DELIVERABLE

Project Acronym: E-ARK

Grant Agreement Number: 620998

Project Title: European Archival Records and Knowledge Preservation

DELIVERABLE DETAILS

DELIVERABLE REFERENCE NO.	**D3.4**
DELIVERABLE TITLE	Records export, transfer and ingest recommendations and SIP Creation Tools
REVISION	**1.0**

AUTHOR(S)	
Name(s)	Organisation(s)
Aadi Kaljuvee	National Archives of Estonia (NAE)
Alex Thirifays	Danish National Archives (DNA) // Magenta ApS (MAG)
Angela Dappert	University of Portsmouth (UPHEC)
Björn Skog	ES Solutions (ESS)
Boris Domajnko (MC)	Slovenian National Archives (SNA)
Jernej Križaj	Slovenian National Archives (SNA)
Jože Škofljanec	Slovenian National Archives (SNA)
Miguel Ferreira	KEEP SOLUTIONS (KEEPS)
Tarvo Kärberg	National Archives of Estonia (NAE)
Thor Dekov Buur	Magenta ApS
Torben Lauritzen	Magenta ApS

REVIEWER(S)	
Name(s)	Organisation(s)
Kuldar Aas	National Archives of Estonia (NAE)
David Anderson	University of Brighton
Miguel Ferreira	KEEP SOLUTIONS

Project co-funded by the European Commission within the ICT Policy Support Programme Dissemination Level		
P	**Public**	X
C	**Confidential, only for members of the Consortium and the Commission Services**	

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Submitted Revisions History

Revision No.	Date	Authors(s)	Organisation	Description
0.1	02.02.2015	Tarvo Kärberg	NAE	First version.
0.2	03.02.2015	Tarvo Kärberg	NAE	Some minor updates.
0.3	09.02.2015	Tarvo Kärberg	NAE	Updating.
0.4	13.03.2015	Tarvo Kärberg	NAE	Updating based on feedback.
0.5	27.03.2015	Aadi Kaljuvee	NAE	Commenting and adding diagrams
0.6	01.04.2015	Tarvo Kärberg	NAE	Adding diagrams.
0.7	14.04.2015	Jože Škofljanec Jernej Križaj	MC	Commenting and updating.
0.72	24.04.2015	Angela Dappert	UPHEC	Commenting and updating.
0.73	06.05.2015	Tarvo Kärberg	NAE	Commenting and updating.
0.74	29.05.2015	Tarvo Kärberg	NAE	Updating the pre-ingest.
0.75	05.06.2015	Tarvo Kärberg	NAE	Updating the ingest part.
0.8	05.10.2015	Tarvo Kärberg	NAE	Integrating feedback from Angela Dappert and Björn Skog.
0.9	17.11.2015	Boris Domajnko		Commenting and updating the use cases.
0.91	20.11.2015	Tarvo Kärberg	NAE	Minor updates.
0.92	29.01.2016	Tarvo Kärberg	NAE	Updating references to documents and diagrams according to previous agreements.
0.93	25.02.2016	David Anderson	University of Brighton	Quality assurance and proof-reading.
0.94	19.12.2016	Tarvo Kärberg	NAE	Changing the template of the document, rearranging the chapters, adding information about tools.
0.95	21.12.2016	Miguel Ferreira	KEEPS	Reviewing and updating the content.
0.96	30.12.2016	Björn Skog	ESS	Reviewing and updating the content.
0.97	30.12.2016	Alex Thirifays	MAG	Reviewing and updating the content.
0.98	18.01.2016	Kuldar Aas	NAE	Technical review and quality assurance.
0.99	30.01.2017	David Anderson	University of Brighton	Quality assurance and proof-reading.
1.0	30.01.2017	Tarvo Kärberg	NAE	Final version of D3.4

Statement of originality:
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

Figures

1. Executive summary

This document describes a software deliverable as it delivers a number of E-ARK tools:

- ERMS Export Module (a tool for exporting records and their metadata from ERMS in a controlled manner);
- Database Preservation Toolkit (a tool for exporting relational databases as SIARD 2.0 or other formats);
- ESSArch Tools for Producer (a tool for SIP creation);
- ESSArch Tools for Archive (a tool for SIP ingestion);[1]
- RODA-in (a tool for SIP creation);
- Universal Archiving Module (a tool for SIP creation).

In addition, an overview of Pre-Ingest and Ingest processes will be provided by this document which will help to understand the tools and their use.

The document is divided into the following chapters:

- **Chapter 2 Introduction** explains the work from the E-ARK Pre-Ingest and Ingest processes perspective.
- **Chapter 3 Tools overview** provides a technical overview of software components and information flow, displaying the main building blocks of the Pre-Ingest and Ingest entities.
- **Chapter 4 Glossary** lists definitions, acronyms, and abbreviations.

---

[1] According to the DoW the initial plan was to develop a single tool – ESSArch Tools – to cover both the creation and ingest. However, in the course of the project it was decided to split it into two separate tools: the ESSArch Tools for Producer; and ESSArch Tools for Archive.

2. Introduction

The purpose of this document is to describe the E-ARK workflow for Pre-Ingest and Ingest together with related software tools. This document is based on existing work, and all the process steps introduced in this document have been identified using both a bottom-up and a top-down approach.

The bottom-up approach consisted of gathering relevant requirements by analysing best practices and requirements from the pilot sites, E-ARK partners, external bodies and their tools. This work started early in the project with the deliverables D2.1 E-ARK General Pilot Model and Use Case Definition,[2] D3.1 E-ARK Report on Available Best Practices[3] and continued to be improved throughout the project in order to support the development.

The top-down approach consisted of identifying the generic steps in Pre-Ingest and Ingest processes and creating high-level illustrations and descriptions from these.

The two approaches are complementary, and were adopted to secure that all requirements were taken into account and that the communication between all relevant parties would run smoothly.

As a subsequent action, a set of E-ARK tools (Chapter 3) were developed/updated in order to provide support for the E-ARK Pre-Ingest and Ingest processes described below.

Pre-Ingest

The high-level illustration of the E-ARK Pre-Ingest process (Figure 1) consists of a number of steps (see also the General Model for E-ARK):[4]
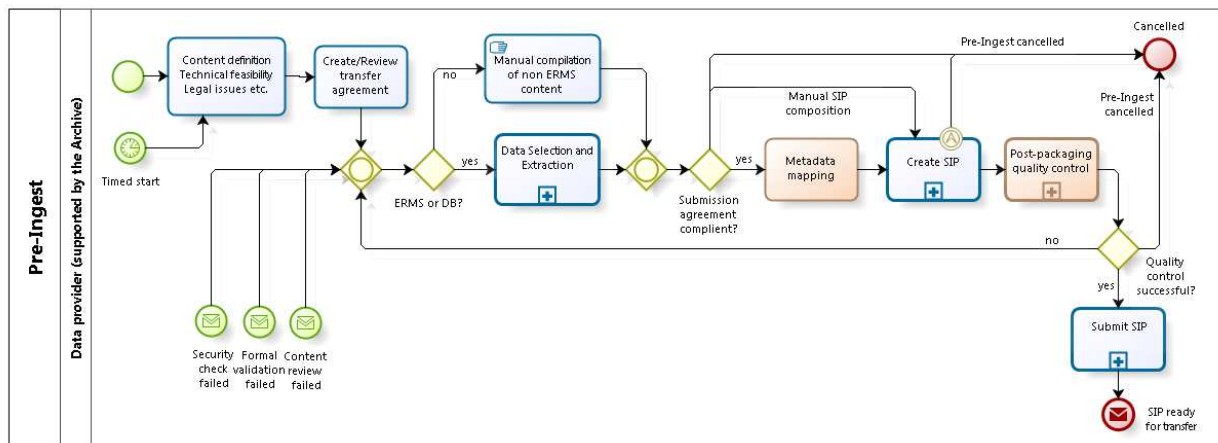
Figure 1: E-ARK Pre-Ingest (High-level process)

Content definition, Technical feasibility, Legal issues – Defining the scope of the transfer

In general, this step represents the process of defining the content for archiving. It could consist of

- defining contact persons and responsibilities;

- defining content (including file formats) to be transferred;

- defining metadata (to be transferred);

- estimating the size of the SIP/transfer;

- checking access and copyright restrictions;

- estimating and setting a timeline for the transfer;

Although it may result in a written transfer agreement (see next step), it is usually a manual process and is therefore unsupported by the E-ARK tools.

Create/Review transfer agreement

The information gathered in the previous step might be formalised as a written transfer agreement. If transfer agreement(s) exist from previous transfers, it might be reviewed and updated in order to describe the current transfer. If no previous transfer agreements exist, it can be created based on the agreement made in the previous step.

This step is usually a manual process and not mandatory and as such, is not supported by the E-ARK tools introduced in this document. However, the E-ARK project has provided a possible structure for the transfer agreement (see the E-ARK SIP Pilot Specification[5] for further details).

Data selection and extraction

Data selection is the process of identifying what content and how should be extracted from the IT system.

In the case of an ERMS, data selection can be implemented either inside an ERMS, as a piece of native functionality, or in a separate "archiving module". Either way, the ERMS data selection component usually includes functionality for defining criteria for data extraction and marking the chosen data (or just metadata, or data with any relevant metadata) for extraction. The data selection part could be built around queries which use criteria like retention periods and appraisal decisions; time passed since the creation of data etc. For example: "Find me all records which have archival value and have been created more than 8 years ago".

In case of databases, data selection can be rather difficult as the task of identifying a reasonable subset of meaningful data across an (often) complicated data model is not straightforward. As such, in most cases native DBMS functionality for selecting and creating the "archivable subset" must be preferred for data selection.

However, one pragmatic case of data selection, often used in practice, is just to limit the data selection to the level of whole tables (i.e. define only which tables need to be archived, no selection done on lower column and row levels). This is a functionality supported also by some external data export tools like the Database Preservation Toolkit.[6]

---

[5] D3.3 E-ARK SIP Pilot Specification, 2016, http://eark-project.com/resources/project-deliverables/51-d33pilotspec/file
[6] http://keeps.github.io/db-preservation-toolkit/

After selection, we need to extract the data from the producer system. There are multiple possibilities (pull vs push; automatically triggered vs manually initiated) for doing that. How exactly it is done depends on each individual case. Note that E-ARK offers only tools which execute the "pull" method manually (see ERMS Export Module on page 13 and Database Preservation Toolkit on page 19).

Manual compilation of non-ERMS (electronic records management systems) content
Manual compilation includes everything involved with the process of identifying what content and how should be collected for the SIP creation process without using any specific tool for that purpose. This task is supported by E-ARK tools UAM (see page 25), ETP (see page 22) and RODA-in (see page 24) by providing means for including computer files and folders into SIPs.

SIP Creation (Metadata mapping, Create SIP)
The SIP can be created

- either by the extract data function directly;
- or in a stand-alone tool (SIP creation tool) which uses the data/metadata extracted or selected/created manually in the previous step.

In case the original metadata is not in a format suitable for the receiving repository then the "Create SIP" step must be preceded by a "Metadata Mapping" step in which a mapping is created between the original metadata format and the format expected by the repository. There is no separate "pre-packaging quality control" step, because all quality controls are included in the "Create SIP" step. This task is supported by E-ARK tools: UAM, ETP and RODA-in.

Post-packaging quality control
Validation and quality control should be carried out to confirm that the SIP which has been created by the producer includes exclusively content that is intended to be exported, and that the content and its metadata are complete, well-formed and valid.

A minimal validation and quality assurance process requires only that a SIP conforms to an XML schema, but may optionally include the checks for:

- data integrity;
- information package structure validity;
- data transfer conformance to the requirements of the submission agreement;
- inclusion of file format specifications;
- file format validity;
- metadata completeness;
- metadata conformity to the requirements of the submission agreement;
- the absence of viruses, worms, Trojans, etc.

It should be noted that our recommendation that validation and quality control should exist on both sides – the producers should check whether the SIP they created is valid, and the archives should double check the SIP during the Ingest step.

All E-ARK SIP Creation tools provide at least minimal required validation during the SIP creation process.

Submit SIP

Once the SIP has been created it must be submitted to the repository. There are many methods (offline (including a series of different media), secure network transfer protocols) available for transfer. Most E-ARK SIP Creation tools allow users to simply save SIPs on their hard drives and use any transfer method or protocol available to the users. However, some tools (most prominently UAM) provide also options to transfer SIPs over secure nationally established networks.

Please also note, that E-ARK tools do not currently offer a generic solution for ensuring the integrity of the whole transfer in case multiple SIPs have been included into a transfer. As such, it is at the discretion of each user or tool developer to implement transfer integrity assurance methods most suitable in their environments.

Ingest

The high-level illustration of the E-ARK Ingest process consists of a number of steps (Figure 2):
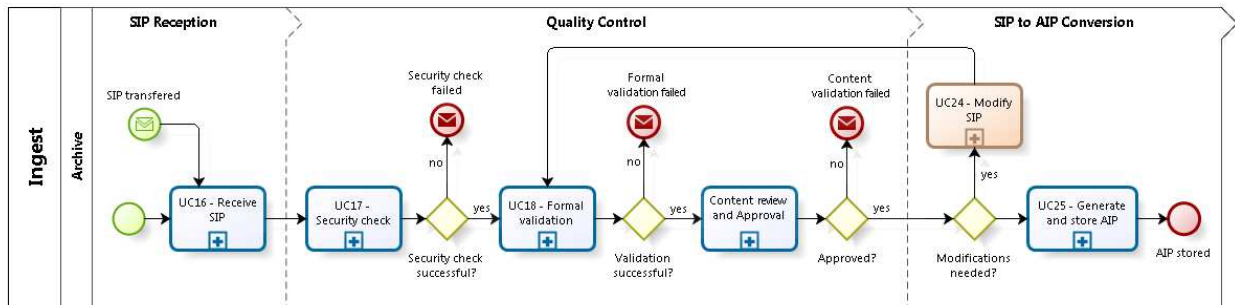
Figure 2: E-ARK Ingest (High-level process)

SIP Reception

The process by which an SIP arrives at an agreed location is outside the scope of the E-ARK project. Therefore, for the purposes of the present document we assume simply that a SIP has been received and is saved to the appropriate processing area.

Quality control

SIP quality[7] assurance includes both manual and automated processes. It consists of:

1. **Security checking** – this includes the various checks to which SIPs are exposed after being received, to ensure they are free of malware etc, and have valid technical integrity

2. **Formal validation** – the process of confirming that SIPs meet the quality criteria stipulated under "post-packaging quality control".

3. **Content review and approval** – a series of checks which result in approval for further processing, or a request to the producer to modify and re-submit the SIP.[8]

The quality control is similar to the Pre-Ingest, but depending on the available resources and needs, archives may apply additional quality checks and even perform some manual validation next to the automated controls.

---

[7] The quality assurance tasks done during the Pre-Ingest are repeated in the Ingest process as well in order to guarantee maximum quality.

[8] Any necessary modifications which are the result of archival repository issues or policies may, optionally, be carried out during the "Modify SIP" stage.

3. Description of tools and tool documentation

Figure 3 below shows a diagram with the components used to select, extract/export the information from a producer system, prepare and create the SIP. The diagram shows the main components:

- Content and Records Management Systems (e.g. Alfresco platform);
- Relational Databases (e.g. PostgreSQL);
- Simple file-system based records (e.g. computer files organised in folders);
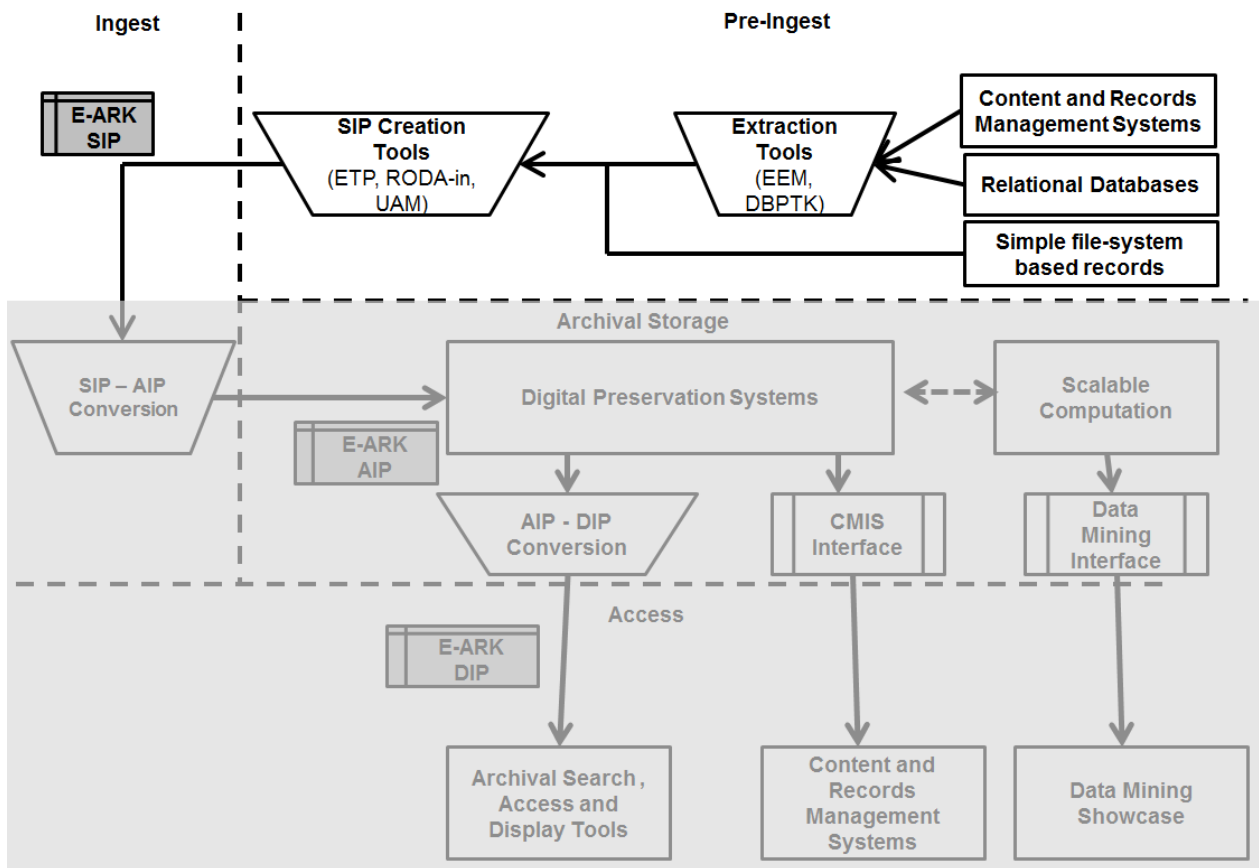- Extraction Tools (see page 13);
- SIP Creation Tools (see page 22).

Figure 3: Overview about tools and Pre-Ingest areas

E-ARK Extraction Tools allow the content to be exported using the pull method and imported into one of the SIP Creation Tools, where the content can be prepared (e.g. edit metadata, etc.) and then a SIP will be created which will be transferred to the repository as the input to the SIP to AIP conversion component.[9]

---

[9] Please note that the SIP to AIP conversion component or any other component in the remaining workflow is not in the scope of this document. Please refer to D4.4 and D5.4 for detailed descriptions of E-ARK ingest and access components!

Producer System Extraction Tool

The Producer System Extraction Tool is the system where all the information is located before creating a SIP for archiving purposes. The format of the data and metadata that is extracted from the producer system does not necessary have to be represented using the full E-ARK SIP as not all required information is available at the point of extraction. The export format will be an intermediate format (e.g. SIARD 2.0 in case of relational databases). The general workflow about export and import is illustrated in Figure 4.
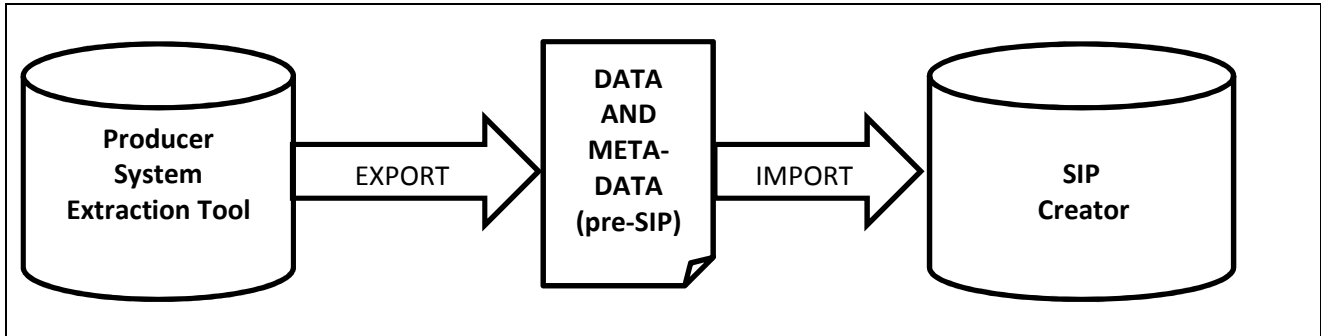
Figure 4: Relations between extraction and import

Work Package 3 delivered 2 tools that are meant for export from a producer system:

1. **ERMS Export Module (EEM)**
   The EEM helps to export records and their metadata from ERMS in a controlled manner. The module is composed of a backend and a frontend. The backend is a RESTful service (running on a Tomcat server) written in Java using the framework Jersey,[10] and the frontend is a web application (running on an Apache HTTP Server) written in AngularJS. The EEM is capable of connecting to and exporting content from any repository that supports the CMIS (Content Management Interoperability Services) protocol[11] (CMIS version >= 1.0 is supported) as seen in Figure 5.
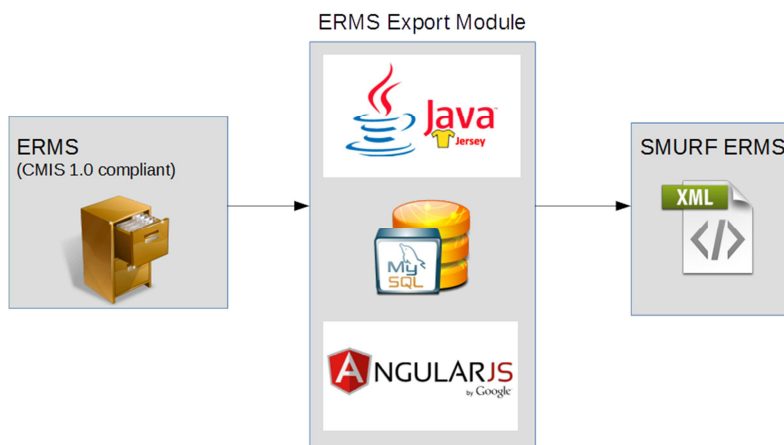

Figure 5: ERMS Export Module

---

[10] Jersey RESTful Web Services framework is open source, production quality, framework for developing RESTful Web Services in Java that provides support for JAX-RS APIs and serves as a JAX-RS (JSR 311 & JSR 339) Reference Implementation, https://jersey.java.net/.

[11] Content Management Interoperability Services (CMIS) Version 1.1, 2015, http://docs.oasis-open.org/cmis/CMIS/v1.1/CMIS-v1.1.html

As such, the module is not only applicable with the Alfresco content management system as envisaged in the Description of Work, but can be used by any other system that supports CMIS 1.0 (or higher) standard like SharePoint 2013,[12] OpenText,[13] eXo Platform,[14] Nuxeo,[15] IBM FileNet Content Manager,[16] OpenKM[17] etc.

The documentation for the ERMS Export Module is available at https://github.com/eark-project/erms-export-ui-module/tree/master

The source code for the backend is available at https://github.com/magenta-aps/E-Ark-Alfresco-export-bridge and for the frontend at https://github.com/magenta-aps/erms-export-ui-module.

The main features are the following:

- Repository profile creation and manipulation. In order to create a profile for the connection the following details must be provided (Figure 6):
  - Name – name for the profile (e.g. Alfresco CMS);
  - URL – CMIS endpoint for the repository;
  - ID – username for the connection to the repository;
  - Password – password for the connection to the repository.
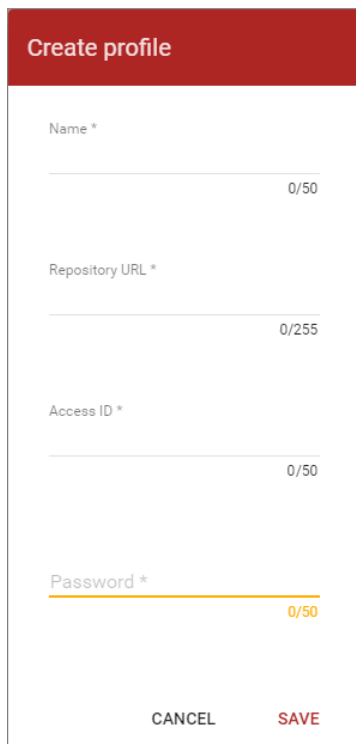

Figure 6: Profile creation

It is also possible to edit the details for the connection or delete the profile (Figure 7).
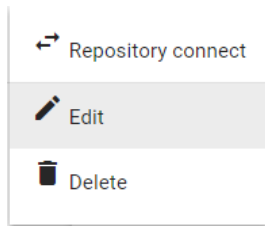

Figure 7: Edit or delete profile

If clicked on the "Edit" menu then the connection profile metadata will be reopened for editing (Figure 8).
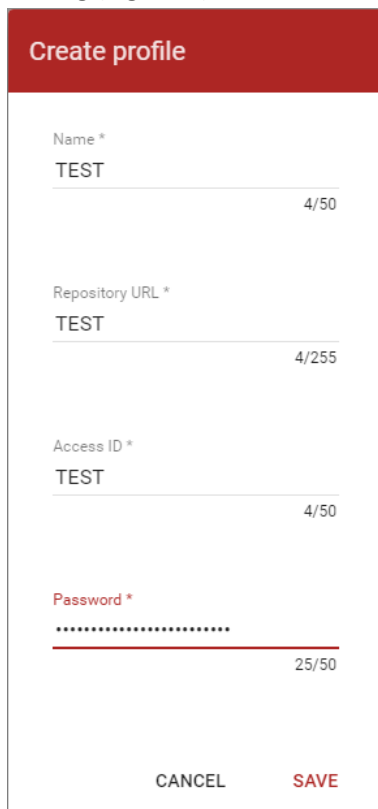

Figure 8: Editing profiles

- Mapping upload and manipulation. The EEM needs to know how to map the metadata associated with the content in the ERMS to the SMURF ERMS format. This mapping is facilitated via an XML mapping file[18] which has to be defined for each repository (Figure 9)
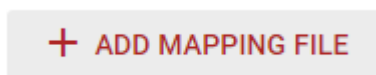

Figure 9: Adding mapping files

---

[18] The documentation on how to create a mapping file is available at https://github.com/eark-project/erms-export-ui-module/tree/master

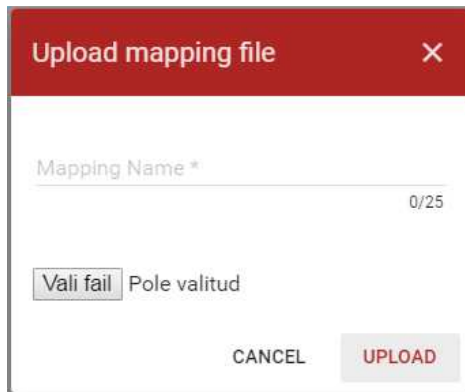The XML mapping file has to be uploaded to the portal (Figure 10).

Figure 10: Uploading the mapping file

All available mapping file are visible when pressed "MANAGE MAPPING FILES" button.

Figure 11: Managing mapping files

If a mapping file is not needed anymore then it can be deleted from the system (Figure 12).

Figure 12: Removing mapping files

- Connecting with the repository. Double click on the name of the repository or "Repository connect" command (Figure 13) establishes a connection with the repository.

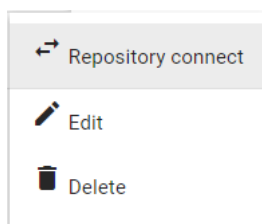**Figure 13: Connecting with the repository**
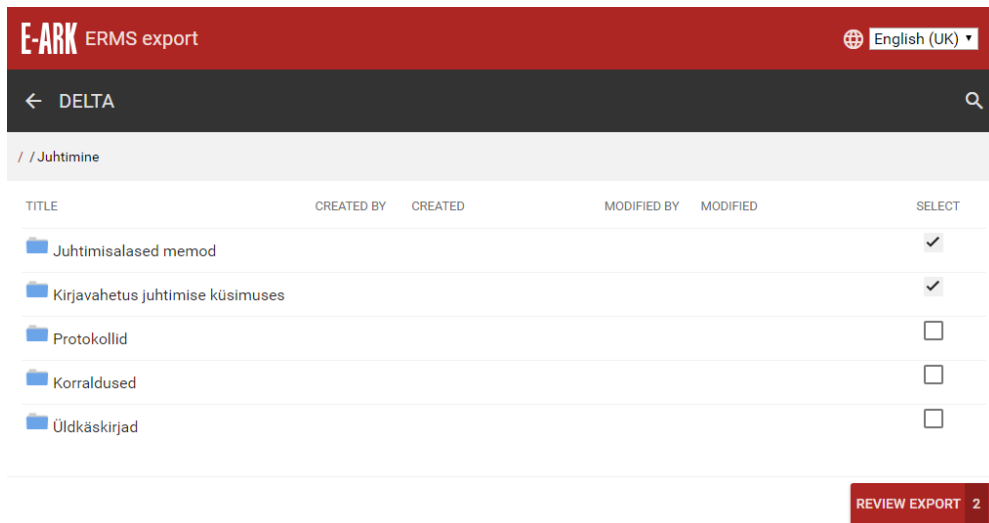
- Selection of contents to export (Figure 14).

- Export of data and metadata from a repository. It is possible to review the selection before export and make any modifications if necessary (Figure 15).
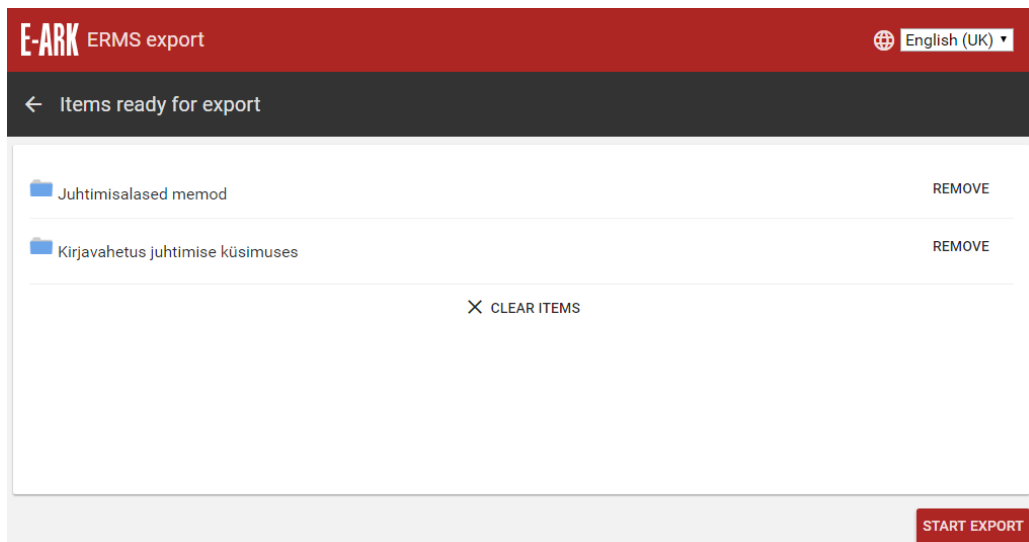
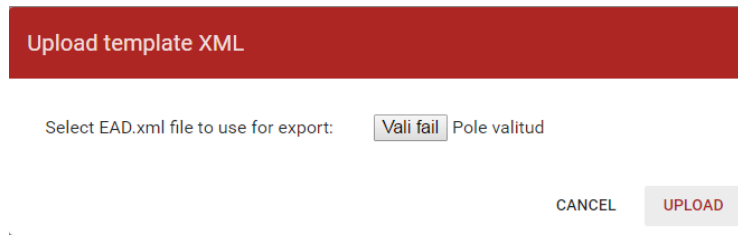If required, upload an EAD template file[19] (Figure 16).

Figure 16: Uploading EAD template

The user will be informed about any problems occurred in the export process (Figure 17)
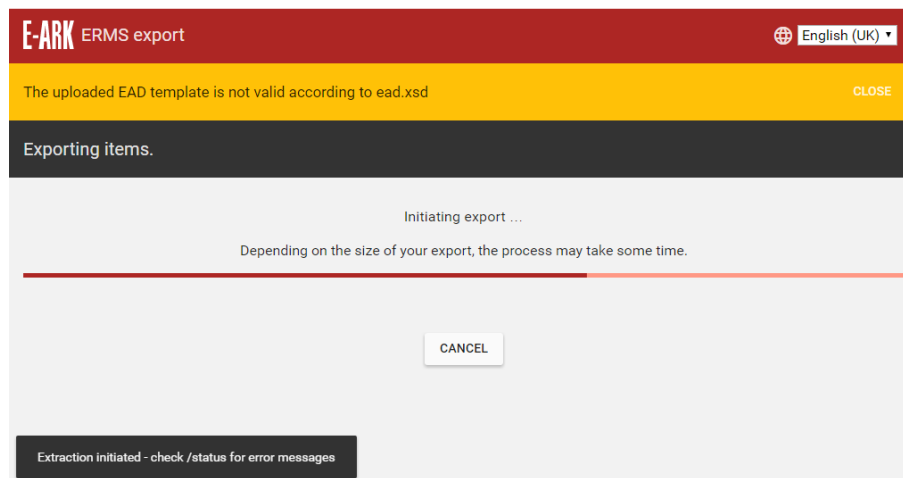
Figure 17: Error during the export

and also about the success of the export (Figure 18).
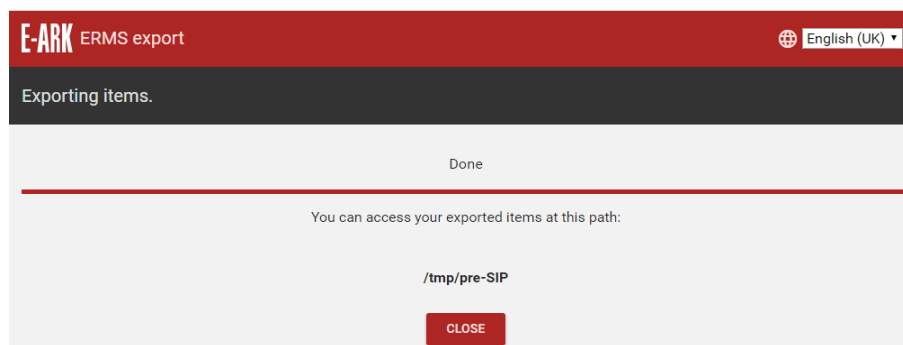
Figure 18: Export done

NB! The ERMS Export Module will also be piloted by a third party vendor

The source code and documentation regarding the EEM was also additionally delivered to a Slovenian Alfresco provider Agenda Open Systems (http://www.agenda.si/en/) which was interested in the tool, but is not an E-ARK project partner. Exact extent of the impact to future Slovenian archival practice can only be seen beyond the E-ARK project.

---

[19] The ead_template.xml is available at https://github.com/magenta-aps/E-Ark-ERMS-export-bridge/blob/master/src/main/resources/ead_template.xml.

2. **Database Preservation Toolkit**

   This toolkit allows conversion between database formats, including connection to live systems, for purposes of digitally preserving databases. The toolkit allows conversion from live databases into preservation formats such as SIARD 2.0, a XML-based format created for the purpose of database preservation (Figure 19). The toolkit also allows loading the archived records in SIARD format back into a live DBMS (Oracle, MySQL etc.).
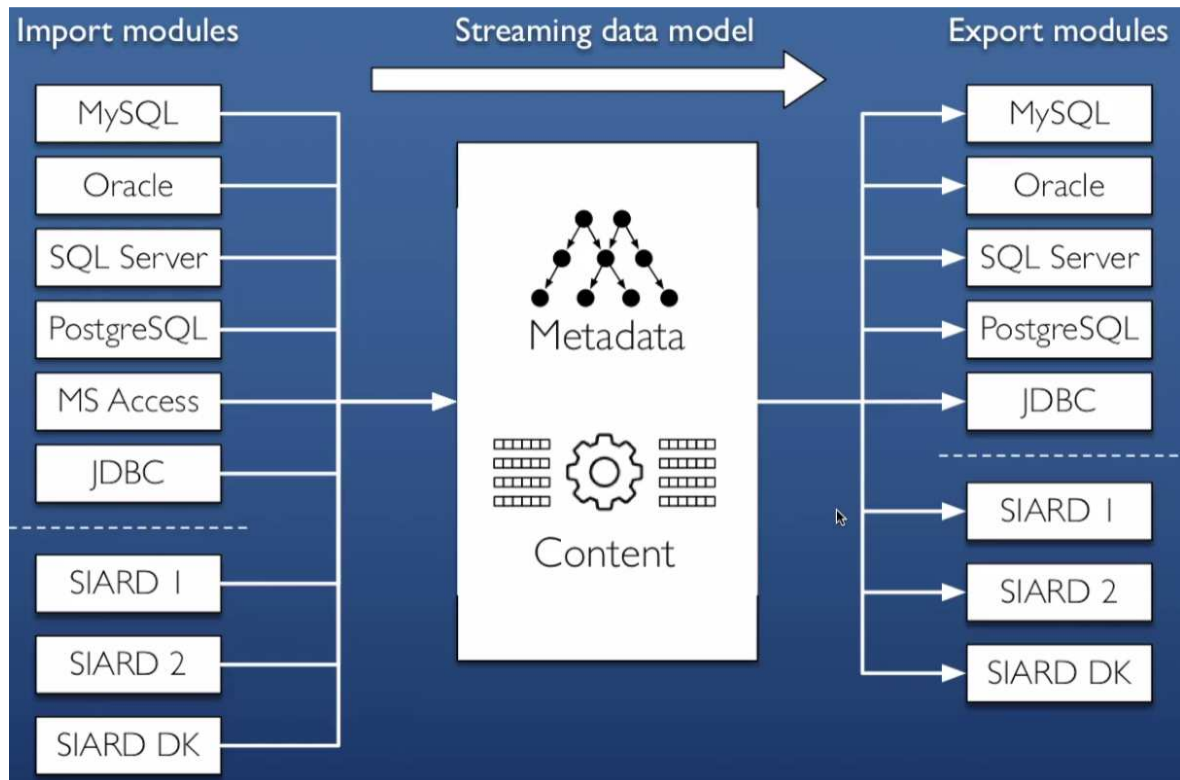
Figure 19: Database Preservation Toolkit

The tool can be used either as a command line tool, or one can choose to plug its graphical user interface on top of it, for the less technical users (see the 2.1 chapter below).

The toolkit is created as a platform that uses input and output modules. Each module supports read or write to a particular database format or live system. New modules can easily be added, providing the ability to convert to or from a new database format or live system.

The currently supported RDBMSs are:

- MySQL/MariaDB;
- PostgreSQL;
- Oracle;
- Microsoft SQL Server;
- Microsoft Access (only as input);
- And other databases (using JDBC).

Using this tool, any database maintained in any of these DBMSs can be converted to SIARD. The same tool can also be used to load a database in SIARD format back into a live DBMS (from the above list), in order to provide access to the database using the capabilities of the database system, like SQL querying and data analysis.

This tool is a command line application, so IT personnel can execute it on servers efficiently, without the overhead of a Graphical User Interface.

2.1 The graphical user interface for the Database Preservation Toolkit

A web-based GUI for the DBPTK has been written in Python (backend) and Material Design Lite (frontend). The frontend for the GUI communicates with the Python backend which initiates the DBPTK by making system calls to the DBPTK jar-file using Java.

The GUI works as follows.

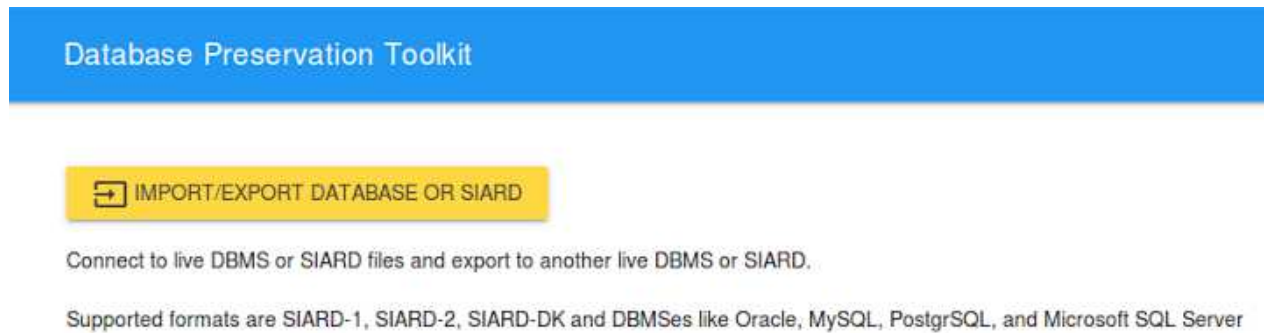First, the user is presented with a page as the one shown below (Figure 20):

Figure 20: Main view

Clicking the yellow button will load a page that enables the user to specify the import settings (Figure 21):
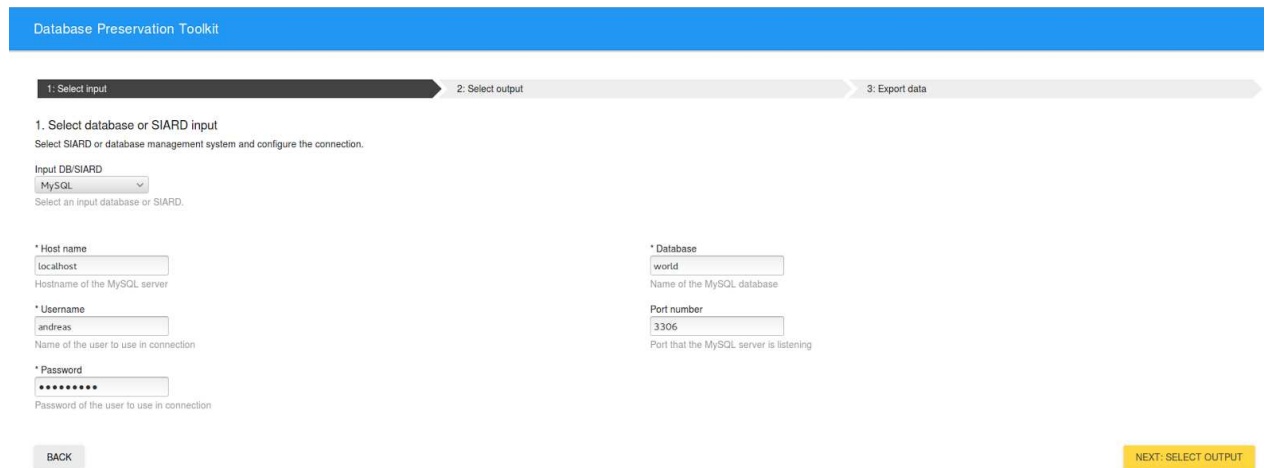

Figure 21: Import settings

The user can select where to import data from. As seen in the screenshot, the user has in this case selected to import data from a MySQL database named "world". When clicking the "Next"-button, the user will be directed to the export page, where (s)he can specify the export settings (Figure 22):

Figure 22: Export settings

In this example, the user has chosen to export to the SIARD-2 format and that the resulting archive should be stored in a file named "world.siard" in the temporary folder of the operating system. It is also possible to compress the SIARD file using a lossless data compression algorithm by selecting 'Compress'; or to choose the 'PrettyXML' option in order to make the XML human readable.

To start the export, the user clicks the "Start Exporting" button and when the export process has finished, the user will be notified of this as shown here (Figure 23):

Figure 23: Export complete

The export has now been completed, and the SIARD-2 archive can be accessed in the file location specified on the export settings page.

The application is available at http://www.database-preservation.com/.
The source code for the DBTK is available at https://github.com/keeps/db-preservation-toolkit and for the DBPTK GUI at https://github.com/eark-project/dbptk-gui-backend

Note: As agreed in the second year of the project, the DBExtractor, which was described as the intended database export tool in the Description of Work, was replaced by a better solution Database Preservation Toolkit. The database preservation related pilots carried out in the E-ARK project have justified this decision.

SIP Creation Tool

SIP Creation Tool is responsible for creating and validating the SIP from various sources as seen in Figure 24 (in the middle). The SIP Creation Tool provides an input to the SIP Receiving Tool and ultimately the SIP to AIP conversion component in the Ingest.
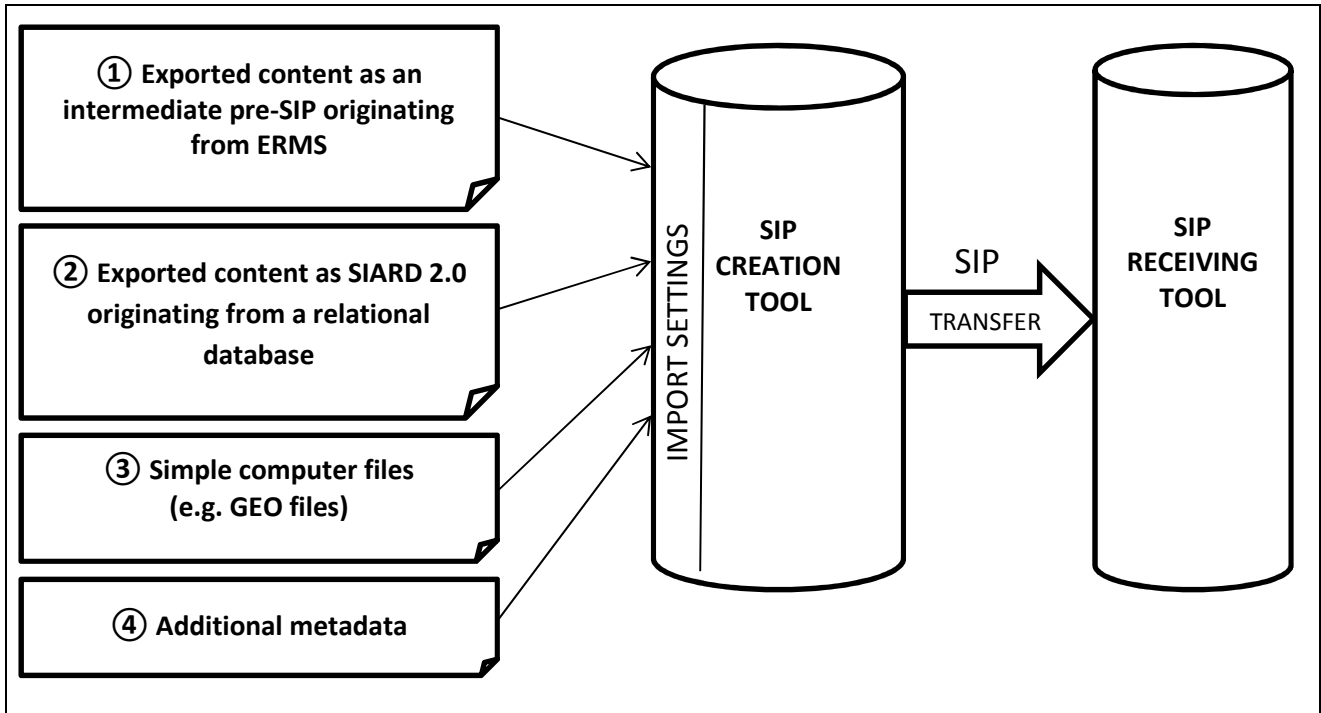

Figure 24: Context of the SIP creation tool

Work Package 3 delivered 3 tools that are meant for SIP creation:

1. **ESSArch Tools for Producer (ETP)**

    ETP is a SIP creator tool which provides functionality to facilitate the preparation of IP's, the creation of SIP's and the submission of SIP's.[20] The tool supports the E-ARK general model but can easily be configured to support any other processes and workflows. Features provided are:
    • Authorizations based on users, groups and permissions;
    • Integration possibility with Active Directory (AD);
    • Translations, e.g. different languages;
    • Submission Agreements (SA) are supported;
    • SA related profiles like SIP profiles, Submit description profiles, Transfer project profiles etc.;
    • Different parallel tasks, steps and workflows can be managed, e.g. parallel work capabilities (Figure 25);
    • Events are logged during every task and step and event types can easily be configured;
    • Locking, unlocking and reuse/removal of IP's in conjunction with authority models;
    • Filter functions is provided, default Archivist organisation and Archival institution;

---

[20] In the course of the project it was decided to develop also the ETA (ESSArch Tools for Archive) as an SIP-receiving counterpart for ETP – as a tool performing the quality control of E-ARK SIPs before delivering these to EPP (ESSArch Preservation Platform) for further processing.

- Search functionality, can be adjusted easily;
- Preparation of the physical and logical structure of the IP in order to be able to add content;
- Upload content (exports/file/folders etc.) into the selected prepared IP;
- Add descriptive metadata for the selected prepared IP;
- Create SIPs e.g. create a container files (tar/zip) of prepared IPs;
- Create submit descriptions in order to facilitate the submission of the SIP;
- Submit SIP to archivist organization together with associated submit description;
- Different metadata standards are supported/used, like METS and PREMIS;
- An API (REST) can be used to easily interact with the tool;
- Quality control (automatic/manually), validators, different conformance checkers etc.
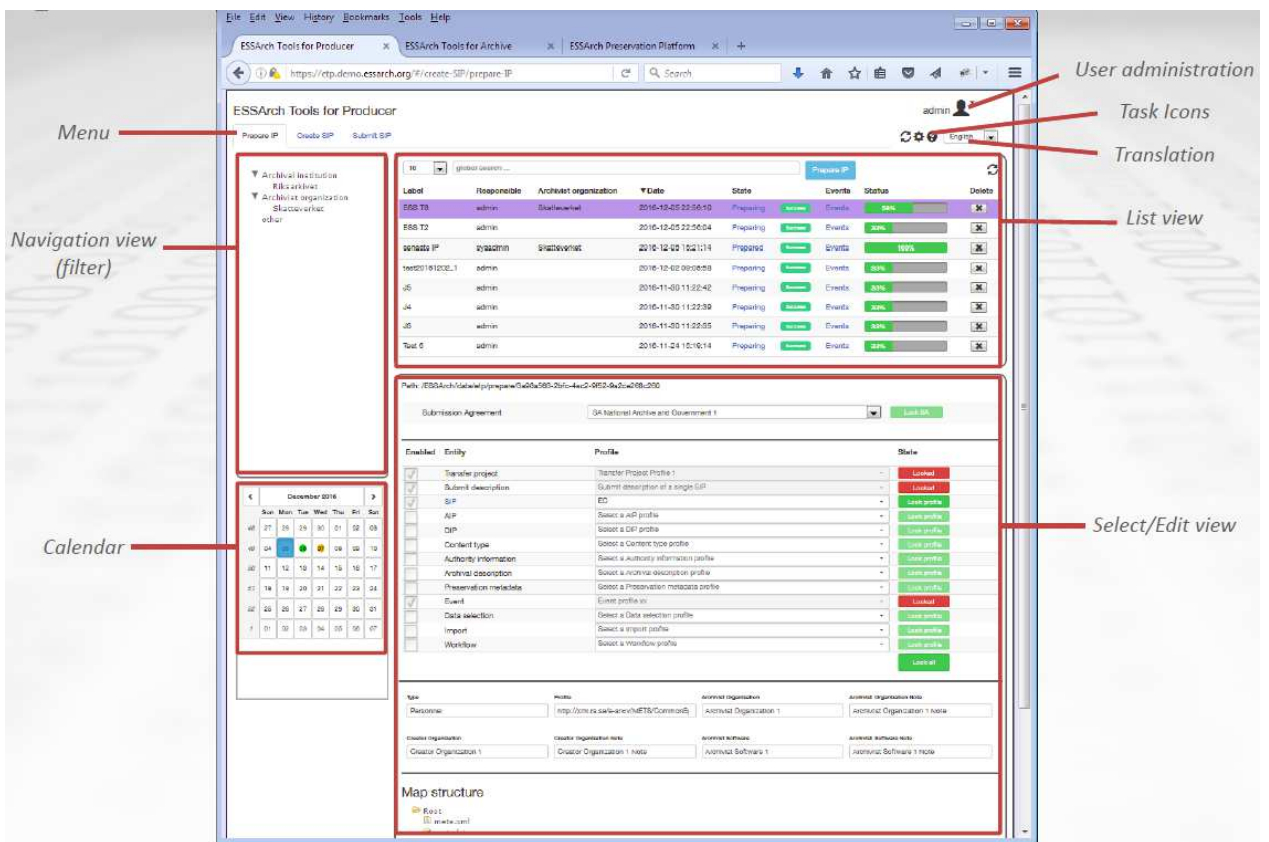
Figure 25: ESSArch Tools for Producer

The application ETP is available at http://etp.essarch.org.

The source code is available at https://github.com/eark-project/ESSArch_Tools_Producer.

More information about the behaviour and functionality of the tool can be found in the documentation at http://doc.essarch.org/.

2. **RODA-in**

It is a tool specially designed for producers and archivists to create SIPs from files and folders available on the local file system. RODA-in satisfies the need for mass processing of data by providing a quick and simple way for creating thousands of valid SIPs with just a few clicks, complete with data and metadata (Figure 26).

The tool includes features such as:

- Create, load and edit classification schemas;
- Automatic association of files/folders to SIP;
- Automatic association of metadata to SIP;
- Definition of metadata templates;
- Support for various metadata formats (EAD, DC, etc.);
- Creation of SIP of unlimited size;
- Creation of SIP in various formats.
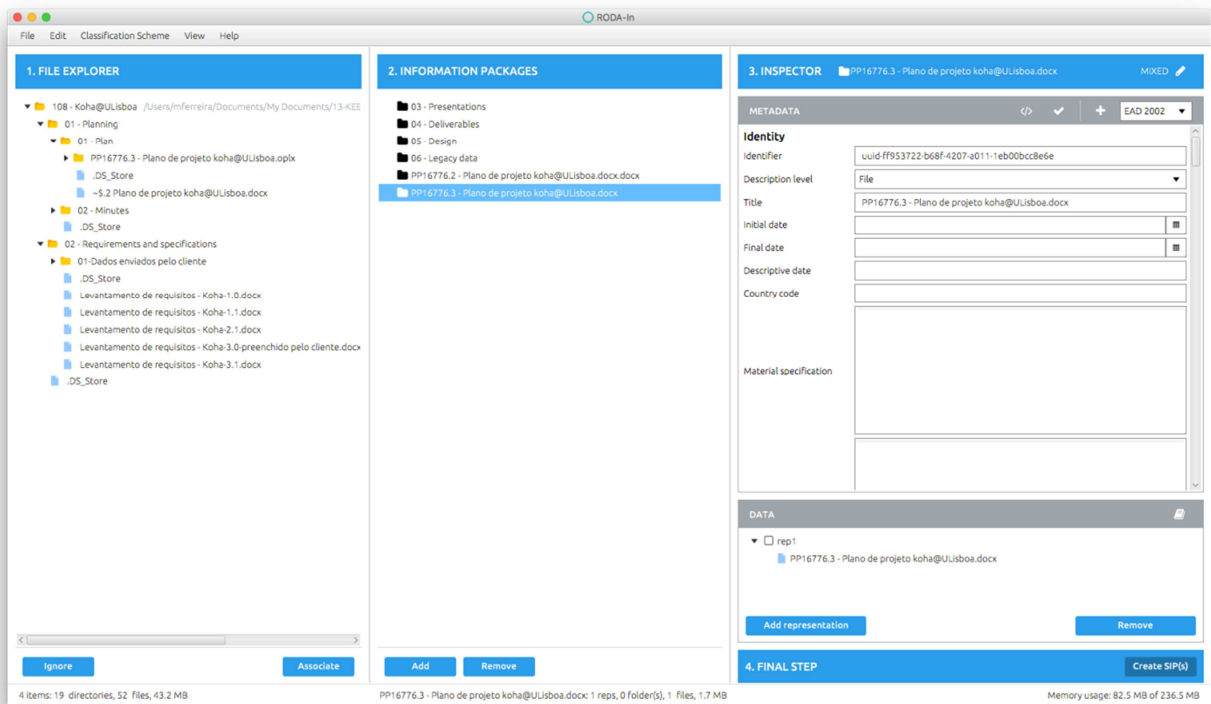
Figure 26: Screenshot of the RODA-in tool in action.

RODA-in application is available at http://rodain.roda-community.org/.

The source code is available at https://github.com/keeps/roda-in.

More information about the behaviour and functionality of the tool can be found in the "How to use" section and video tutorials at https://github.com/keeps/roda-in.

3. Universal Archiving Module (UAM)

The Universal Archiving Module (UAM) is software originally created by the National Archives of Estonia (NAE) for the preparation and transfer of digital documents exported from electronic records managements systems. The software was updated to support the E-ARK SIP format and workflows during the E-ARK project.

Use of UAM requires the ability of an institution's ERMS to export documents and their metadata in XML format. In addition to importing data from records and content management systems, UAM allows users to rearrange, classify and further describe the contents of the transfer; validate the transfer according to the rules established by the archives and finally create SIP packages to be transferred to the digital archives. UAM addresses specifically the normalisation and preparation of records as individual entities.

UAM also allows describing archival creators (Figure 27).
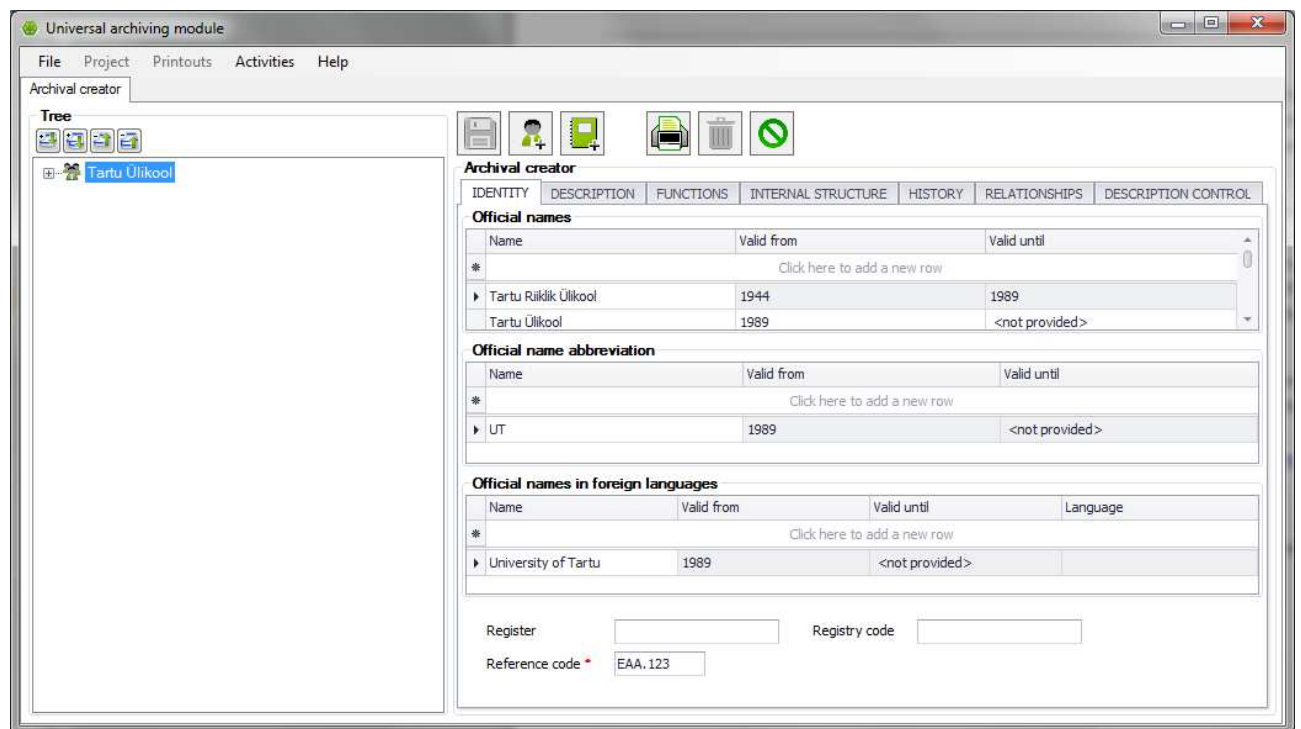
Figure 27: Archival creator in UAM.

More precisely, UAM supports:

- Describing the agents (data producer/archival creator);
- Describing a transfer project;
- Selecting import format;
- Importing classification schemas;
- Building and managing classification schemas;
- Importing records/items, computer files and metadata;
- Adding computer files (manually);
- Enriching metadata (adding, editing and deleting metadata) of records/items or computer files;
- Managing (adding, deleting or relocating/moving) objects (records/items or computer files);

- Creating SIPs;
- Transferring SIPs;
- Removing SIPs from the tool.

Level of maturity: the first version of UAM was released in 2008 and has been used for preparing archival transfers since then. While not mandatory to be used it is officially released on the NAE web site and all public entities are encouraged to use it for their records.

The documentation for UAM is available at http://www.ra.ee/en/information-management/universal-archiving-module/. The source code is available upon request, but the application is freely downloadable from http://www.ra.ee/UAM/UAM.Setup.msi.

Note: As such, as the WP3 promised to release only one SIP Creation Tool in the Description of Work then the objective was greatly exceeded by providing 3 SIP creation tools instead of one.

SIP Receiving Tool

ESSArch Tools for Archive (ETA)

ETA is a set of tools for the archival institution to facilitate receiving SIPs, perform quality controls and to prepare SIPs for ingest into the preservation platform. The tool supports the E-ARK general model but can easily be configured to support any other processes and workflows. Features provided are:

- Authorizations based on users, groups and permissions
- Integration possibility with Active Directory (AD)
- Translations, e.g. different languages
- Different parallel tasks, steps and workflows can be managed, e.g. parallel work capabilities
- Events are logged during every task and step and event types can easily be configured
- Filter functions is provided, default Archivist organisation and Archival institution
- Search functionality, can be adjusted easily
- Verify submit descriptions in order to facilitate the receiving of the SIPs
- Receive any unidentified SIP (e.g. not created by ETP)
- Transfer SIPs for ingestion into the preservation platform EPP
- An API (REST) can be used to easily interact with the tool
- Perform quality control (automatic/manually, validators, conformance checkers etc.)

ETA is a part of the Ingest process step and can be easily compared to a reception desk where you receive packages, performs the first checks of the packages and then places them at the appropriate shelves behind the reception desk, ready to be picked up by the persons responsible for the next steps of the Ingest process (Figure 28).
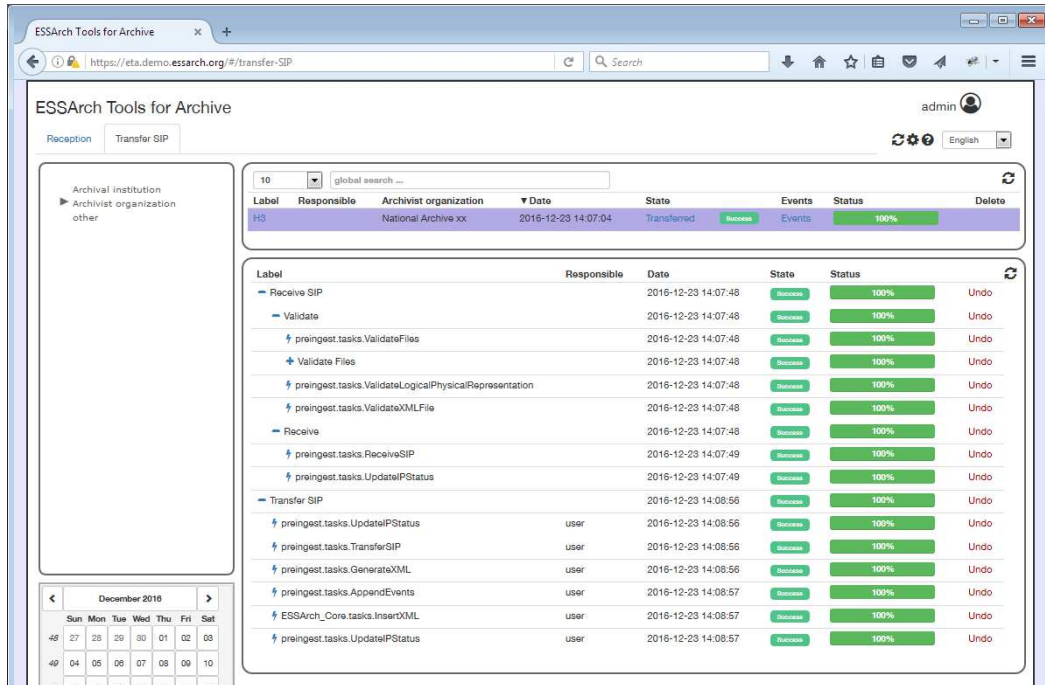
Figure 28: ESSArch Tools for Archive

The application ETP is available at http://etp.essarch.org.

The source code is available at https://github.com/eark-project/ESSArch_Tools_Archive.

More information about the behaviour and functionality of the tool can be found in the documentation at http://doc.essarch.org/.

4. Glossary

Table 1: Glossary

Agent	Same as Archival Creator, Data Provider (definition in the sense of the current document).
AIP	OAIS: An Archival Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.
Archive	An Organisation that intends to preserve information for Access and use by a Designated Community.
Classification schema	The classification system used for archival fonds. Synonym for the aggregation in the archival context (definition in the sense of the current document).
EDRMS or ERMS	Electronic (Documents) and Records Management System is a type of content management system and refers to the combined technologies of document management and records management systems as an integrated system.
Export	Extracting/exporting information from a producer system (definition in the sense of the current document). The extracted content will be imported to a SIP Creation Tool for SIP preparation.
Export format	The specification of the exported information. The format is very close to the full E-ARK SIP if not the same.
General Model	The General Model is an E-ARK deliverable (D2.1), which breaks down complex processes into conceptual level activities, written in plain language, with minimal technical details, so that stakeholders (record managers, archivists, system designers, programmers) have a common understanding of the given examples.
Import	Loading information into the SIP Creation Tool (definition in the sense of the current document). The term import settings refer to the process of selecting one concrete format for loading the information into the SIP Creation Tool.
Information Package	A logical container composed of optional Content Information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information.
Producer / producer system	The person or system that provides the SIP (the information to be preserved and the relevant metadata) to the archive.
Record	Same as the Item in ISAD(G) (definition in the sense of the current document).
SFSB	Simple file-system based records. Unstructured electronic records, documents as computer files organised in folders, not in connection with an ERMS.
Submission	Same as Data Submission Session in OAIS. A delivery of media or a single telecommunications session that provides Data to an OAIS. The Data Submission Session format/contents is based on a data model negotiated between the OAIS and the Producer in the Submission Agreement. This data model identifies the logical constructs used by the Producer and how they are represented on each media delivery or in the telecommunication session.
Submission agreement	A Submission Agreement defines the nature and scope of the records involved in an Ingest Project and delineates the manner in which the Archive and the Producer execute the transfer, validation, and transformation of these records. In addition to guiding the work of the Producer and the Archive for transfer and transformation and serving as the benchmark for validation, it also provides both entities a document describing the terms of an Ingest Project that they can endorse and agree to. The Submission Agreement can cover a single Ingest Project or serial Projects. Source: http://dca.lib.tufts.edu/features/nhprc/reports/ingest/submission.html

Submission Information Package (SIP)	An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.
Transfer agreement	Same as submission agreement.
Transfer project	One submission or delivery.