create user in Nesstar WebView

4.2.13 Management sends username and password to the user via RT

STEPS:
1. Go to ticket display screen
2. Choose Reply at the bottom of the screen (right from the user message)
3. Include article (change if necessary) or manually write a message
4. Optionally change subject
5. Click "Update ticket"
6. Check history at the bottom of the screen

4.3 Creating metadata and publishing data to the web catalog using Nesstar Publisher

http://www.adp.fdv.uni-lj.si/serscida_wp4_lj2013_workshop/presentations/NesstarPublisher.pptx

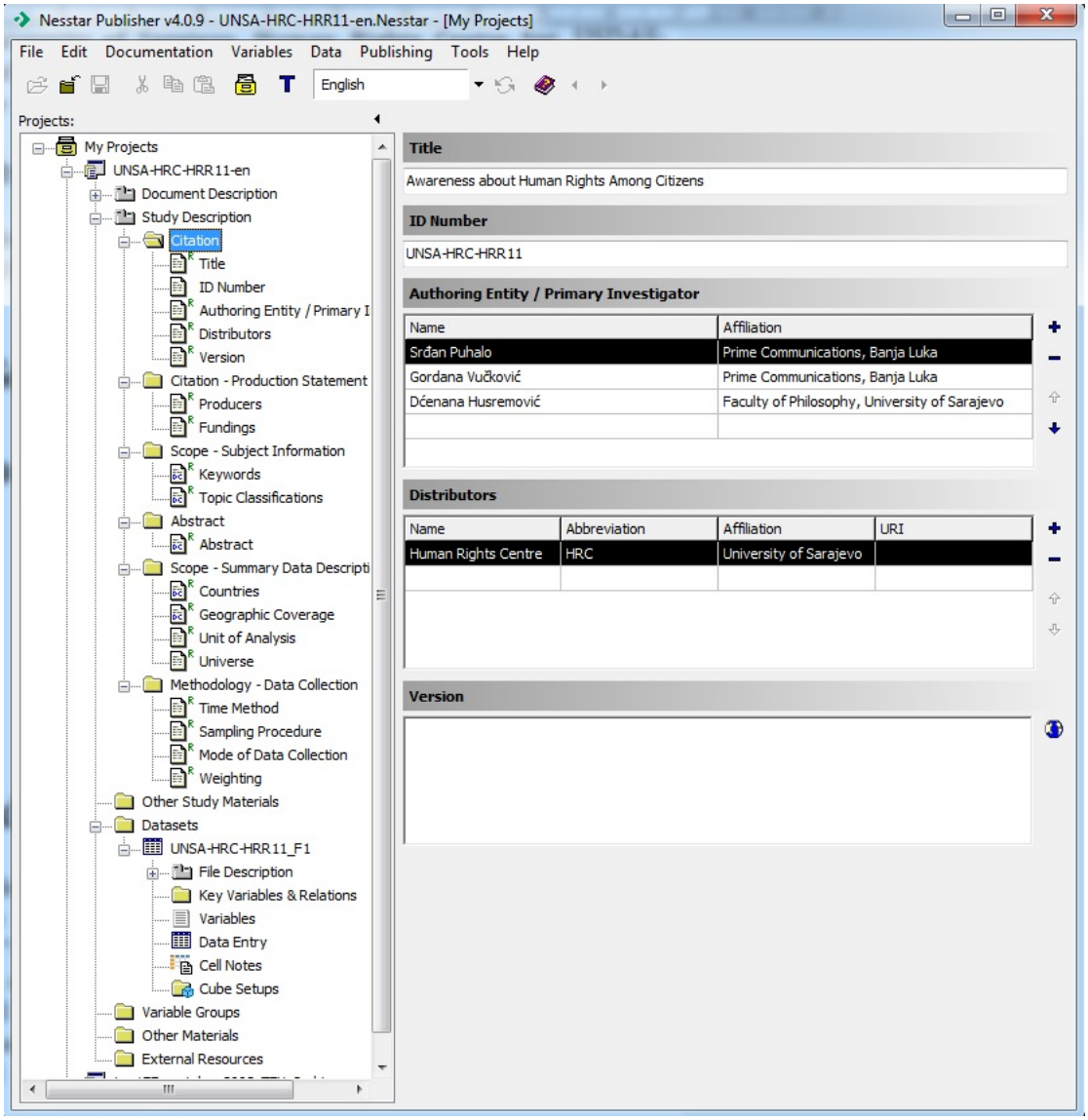

Figure 22: Nesstar Publisher

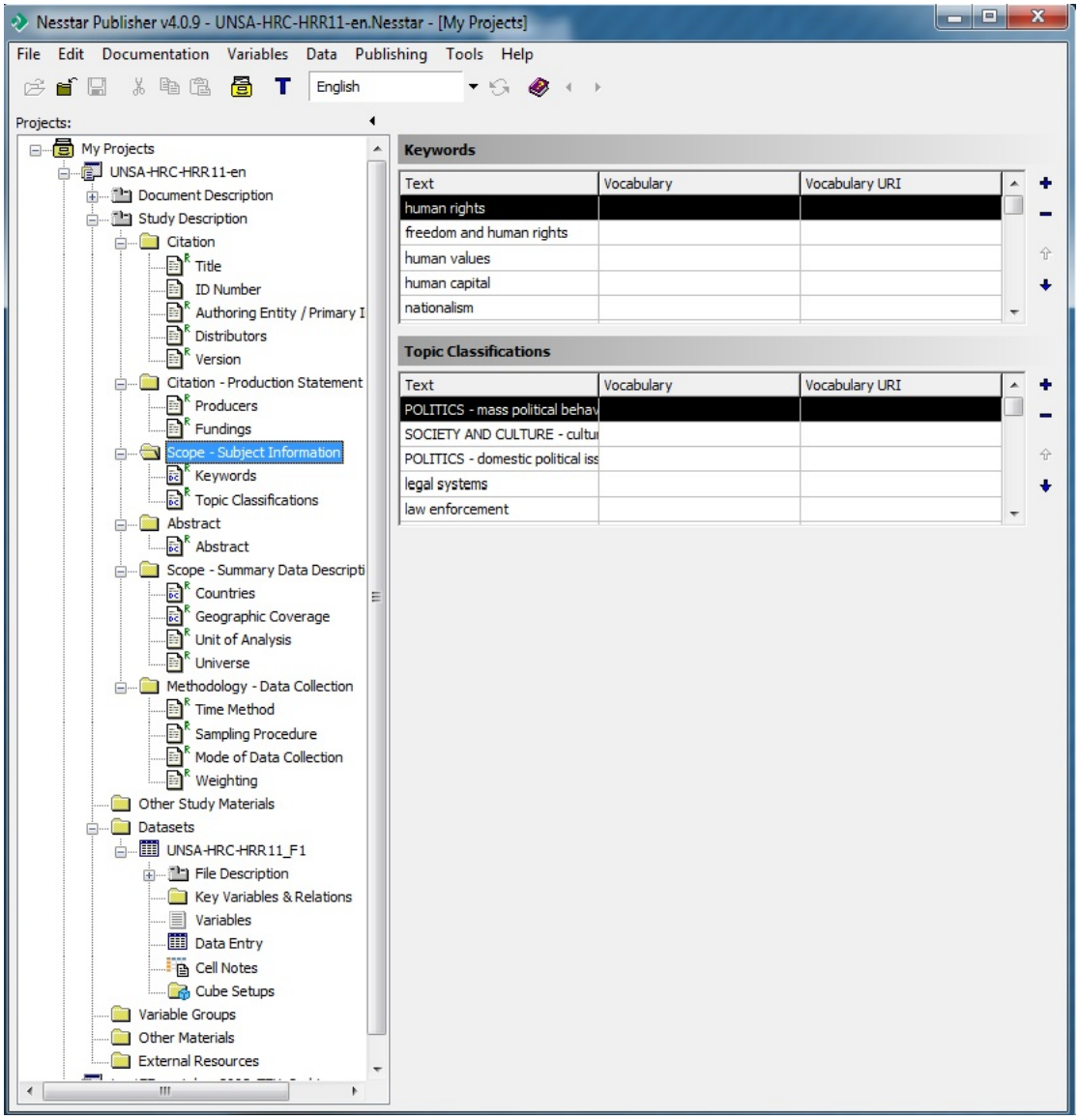

Figure 23: Nesstar Publisher

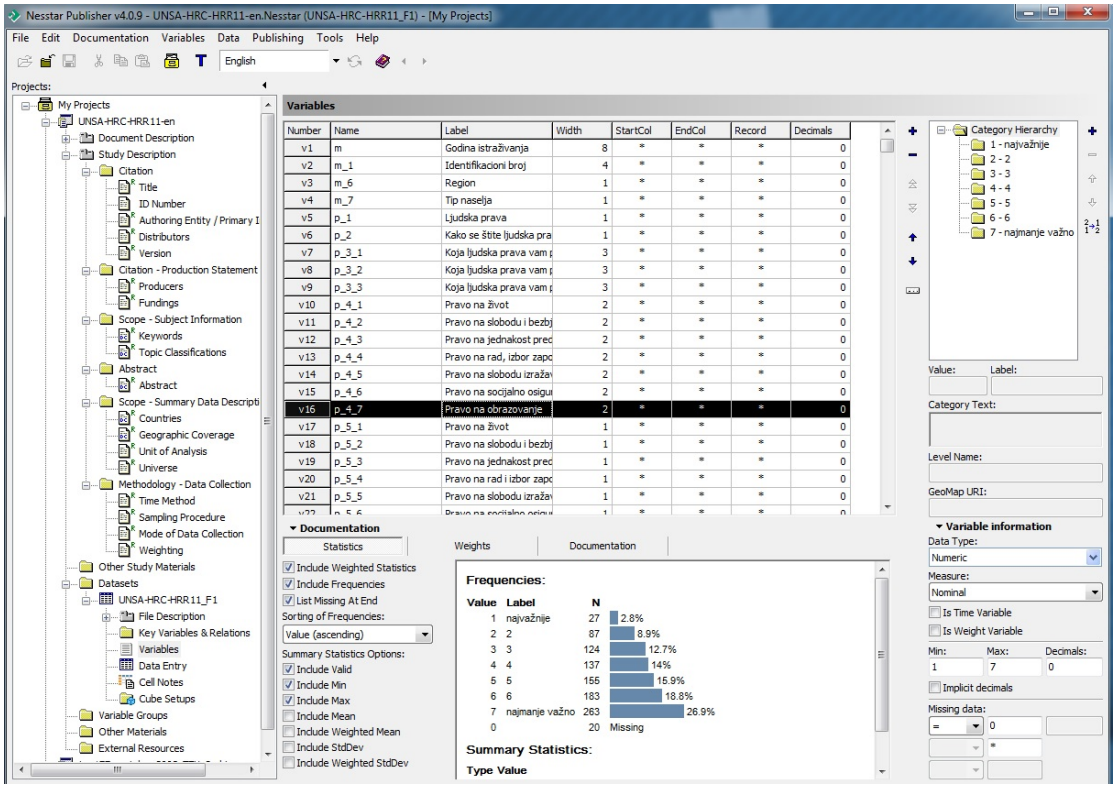

Figure 24: Nesstar Publisher

4.4 Accessing data in Nesstar catalog

To browse data in the Nesstar catalogs:

Bosnia & Herzegovina: http://data.ba.serscida.eu/
Croatia: http://data.hr.serscida.eu/
Serbia: http://data.rs.serscida.eu/

See also:

http://www.adp.fdv.uni-lj.si/media/gradiva/Vodic_po_Nesstarju.pdf

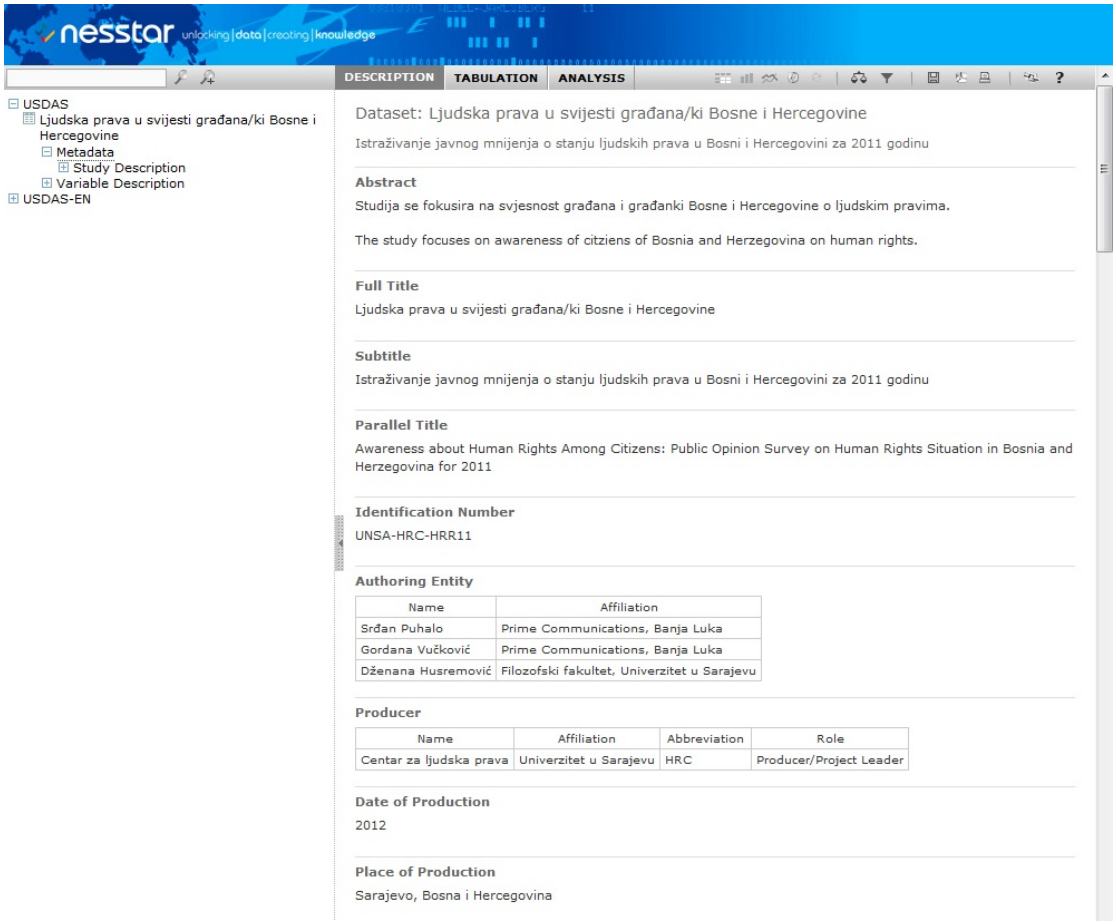

Figure 25: Nesstar WebView

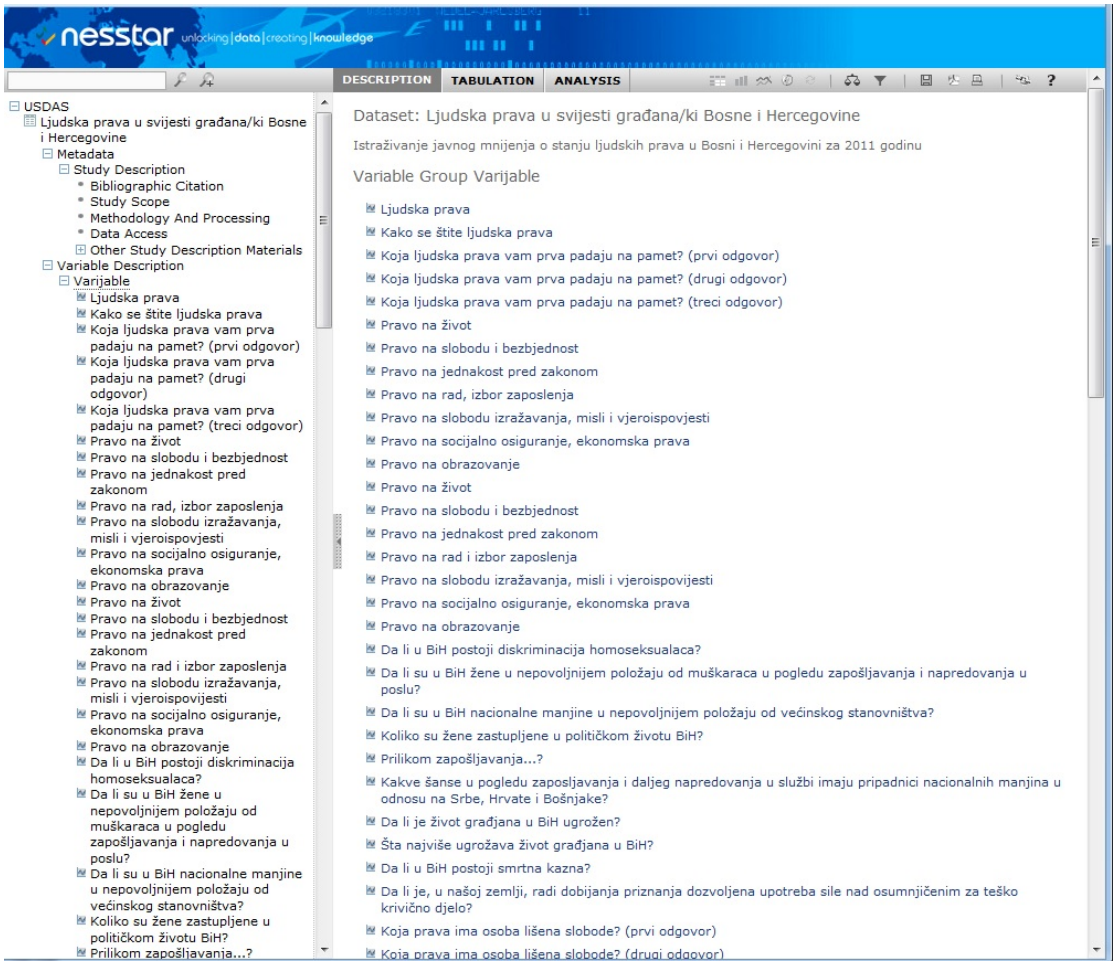

Figure 26: Nesstar WebView

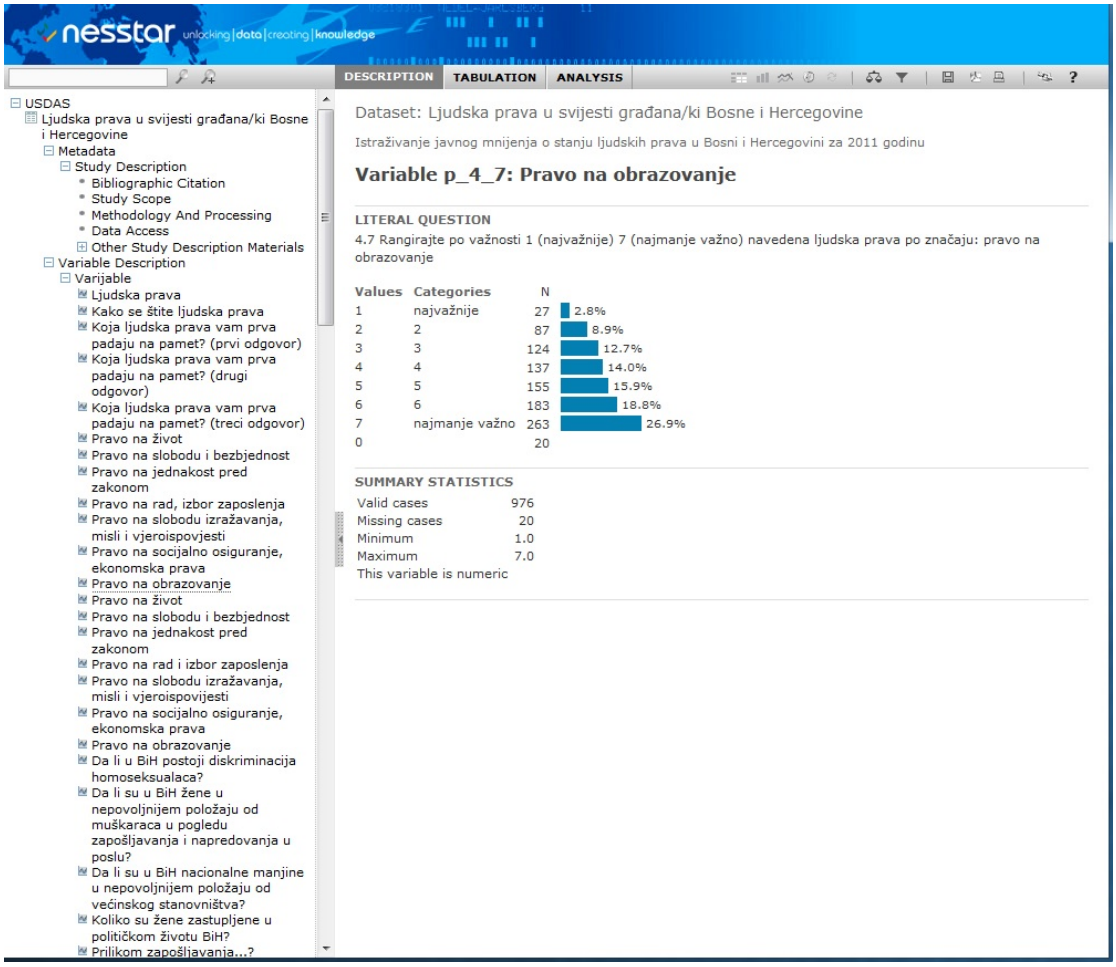

Figure 27: Nesstar WebView

5 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 in Croatia, Serbia, and Bosnia and Herzegovina 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 (for example, case-by case email or FTP distribution of some data, rather than direct downloading online).

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

6 Annex

6.1 Criteria for choosing archiving tool

Criteria for choosing data archiving tool:

1. **Learning Curve – Developers** – How long does it take to be productive?
2. **Learning Curve – Users**
3. **Degree of maturity** – Is the technology ready for industry-strength applications? Is it already fading out?
4. **Roadmap perspective, community** – What is the long-term perspective? Is there a strong and vivid developer and user community?
5. **Popularity** – How widespread is the adoption?
6. **Database Support** – What database systems are supported?
7. **Active maintenance and support** – How easy is it to find know-how? How comprehensive is the documentation?
8. **Local Know-How and expertise**
9. **OAIS compliant with combination of tools**
10. **DDI2.5**
11. **Supports OAI-PMH**
12. **Free / open source**
13. **Platform openness** – What platforms are supported (Linux, Windows, etc.)
14. **Interfaces and standards**
15. **Feedback from other institutions** (CESSDA members)
16. **Effort / benefit** of additional development
17. **Multilingual support**
18. **Reliability and robustness, integrity**
19. **Flexibility** – How easy is to customize/extend

6.2 SERSCIDA WP4 - Standards and Tools Table

	Name	Type	Description	Availability	URL
Standard	DataCite	DOI + metadata standard	A widely adopted service providing DOIs for data services. It also specifies a level of minimum metadata to create new DOIs	Priced subscription	http://www.datacite.org/
Standard	ISO27001	Information security standard	An international standard for information security, and badge of trust for any organisation hosting sensitive data	Cost for ISO document and assessment	http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=54534
Standard	DDI	Metadata: data description	DDI is the foremost metadata specification for social science data, and has a large and active internation community of users and developers	Free	http://www.ddialliance.org/
Standard	QuDEx	Metadata: qualitative data	Complementary to DDI (see above) this metadata standard aids in creating structured descriptions of complex collections, capturing for example, object relationships	Free	http://www.data-archive.ac.uk/create-manage/projects/qudex?index=1
Standard	Text Encoding Initiative (TEI)	Metadata: qualitative data	A metadata standard for marking up interview transcripts	Free	http://www.tei-c.org/index.xml
Standard	Statistical Data and Metadata eXchange (SDMX)	Metadata: statistical data	A metadata standard for the exchange of statistical information	Free	http://sdmx.org/

Standard	HASSET	Social science thesaurus	The UK Data Archive's HASSET thesaurus has been many years in development, and is among the most complete of its kind. Used for keyword indexing	Available from UKDA on request	http://data-archive.ac.uk/find/hasset-thesaurus/hasset-licence
Tool	Confluence	Collaborative editing	A Wiki-like platform for collaborative resource editing	Priced licencing	https://www.atlassian.com/software/confluence
Tool	SPSS	Data processing	SPSS, a statistical analysis package, has a scripting system that the UK Data Archive uses to handle certain parts of the Ingest process. These are currently being ported to Python	Priced licencing	
Tool	Refine	Data quality control	OpenRefine, the latest iteration of what was once called Google Refine, is a data manipulation tools that the UK Data Archive uses for data cleansing and quality control purposes. Is particularly useful for correcting erroneous text values in, for example, a variable containing place names.	Free	http://openrefine.org/
Tool	Renamer	Data quality control	App for batch renaming files, used primarily to ensure AIP compliance with archival storage requirements		
Tool	AxCrypt	Encryption	Encrpytion tools allowing for various types of encryption and secure file shredding after use	Free	http://www.axantum.com/axcrypt/
Tool	SafeHouse Explorer	Encryption	Allows creation of encrypted volumes including hard drive partitions and USB sticks	Free	http://www.safehousesoftware.com/SafeHouseExplorer.aspx

Tool	ZendTo	File transfer	Self-hosted 'Dropbox-like' software, used for secure file drop off including data deposits	Free	http://zend.to/
Tool	StatTransfer	Format conversion	StatTransfer is useful utility for moving between statistical data formats, and supports a very wide range of current software packages.	Free	http://stattransfer.com/
Tool	JIRA	Query tracking	The backbone of the UK Data Archive's services, JIRA enables tracking and reporting of issues from user access requests to the latest technical dev work	Priced licencing	https://www.atlassian.com/software/jira
Tool	NESSTAR	Storage and dissemination	NESSTAR is set of modular tools for disseminating tabular data and DDI metadata via a web interface, including capability for subsetting and analysis	Free	http://www.nesstar.com/
Tool	EPrints	Repository software	A repository platform, originally created for hosting open access publications. With the use of plugins (notably 'ReCollect') EPrints can also be used as a data repository.	Free	http://www.eprints.org/
Tool	CKAN	Repository software	Data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data.	Free	http://ckan.org/
Tool	DSpace	Repository software	DSpace preserves and enables easy and open access to all types of digital content.	Free	www.dspace.org/
Tool	NADA	Data catalog/publisher	NADA is a web-based cataloging system that serves as a portal for researchers to browse, search, compare, apply for	Free	http://www.ihsn.org/home/software/nada

			access, and download relevant census or survey information. NADA uses the Data Documentation Initiative (DDI), XML-based international metadata standard. NADA requires a server running PHP 5.2 or later with a MySQL database. Features: Instantly publish DDI documents to the web, Search surveys for rich metadata, Rich web interface to manage your DDI based codebooks, Distribute external resources and microdata with ease.		
Tool	Invenio	Repository software	Software suite enabling you to run your own digital library or document repository on the web. The technology offered by the software covers all aspects of digital library management from document ingestion through classification, indexing, and curation to dissemination.	Free	http://invenio-software.org
Tool	R	Data processing	Software environment for statistical computing and graphics	Free	http://www.r-project.org
Tool	Trac	Query tracking	Enables tracking and reporting of issues from user access requests to the latest technical dev work	Free	http://trac.edgewall.org
Standard	OAIS	Open Archival Information System	Framework for the understanding and increased awareness of archival concepts needed for long term digital information preservation and access	Free	http://www.dpconline.org/index.php?option=com_docman&task=doc_download&gid=347

Tool	Greenstone	Repository software	The basic unit to be stored to the system is a document, (the system supports various formats of digital objects), which is stored to previously prepared collections with defined information which documents they will contain and these collections then will be linked to the libraries. To process individual types of documents, the system has to contain programming support, the so-called plugins. It can be used to browse and search collections. The search of collections is based on categorising of similar documents to the same folder that can be hierarchically linked and in this way create a logical search structure. The system contains management of users and assigning of user rights.	Free	http://www.greenstone.org

6.3 End user Licences

University of Sarajevo - Data Archive and Services of Social Sciences
Human Rights Centre
Zmaja od Bosne 8, 71000 Sarajevo

END USER LICENCE	
Your details	
Title:	
First name:	
Last name:	
Address:	
Postcode and city:	
Country:	
Telephone no:	
Email address:	
Your institution	
Institution:	
Address:	
Postcode and city:	
Country:	
Position:	
Discipline:	
Required study details	
Study ID:	
Year of the study:	
Title of the study:	
Purpose of use:(scientific research, academic education or other)	
Please, specify purpose of use in more details:	

I am confirming that personal data submitted above is accurate: [X]

Terms of use

The data stored in the archives is available only to the users who specify their purpose of the use. Based on this agreement users are obliged to respect the professional ethical code and are further obliged, to fully cite the author and Archives, and pay for the commercial usage of the data. Each user is bound to warn USDAS for any errors discovered in the Data Collections. User will supply the USDAS with two copies of any published work based wholly or in part on the Data Collections or alternatively with references to said publications.

I declare that:

- I will use given dataset only one year after receiving it and will then destroy it,
- I will not hand over any copies of the data and accompanying material without written permission from archives
- I will use data only for specified purpose and will refrain myself from any attempts to identify individuals or households participated in the survey.

I agree to the above conditions and will fully respect them: [X]

To fulfil conditions for status of USDAS user, please send a scanned copy of completed and signed form on contract@ba.serscida.eu.

_____ _____

Date and place Signature

Croatian Data Archive for the Social Sciences
Library of the Faculty of Humanities and Social Sciences University of Zagreb
I. Lučića 3, HR-10000 Zagreb

END USER LICENCE	
Your details	
Title:	
First name:	
Last name:	
Address:	
Postcode and city:	
Country:	
Telephone no:	
Email address:	
Your institution	
Institution:	
Address:	
Postcode and city:	
Country:	
Position:	
Discipline:	
Required study details	
Study ID:	
Year of the study:	
Title of the study:	
Purpose of use:(scientific research, academic education or other)	
Please, specify purpose of use in more details:	

I am confirming that personal data submitted above is accurate: [X]

Terms of use

The data stored in the archives is available only to the users who specify their purpose of the use. Based on this agreement users are obliged to respect the professional ethical code and are further obliged, to fully cite the author and Archives, and pay for the commercial usage of the data. Each user is bound to warn CRODAS for any errors discovered in the Data Collections. User will supply the CRODAS with two copies of any published work based wholly or in part on the Data Collections or alternatively with references to said publications.

I declare that:

- I will use given dataset only one year after receiving it and will then destroy it,
- I will not hand over any copies of the data and accompanying material without written permission from archives
- I will use data only for specified purpose and will refrain myself from any attempts to identify individuals or households participated in the survey.

I agree to the above conditions and will fully respect them: [X]

To fulfil conditions for status of CRODAS user, please send a scanned copy of completed and signed form on contract@hr.serscida.eu.

_____ _____

Date and place Signature

Serbian Social Science Data Centre
Institute of Economic Sciences
Belgrade, Zmaj Jovina 12

END USER LICENCE	
Your details	
Title:	
First name:	
Last name:	
Address:	
Postcode and city:	
Country:	
Telephone no:	
Email address:	
Your institution	
Institution:	
Address:	
Postcode and city:	
Country:	
Position:	
Discipline:	
Required study details	
Study ID:	
Year of the study:	
Title of the study:	
Purpose of use:(scientific research, academic education or other)	
Please, specify purpose of use in more details:	

I am confirming that personal data submitted above is accurate: [X]

Terms of use

The data stored in the archives is available only to the users who specify their purpose of the use. Based on this agreement users are obliged to respect the professional ethical code and are further obliged, to fully cite the author and Archives, and pay for the commercial usage of the data. Each user is bound to warn SR-SSDC for any errors discovered in the Data Collections. User will supply the SR-SSDC with two copies of any published work based wholly or in part on the Data Collections or alternatively with references to said publications.

I declare that:

- I will use given dataset only one year after receiving it and will then destroy it,
- I will not hand over any copies of the data and accompanying material without written permission from archives
- I will use data only for specified purpose and will refrain myself from any attempts to identify individuals or households participated in the survey.

I agree to the above conditions and will fully respect them: [X]

To fulfil conditions for status of SR-SSDC user, please send a scanned copy of completed and signed form on e-mailsr.ssdc@ien.bg.ac.rs.

_____ _____

Date and place Signature

6.4 Hardware

6.4.1 CRODAS

4 x Server
DELL PowerEdge R720
Processor: 2x Intel Xeon E5-2620v2, 2, 1GHz, 6 cores, 15MB smart Chace memory, 7, 2GT/s QPI; Chipset: Intel C602; Memory: 96GB, 1866MHz, Dual Rank DDR3, max. mem. 768GB, 24 slots; Hard disk: 8x1TB SATA, 7200 RPM, 6Gb/s, 3,5", Hot Plug

1 x network switch
Dell Networking 6248 Switch

2 x network transceiver
Routerboard SFP+ S+31DLC10D

4 x Workstation
DELL Precision T3610
Processor: Intel Xeon processor E5-1607v2, 4 cores, 10 MB Cache memory, frequency 3,0GHz; Chipset: Intel C602; Memory: 32:GB (4x8) 1600MHz DDR3 Non-ECC memory, max mem. 128 GB; Hard disk: 256GB SSD + 3x 900GB, SAS, 10000 RPM

8 x Monitor
Dell P2414H
24" LCD LED, 16:9, resolution 1920x1080

4 x Loudspekers
Dell AC511

4 x Webcam
Logitech C270HD

1 x multifucntional device
Canon MF 4870dn

1 x laptop
Lenovo Yoga 2
Processor: Intel® Core™ i3 – 4010U, freq. 1,70GHz, 3MB cache memory; Hard disk: 500GB SATA + 16GB SSD; Memory: 4GB DDR3 1600Hz; Screen: 13.3" IPS, 1600x900, touchscreen

1 x projection canvas

6.4.2 SR-SSDC

1 x Server
DELL PowerEdge R720
Processor: 2x Intel Xeon E5-2620v2, 2, 1GHz, 6 cores, 15MB smart Chace memory, 7, 2GT/s QPI; Chipset: Intel C602; Memory: 96GB, 1866MHz, Dual Rank DDR3, max. mem.

2 x Workstation

DELL Precision T3610
Processor: Intel Xeon processor E5-1607v2, 4 cores, 10 MB Cache memory, frequency 3,0GHz; Chipset: Intel C602; Memory: 32:GB (4x8) 1600MHz DDR3 Non-ECC memory, max mem. 128 GB; Hard disk: 256GB SSD + 3x 900GB, SAS, 10000 RPM

2 x Monitor
Dell P2414H
24" LCD LED, 16:9, resolution 1920x1080

2 x Notebook
Acer Aspire S7-391-53314G12aws
Intel Core i5-3317U 1.7GHz (dual-core, 3MB cache); 4GB (1x4GB) DDR3; 128GB SSD; Intel HD 4000; 13" TFT LCD FullHD 1920x1080; Wi-Fi 802.11 a/b/g/n, Bluetooth 4.0; 2x USB 3.0, HDMI
0.3Mpix, with microphone; SD and MMC card reader; Windows 8 (64-bit)

1 x Network switch
Ports - 16 x 10/100/1000, Uplink - 2x SFP, Switch Capacity (Gb/s) – 32, Forwarding rate (Mpps) – 20, Jumbo frames (Bytes) – 9216, Spanning Tree - STP (802.1D), RSTP (802.1w), MSTP (802.1s), QoS L2/L3-aware, Port security - 802.1x: Supports single and multiple host access, MAC-Based port security

6.5 Software

6.5.1 List of software used at SND

- ***Office suite:***
  - MS Office Suite 2010, LibreOffice
- ***Mail/calendar client:***
  - Outlook, Thunderbird
- ***XMPP chat client:***
  - Pidgin, Trillian (Windows)
  - Adium, iChat (Macintosh)
- ***Web browser:***
  - Mozilla Firefox, Chrome
- SIMS aka SND Internal Management System
- Self-developed ”specialist” software
- NESSTAR, STATA, SPSS, SAS
- Notepad++
- Visual Studio, oXygen XML Editor
- SVN/Git
- Adobe C6 Design Standard, Photoshop,
- Acrobat, Dreamweaver

6.5.2 List of software used at ADP

- ***Office suite:***
  - Windows 7 Enterprise x64
  - MS Office Suite 2010
- ***Mail/calendar client:***
  - Outlook, Pegasus, Google Calendar
- ***Chat client:***
  - Skype

- ***Webbrowser:***
  - Mozilla Firefox, Chrome, IE
- Adobe Profesionnal
- Mendeley, Endnote Web
- NESSTAR, STATA, SPSS, SAS
- oXygen XML Editor
- Tortoise SVN, OwnCloud
- Jira
- Gimp, Inkscape

6.5.3 List of software used at FORS

- ***Common***
  - Windows 7 Enterprise x64
  - Office 2010
  - Libre/Openoffice
  - Firefox
  - Internet Explorer 9
  - Adobe Acrobat Pro 10
- ***Scientific***
  - SPSS 20, Stat-Transfer 11 (10) 5 > 15, Stata 12, SAS
  - MLwin 2.25 (2.02) 2 > 7
  - Endnote
  - NESSTAR

7 References

CESSDA Catalogue: http://www.cessda.net/catalogue/

CESSDA Topic Classification: http://www.cessda.net/catalogue/ (see under "Browse by topic")

Data Documentation Initiative, Community: http://www.ddialliance.org/community

Data Documentation Initiative, Controlled Vocabularies: http://www.ddialliance.org/controlled-vocabularies

Data Documentation Initiative, DDI Codebook 2.1: http://www.ddialliance.org/Specification/DDI-Codebook/2.1

Data Documentation Initiative, DDI Tools: http://www.ddialliance.org/resources/tools

Data Documentation Initiative (2001), Making XML files for the new IDC of CESSDA: http://www.ddialliance.org/sites/default/files/cessda-rec.pdf

ELLST: http://ukdataservice.ac.uk/about-us/projects/cessda-elsst/details.aspx

EPrints: http://www.eprints.org/software/

FileSender: https://www.assembla.com/spaces/file_sender/wiki

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

Lavoie, B.F. (2004) The Open Archival Information System Reference Model: Introductory Guide: www.dpconline.org/docs/lavoie_OAIS.pdf

Nesstar, Nesstar Publisher: http://www.nesstar.com/software/publisher.html

Nesstar, Nesstar Server: http://www.nesstar.com/software/server.html
Nesstar, Nesstar Administrator Guide Version 4.0: http://www.nesstar.com/help/4.0/server/index.html

Nesstar, Nesstar WebView User Guide Version 4.0: http://nesstar.com/help/4.0/webview/index.html

Nesstar, Nesstar WebView User Guide Version 3.0: http://www.nesstar.org/Release-free/WebView_Userguide_v3.pdf

Nesstar, Nesstar Server Components: http://www.nesstar.com/help/4.0/server/knowing_server/components.html

Nesstar, Introduction to Access Control Unit: http://www.nesstar.com/help/4.0/server/access-control-unit/acu-introduction.html

Nesstar, Understanding the Nesstar ACU Policies: http://www.nesstar.com/help/4.0/server/access-control-unit/acu-preloaded-policies

Owncloud: http://www.owncloud.org

Pavlinusic, D. and Blaskovic, L. (2014) Build your own cloud using ganeti (kvm, drbd), salt and zfs (Presented at the Croation Linux Users' Convention DORS/CLUC 2014, 16-18 June 2014, Zagreb, Croatia) http://bit.ly/dc14-ganeti

RT Request Tracker: https://www.bestpractical.com/rt/

Reference Model for an Open Archival Information System, OAIS (2012); CCSDS 650.0-M-2; Consultative Committee for Space Data Systems: Washington http://public.ccsds.org/publications/archive/650x0m2.pdf

SERSCIDA WP4 Workshop (2013) 26-27 September 2013, Ljubljana http://www.adp.fdv.uni-lj.si/serscida_wp4_lj2013_workshop/presentations.html

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: http://www.dptp.org/wp-content/uploads/2010/08/UKDAp90.pdf

Technical Registry PRONOM: http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

Ward, M. (2006) Online access to repositories of survey data: Nesstar, (Presented at the 'Best practices in sharing data' conference, 17 March 2006, Regent's college, London www.esds.ac.uk/news/eventsdocs/bestmwmar06.ppt