



D.2.4.P3 Ingest from government agencies

(Full Scale Pilot 3)

DOI: 10.5281/zenodo.1171476

Grant Agreement Number:	620998
Project Title:	European Archival Records and Knowledge Preservation
Release Date:	12 th February 2018
Contributors	
Name	Affiliation
Karin Oolu	National Archives of Estonia
István Alföldi	National Archives of Hungary
David Anderson	University of Brighton
Janet Anderson	University of Brighton

Table of Contents

1. EXECUTIVE SUMMARY.....1

2. PILOT DOCUMENTATION2

2.1 SCENARIOS 2

2.2 INTRODUCTION..... 2

2.3 ORGANISATIONS INVOLVED 2

2.4 SOFTWARE COMPONENTS 3

2.5 DATA CHARACTERIZATION 3

2.6 PILOT WORKFLOW 5

2.7 INSTALLATION INSTRUCTIONS..... 14

1. Executive Summary

This document is part of the deliverable:

D2.4) Pilot documentation

Pilot documentation: This package of documentation will provide technical and end-user guidance to support not only the pilot sites but also possible future deployments thereafter. [month 33] (from DoW)

Structure of this deliverable

The deliverable is a package of linked documents.

This **Summary** contains the common information and short overview of the pilots, along with links to the final version of the Pilot Definition excel files and Pilot Documentation Packages. The **Pilot Definition** excel provides detailed information about scenarios, data sets and step-by-step preparation and process step instructions. The **Pilot Documentation Package** is created by the pilot staff responsible for the pilot execution. This package contains additional information along with screenshots (and videos in some cases) of the tools during the execution of the pilot.

Summary (this document) – Created by WP2

Pilot Package – Pilot 1

- Pilot Definition (Final version) – Created by WP2 and Pilot 1 responsible
- Pilot Documentation files – Created by Pilot 1

Pilot Package – Pilot 2

- Pilot Definition (Final version) – Created by WP2 and Pilot 2 responsible
- Pilot Documentation files – Created by Pilot 2

Pilot Package – Pilot 3

- Pilot Definition (Final version) – Created by WP2 and Pilot 3 responsible
- Pilot Documentation files – Created by Pilot 3

Pilot Package – Pilot 4

- Pilot Definition (Final version) – Created by WP2 and Pilot 4 responsible
- Pilot Documentation files – Created by Pilot 4

Pilot Package – Pilot 5

- Pilot Definition (Final version) – Created by WP2 and Pilot 5 responsible
- Pilot Documentation files – Created by Pilot 5

Pilot Package – Pilot 6

- Pilot Definition (Final version) – Created by WP2 and Pilot 1 responsible
- Pilot Documentation files – Created by Pilot 6

Pilot Package – Pilot 7

- Pilot Definition (Final version) – Created by WP2 and Pilot 7 responsible
- Pilot Documentation files – Created by Pilot 7

2. Pilot documentation

2.1 Scenarios

Scenario 1: **Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica**

and

Scenario 2: **Provide access to records from governmental institution through RESTful services**

2.2 Introduction

The main part of the pilot 3 includes the export of electronic records and their metadata from EDRM systems of Estonian public sector institutions, transfer and ingest to the NAE digital repository. There will be data selected and extracted from the native ERMS (DELTA) Export Module in the Ministry of Justice in Estonia, exported to the Universal Archival Module (UAM) of the National Archives of Estonia (NAE) to create E-ARK SIP and ingested to Preservica (NAE) in the first scenario.

In addition to that, Estonian agencies have the responsibility to make public electronic records with no access restrictions available on their web sites, which means that the pilot will also enable this through standardised linking/access methods that are implemented in the agencies' digital infrastructure / web site. The second scenario will be pilot access to the records from the agency.

Detailed descriptions of the organisations, tools and data will follow in the next chapters.

2.3 Organisations involved

The organizations involved in this pilot are: The Ministry of Justice of Estonia (data provider agency) and The National Archives of Estonia (NAE)

The main duty of the **Ministry of Justice** is to plan and to carry out a legal and criminal policy of the state, which will help ensure an open and secure society, where people may be assured of the use and protection of their rights. Ministry of Justice of Estonia Works as a coordinator of legislative drafting, stands for the systematic development of law and supports the formation of quality legislative drafting. The competence of the Ministry of Justice also includes organizing the professional activities and legal services of the notaries' offices, the work of the service of sworn translators and bailiffs, coordinating crime prevention and harmonization of Estonian legislations with the law of the European Union, processing international rogatory letters, and guaranteeing the legality of the activities of the Bar Association, and organizing the representation of the state in judicial proceedings.

A total of more than 3,300 people are employed in the administrative field of the ministry, 160 of whom work in the ministry (more information can be found from this link: <https://www.just.ee/et/ministry-justice>)

The **National Archives of Estonia** (NAE) is the central governmental archival agency in Estonia. The Digital Archives of NAE is the leading national authority on electronic records management and long-term preservation. It has also established a fully functional digital preservation system and issues relevant national guidelines.

2.4 Software components

The main software products involved in this pilot are: native ERMS (DELTA) Export Module, UAM – Universal Archival Module, E-ARK CMIS Browser.

ERMS (DELTA) Export Module

Delta is a content management system, where administrators can define different types with complex structure for documents and types for compound workflows. The users can create (modify, delete, search) documents based on the types defined by administrators and start (postpone, finish, search) compound workflows for/with documents, assign (execute, search) tasks to users (including himself). The system usually has one active store for documents and 1 to n archive stores. There are also included several data migration utilities from different systems to DELTA (e.g. SharePoint). Also Delta has event logs monitoring system.

DELTA is used in many departments and official organizations in Estonia. Performance tests has been done for amount of users from 100 until 10 000. DELTA is also operating with about 10 000 000 documents per store. All of it covering all requirements of Estonian official organizations.

At Ministry of Justice it is used Delta version 5.2.2.150. The Delta Export Module allows pulling out the data and metadata from source systems for archiving purposes.

Universal Archiving Module (UAM)

The universal archiving module (UAM) is a software created by the NAE for the preparation and transfer of digital documents extracted from electronic records managements systems. Use of UAM requires the ability of an institution's ERMS to export documents and their metadata in XML format.

In addition to import data from records and content management systems UAM allows to rearrange, classify and further describe the contents of the transfer; validate the transfer according to the rules established by the National Archives of Estonia and finally create SIP packages to be transferred to the digital archives. UAM addresses specifically the normalisation and preparation of records as individual entities.

Level of maturity: The current version of UAM was released in autumn 2012 and has been used for preparing two transfers since then. While not mandatory to be used it is officially released on the National Archives of Estonia web site and all public entities are encouraged to use it for their records.

Technical parameters of UAM:

- a client of Windows;
- uses .NET framework;
- uses programming languages C# and LINQ;
- requires SQL (Express) server;
- uses JHove component.

E-ARK CMIS Browser

The E-ARK CMIS Browser is a lightweight access portal capable of connecting to a CMIS-compliant digital repository (piloted with Preservica). The tool allows for the simple browsing and searching of content stored in the repository, being at the same time installed at the (internal) network of the agency

2.5 Data characterization

The first data-set used in the pilot 3 was comprised of a different series of the Ministry of Justice of Estonia documents, of total number of 15 files. Document formats are in ddoc, docx, PDF, TIFF formats.

Jrknr Tähis Pealkiri Kehtiv alates Kehtiv kuni
 1 1.1-8 Ministri, kantsleri ja hallatavate asutuste juhtide vahetumisel üleandmise aktid 01.01.2010 31.12.2010

1.1-8 Ministri, kantsleri ja hallatavate asutuste juhtide vahetumisel üleandmise aktid (dokumendid)									
<input type="checkbox"/>	Viit	Reg kpv	Pealkiri	Liik	Saatja/Saaja	Vastutaja	Tähtaeg	Vastamise kpv	Failid
<input type="checkbox"/>			JM-RIK_üleandmise-vastuvõtmise ...	Väljaminev kiri		Liis Laaneloog			
<input type="checkbox"/>	1.1-8/17728	31.12.2010	Akt JUM RIK 30.12.2010 ...	Akt	RIK	Liis Laaneloog			
<input type="checkbox"/>	1.1-8/17728	31.12.2010	RTO-Jum_üleandmise-vastuvõtmise ...	Akt	Riigikantselei	Liis Laaneloog			
<input type="checkbox"/>	1.1-8/17728	05.01.2011	Allkirjastatud vara üleandmise ...	Akt	RIK	Kristel Marksalu			

2 1.1-8 Juhtide üleandmise-vastuvõtmise aktid 01.01.2008 31.12.2008

1.1-8 Juhtide üleandmise-vastuvõtmise aktid (dokumendid)									
<input type="checkbox"/>	Viit	Reg kpv	Pealkiri	Liik	Saatja/Saaja	Vastutaja	Tähtaeg	Vastamise kpv	Failid
<input type="checkbox"/>		01.03.2008	KEKK akt	Akt		Liis Laaneloog			
<input type="checkbox"/>	[AK] /25	02.01.2008	Ämari Vangla asjaajamise ...	Sissetulev kiri	Ämari Vangla	Liis Laaneloog			
<input type="checkbox"/>	[AK] 14132	10.12.2008	Hendrik R. Bott praktika ...	Sissetulev kiri	Hendrik R. Bott	Gerly Kuivkaev			

Jrknr Tähis Pealkiri Kehtiv alates Kehtiv kuni
 1 1.4-3 Justiitsministeeriumi tegevuse auditeerimise toimik 01.01.2004 31.12.2004

<input type="checkbox"/>	Viit	Reg kpv	Pealkiri	Liik	Saatja/Saaja	Vastutaja	Tähtaeg	Vastamise kpv	Failid
<input type="checkbox"/>	5-04/6734	25.05.2004	Auditi lõpparuande koopia ...	Väljaminev kiri	Ravala pst 3	Liis Laaneloog			

2 3.1-18 Personalialane aruanne 01.01.2003 31.12.2003

<input type="checkbox"/>	Viit	Reg kpv	Pealkiri	Liik	Saatja/Saaja	Vastutaja	Tähtaeg	Vastamise kpv	Failid
<input type="checkbox"/>	1-3-13/10068	08.08.2003	Aruanne tööjõu liikumine ...	Väljaminev kiri	Statistikaamet, Palgastatistika ...	Liis Laaneloog			
<input type="checkbox"/>	1-3-13/12189	03.10.2003	Aruanne "Tööjõu liikumine ...	Väljaminev kiri	Palgastatistika sektor	Liis Laaneloog			

3 3.1-18 Personalialane aruanne 01.01.2004 31.12.2004

<input type="checkbox"/>	Viit	Reg kpv	Pealkiri	Liik	Saatja/Saaja	Vastutaja	Tähtaeg	Vastamise kpv	Failid
<input type="checkbox"/>	1-3-13/1104	23.01.2003	Aruanne "Tööjõu liikumine ...	Väljaminev kiri	Statistikaamet, Palgastatistika ...	Liis Laaneloog			
<input type="checkbox"/>	1-3-13/5668	30.04.2004	Aruanne 2004. a. I kvartal	Väljaminev kiri	Statistikaamet, Palgastatistika ...	Liis Laaneloog			
<input type="checkbox"/>	260	08.01.2004	2004. a. tööjõu liikumine ...	Sissetulev kiri	Statistikaamet	Liis Laaneloog			
<input type="checkbox"/>	261	08.01.2004	2004. a. palagaaruande ...	Sissetulev kiri	Statistikaamet	Liis Laaneloog			

4 1.4-3 Justiitsministeeriumi tegevuse auditeerimise toimik 01.01.2005 31.12.2005

<input type="checkbox"/>	Viit	Reg kpv	Pealkiri	Liik	Saatja/Saaja	Vastutaja	Tähtaeg	Vastamise kpv	Failid
<input type="checkbox"/>			Siseauditi eest vastutava ...	Majasisene dokument		Liis Laaneloog			

2.6 Pilot workflow

“Data selection and extraction” displays the step of the E-ARK Pre-Ingest process where the content to be exported will be selected and extracted from the producer system. Extraction activity must be preceded by the data selection process.

Workflow steps in DELTA:

In case of ERMS, the **data selection** must be implemented inside an ERMS, as a piece of native functionality, or in an “archiving module”. In DELTA this functionality is configured on the basis of lifecycle management. The first step is to fix the lifecycles, then assign concrete lifecycle to a case file. Subsequently it is possible to choose the command to show the case files to be exported.

Please find information about case file life-cycle from the Figure 1. Chosen lifecycle and the next action with date and other involved info are shown with red ovals.

Figure 1. Case file life-cycle

After selecting the life-cycle it is possible to see the list of case-files to be exported, e.g in Oct 2016.

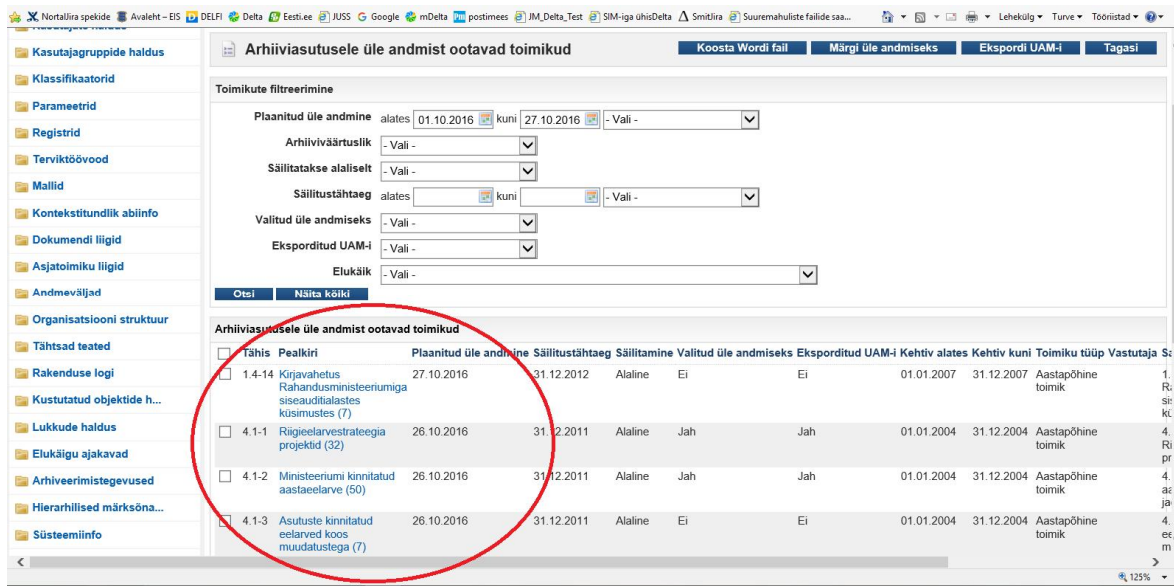


Figure 2. List of case-files of one life-cycle to be exported in Oct 2016

The next step shows the selected case-file to be exported in Oct 2016. After clicking the confirmation button the system creates an xml-file which will be exported to UAM.

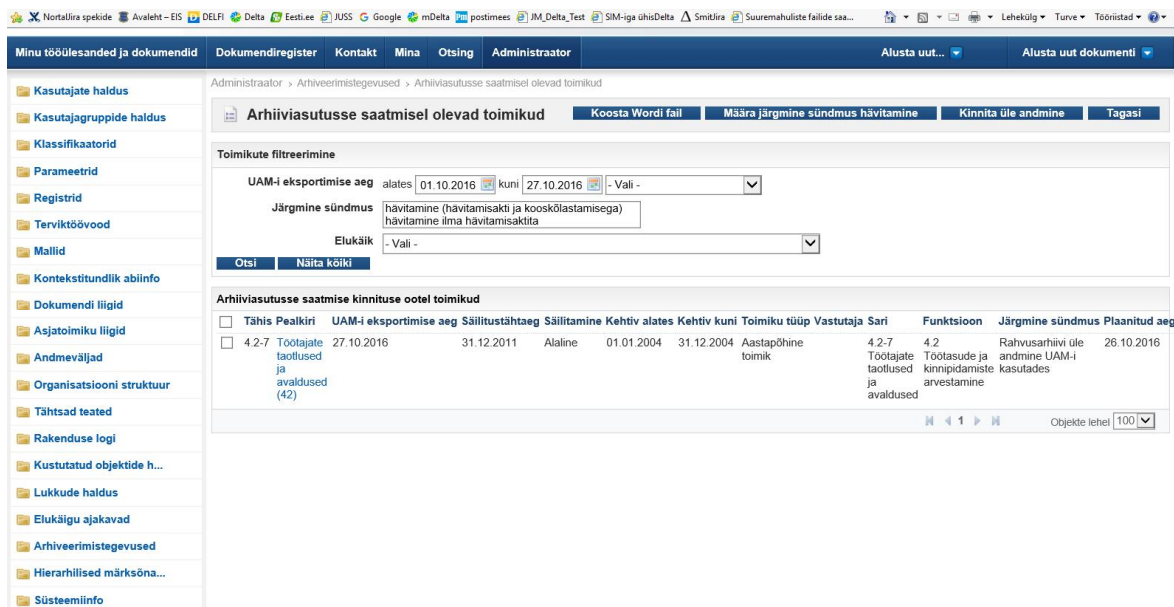


Figure 3. Selection waiting for export to UAM, not confirmed yet

Performed archival actions in DELTA are seen in the Figure 4. We can see that selection process has been documented in DELTA Export Module. The red oval shows the title of the actual .xml file which will be transferred to UAM.

Sooritatud arhiveerimistegevused Tagasi

Tegevuste filtreerimine

Teostatud alates 27.10.2015 kuni - Vali -

Tegevus - Vali -

Otsi Näita kõiki

Arhiveerimistegevused ja tulemused

Tegevus	Teostatud	Teostaja	Staat	Tulemusfail	Dokument
Hävitamisakti kavandi loomine	17.05.2016 14:31	Liis Laaneloog	lõpetatud		Loo uus dokument
Toimikute märkimine arhiivasutusele üle andmiseks	17.05.2016 14:41	Liis Laaneloog	lõpetatud		Loo uus dokument
Toimikute märkimine arhiivasutusele üle andmiseks	17.05.2016 14:42	Liis Laaneloog	lõpetatud		Loo uus dokument
Arhiivasutusse üle andmiseks UAM-i formaadis XML-i eksportimine	17.05.2016 14:43	Liis Laaneloog	lõpetatud	UAM20160517.xml	Loo uus dokument
Arhiivasutusse saatmisel olevate toimikute üle andmise kinnitamine	17.05.2016 14:44	Liis Laaneloog	lõpetatud		Loo uus dokument
Toimikute märkimine arhiivasutusele üle andmiseks	17.05.2016 16:06	Liis Laaneloog	lõpetatud		Loo uus dokument
Arhiivasutusse üle andmiseks UAM-i formaadis XML-i eksportimine	17.05.2016 16:07	Liis Laaneloog	lõpetatud		Loo uus dokument
Arhiivasutusse saatmisel olevate toimikute üle andmise kinnitamine	17.05.2016 16:18	Liis Laaneloog	lõpetatud		Loo uus dokument
Hävitamisakti kavandi loomine	17.05.2016 16:19	Liis Laaneloog	lõpetatud		Loo uus dokument
Hävitamine koos hävitamisakti koostamisega	17.05.2016 16:20	Liis Laaneloog	lõpetatud		Loo uus dokument
Hävitamisakti kavandi loomine	18.05.2016 11:32	Liis Laaneloog	lõpetatud		Loo uus dokument
Hävitamisakti kavandi loomine	25.05.2016 11:15	Liis Laaneloog	lõpetatud		Loo uus dokument
Hävitamine koos hävitamisakti koostamisega	25.05.2016 11:15	Liis Laaneloog	lõpetatud		Loo uus dokument
Hävitamisakti kavandi loomine	31.05.2016 10:39	Liis Laaneloog	lõpetatud		Loo uus dokument
Lihtsustatud hävitamine ilma hävitamisaktita	31.05.2016 10:40	Liis Laaneloog	lõpetatud		Loo uus dokument
Toimikute märkimine arhiivasutusele üle andmiseks	31.05.2016 11:29	Liis Laaneloog	lõpetatud		Loo uus dokument
Arhiivasutusse üle andmiseks UAM-i formaadis XML-i eksportimine	31.05.2016 11:29	Liis Laaneloog	lõpetatud		Loo uus dokument
Ülevaatamist vajavate toimikute nimekiri	10.06.2016 14:13	Liis Laaneloog	lõpetatud		Loo uus dokument
Hävitamisakti kavandi loomine	10.06.2016 14:33	Liis Laaneloog	lõpetatud		Hävitamisakt 5 aastastele toimikutele kuni 31.11.2011
Hävitamine koos hävitamisakti koostamisega	10.06.2016 14:41	Liis Laaneloog	lõpetatud		Toimikute hävitamise akt
Ülevaatamist vajavate toimikute nimekiri	30.08.2016 08:02	Liis Laaneloog	lõpetatud		Loo uus dokument
Toimikute märkimine arhiivasutusele üle andmiseks	13.09.2016 10:48	Liis Laaneloog	lõpetatud		Loo uus dokument

https://test-jmdhs.just.sise/dhs/d/a/workspace/SpacesStore/68162ce9-fe13-4525-9111-9dad44fb3fc1/UAM20160517.xml 125%

Figure 4. Performed actions in Delta

Workflow steps in UAM:

Universal Archiving Module (UAM) is capable of importing the extraction to the tool by placing all the metadata into a database and computer files on to a drive. Data provider can now **prepare the records for archiving, validate the work and create submission information packages (SIPs)**.

First action in UAM after downloading it is the configuration of input and output formats, as seen in the Figure 5.

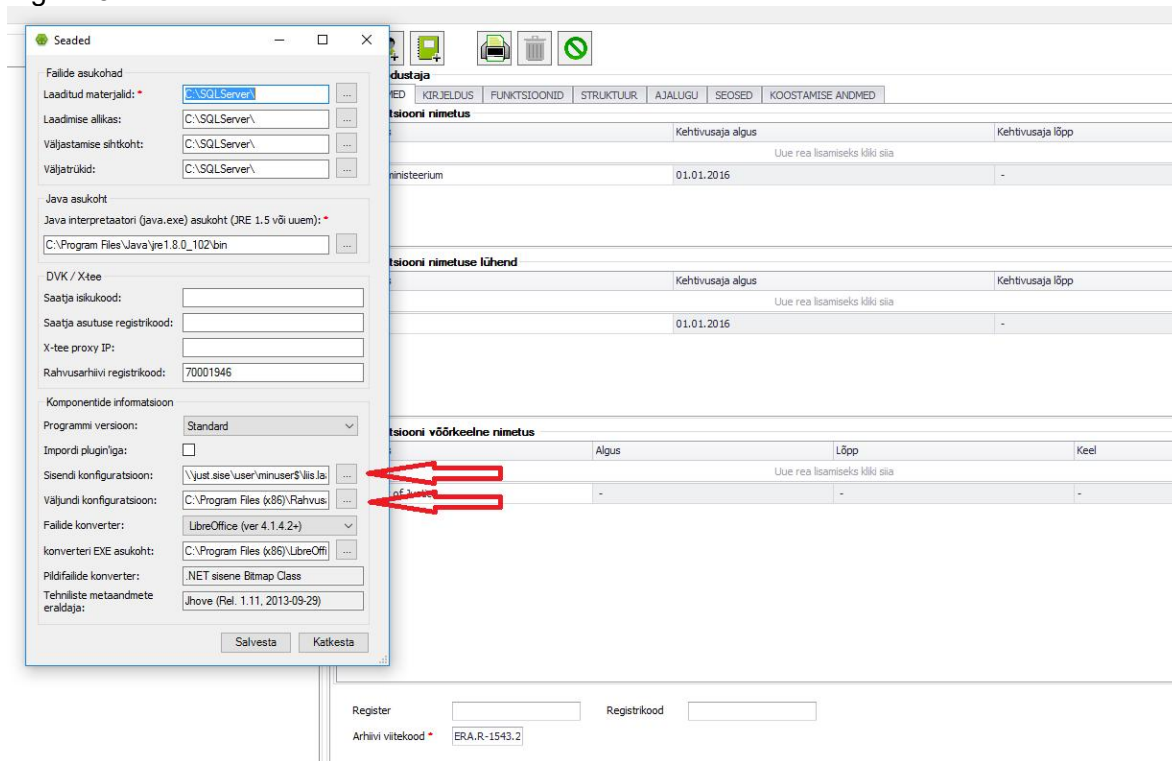


Figure 5. UAM settings

The input configuration (“sisendi konfiguratsioon” in Estonian) was prepared as an XSLT file. The file helped to transform the native Delta export into a suitable format for UAM. The support for an E-ARK SIP was created as one of the options of the transfer to digital archives (Figure 14), but the output configuration (“väljundi konfiguratsioon” in Estonian) shown in Figure 5 provides a way to make additional changes to the predefined export format.

The description of an archive creator will follow as the next activity in UAM. This must be done manually, but data inserted by one agency just once. This step is based on EAC standard, more info from this link: <http://www2.archivists.org/groups/technical-subcommittee-on-eac-cpf/encoded-archival-context-corporate-bodies-persons-and-families-eac-cpf>

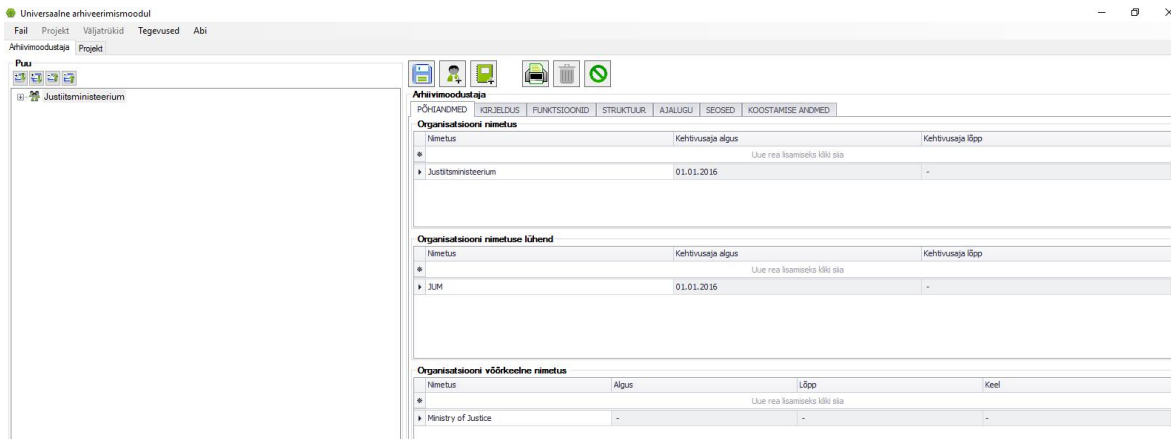


Figure 6. Description of archive creator

SIP creation starts by approving the archival project. In project creation it is important to save the details of the creator, marked with red oval in the next figure. The information will help to identify the right transfer project in case there is more than one project related to the archival creator at some point in the future.

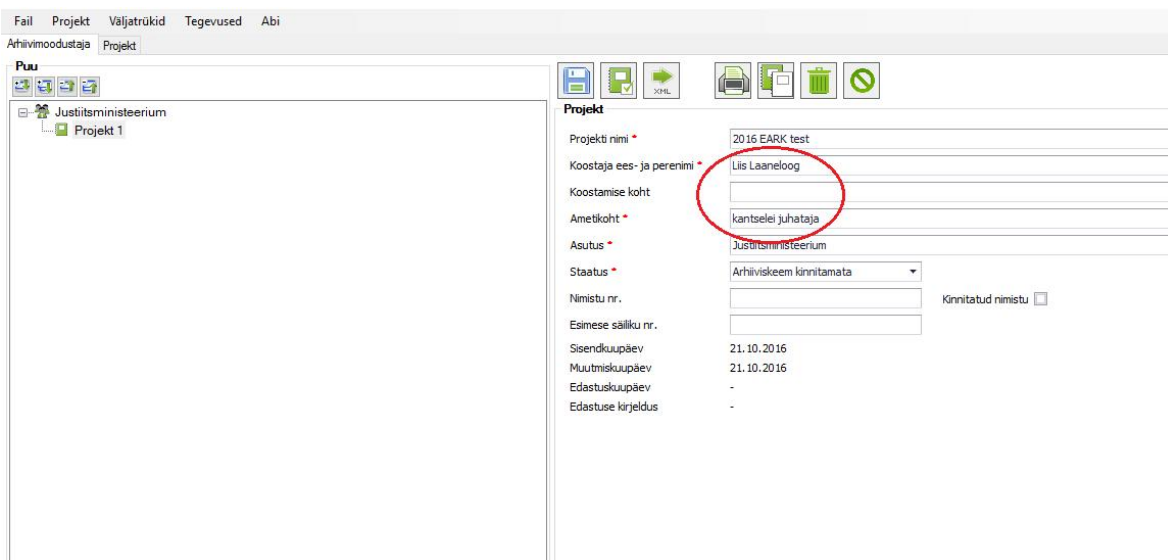


Figure 7. UAM Project description

After opening the project in UAM it is possible to import the .xml files from DELTA:

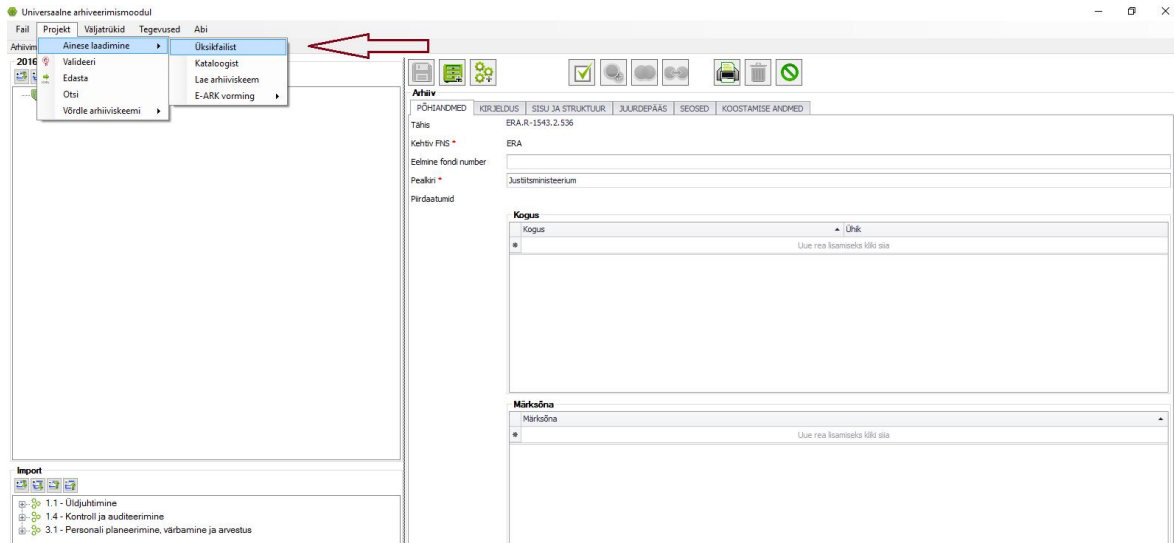


Figure 8. Import to UAM

Next step is to choose the XML files we want to import to UAM. One XML can contain many records with their metadata and computer files encoded as Base64. For example, there are only 4 XML files seen in Figure 9 although they contain 15 records.

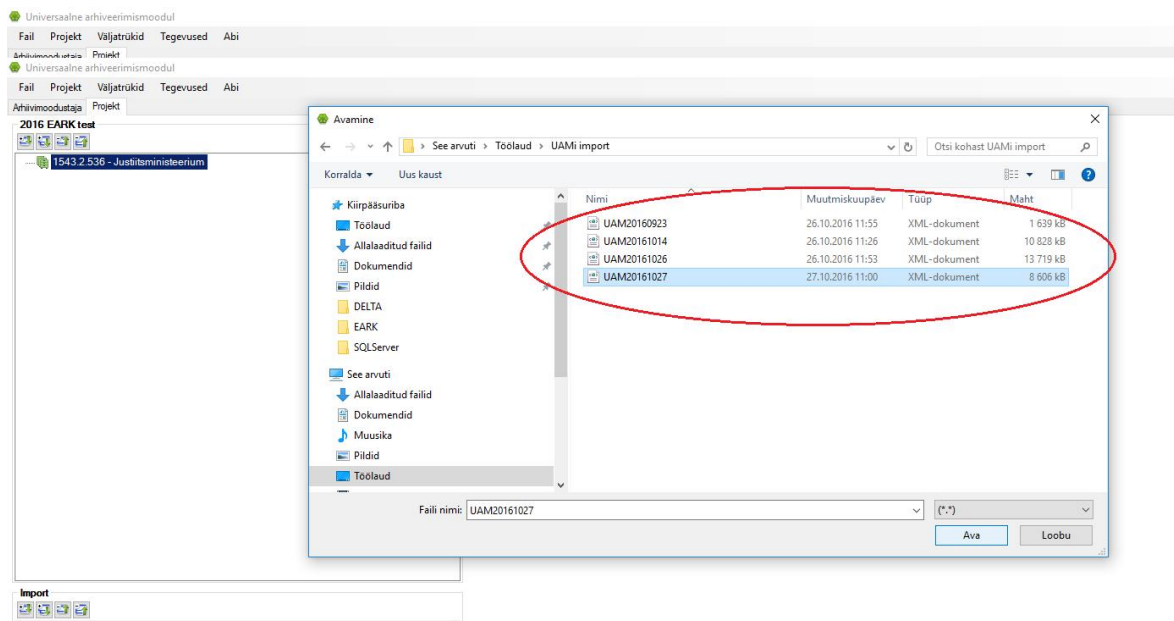


Figure 9. UAM import 2

In the case the import is performed correctly the system creates the report of the success or failure. E.g, in our pilot the report says that 8 files were imported after removing digital signature, 6 files did not correspond to archival formats, 6 files were successfully converted to archival formats, etc. as seen from the Figure 10. There are several checks done in that phase.

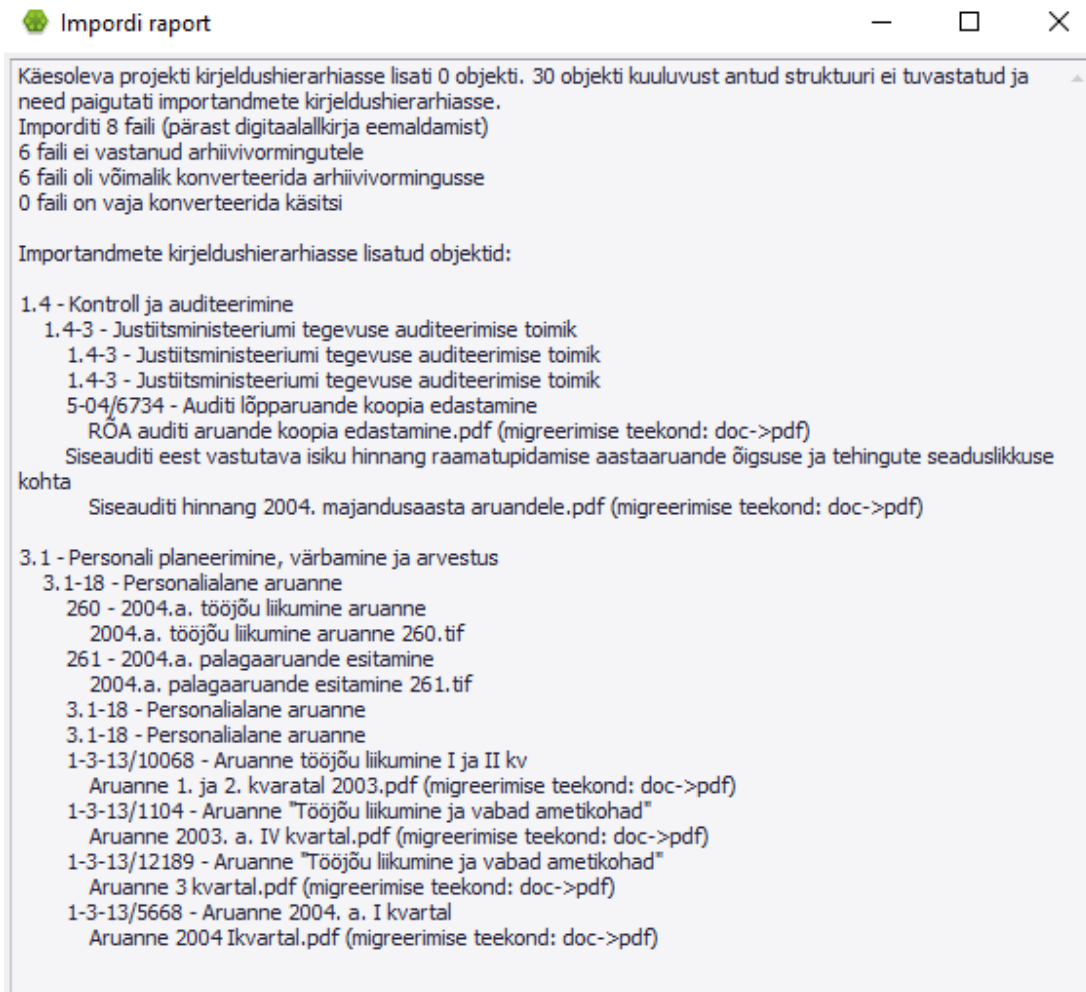


Figure 10. Report of the import to UAM

Subsequently it is possible to re-arrange the imported data by UAM user, e.g re-order case files, add files or metadata according to EAD, etc:

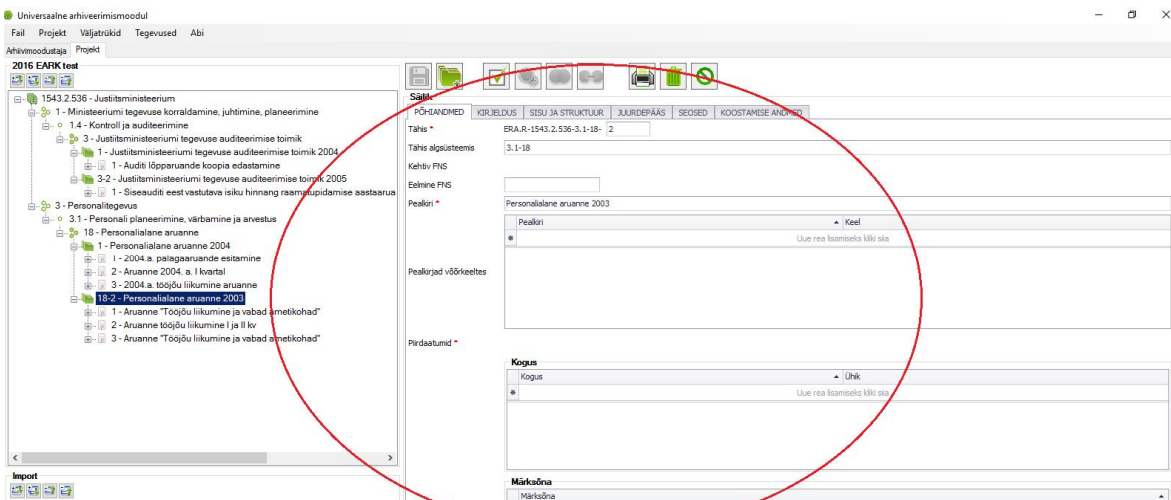


Figure 11. Archival schema after re-arrangements

The next step in UAM is validation, shown in figure 12. In our pilot the system reports about error in transmitter restriction, the cause for access restriction is missing.

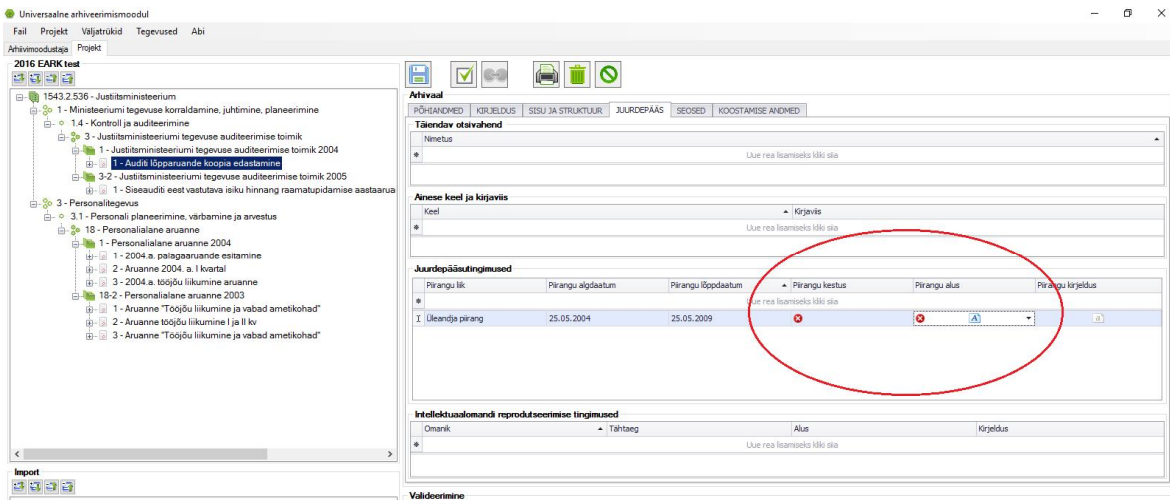


Figure 12. Validation

After validation check it is possible to choose the data carrier, but only in the case of no errors in previous steps. There are two possibilities to export data from UAM to the digital repository of NAE, either via DEC (Document Exchange Center) or using external data carrier.

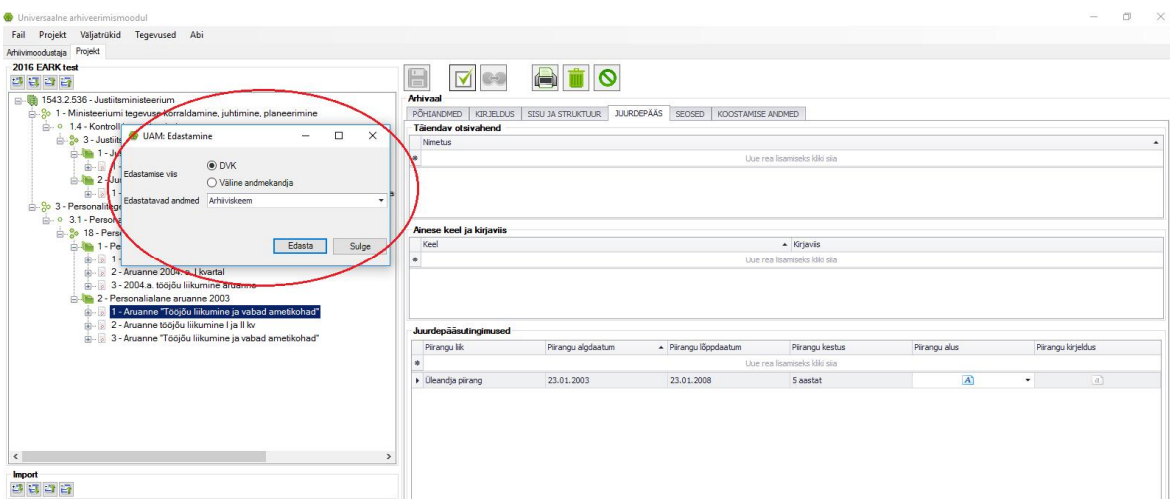


Figure 13. Transfer from UAM

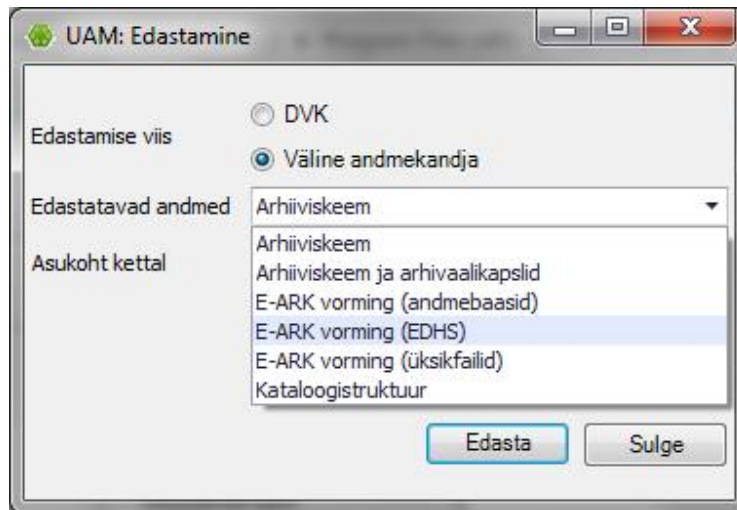


Figure 14. Transfer as E-ARK SIP

Workflow steps in Preservica:

If data has arrived to the repository two workflows in Preservica will be activated automatically:

1. E-ARK Import, which has been configured during E-ARK project. As an example it is seen in the figure 15 that EARK Import has been activated repeatedly, red shows the failed one.
2. Ingest AIP, standard workflow

Date Completed	Workflow Context	Creator	Collection Code	Top Level Record	Size	Files
27.10.16 16:10:15	Ingest AIP (w/o migr.)	Ingest AIP (w/o migr.)	2016-10-27-16-09 +	data	116 KB	2
27.10.16 16:09:51	EARK Import	EARK Import				0
27.10.16 16:09:21	EARK Import	EARK Import				0
27.10.16 16:09:15	Ingest AIP (w/o migr.)	Ingest AIP (w/o migr.)	2016-10-27-16-08 +	data	109 KB	2
27.10.16 16:08:45	EARK Import	EARK Import				0
27.10.16 16:08:21	EARK Import	EARK Import				0
27.10.16 12:38:45	Ingest AIP (w/o migr.)	Ingest AIP (w/o migr.)	2016-10-27-12-37 +	data	769 KB	3
27.10.16 12:37:51	EARK Import	EARK Import				0
27.10.16 12:33:15	Ingest AIP (w/o migr.)	Ingest AIP (w/o migr.)	2016-10-27-12-30 +	data	90 KB	2
27.10.16 12:32:27	Ingest AIP (w/o migr.)	Ingest AIP (w/o migr.)	2016-10-26-17-45 +	data	769 KB	3

Figure 15. Transfer from UAM

After all exported E-ARK SIPs are in NAE repository we can start the second scenario. In the **second scenario (will be piloted in November 2016)** there will be provided access to the exported files. Producers that have ingested data into the archive will have to be able to retrieve them again from the Preservica repository using CMIS protocol and a lightweight access GUI named E-ARK CMIS Browser at the Producer site. Users can now search by browsing exported case files, documents etc in the

whole extent of the exported repository or use “search” command. Link to the tool can be found here: <http://cmis.eark.magenta.dk/#/repository>

Estimated time for the pilot:

The entire process from pre-ingest to ingest and access with 15 files piloted in OCT 27 took 3,5 hours. Time expenditure will not increase remarkably in the bigger amount of files as most of the time was spent to preparation and description activities. Time expenditure per one file will be relatively smaller, estimated time for 1000-1200 documents export will grow up to 4 hours.

The pilot will be continued with testing of bigger data sets in near weeks, also the access part will be tested live.

2.7 Installation instructions

ERMS (DELTA) Export Module

Instructions only available in Estonian as the system is not marketed and used in other countries.

UAM

Before installing UAM, it is advisable to get acquainted with:

- UAM user manual (in English): http://www.arhiiv.ee/public/Digiarhiiv/UAM/UAM_user_manual.pdf
- UAM install manual (in English);
http://www.arhiiv.ee/public/Digiarhiiv/UAM/UAM_install.pdf
- Adjustment instructions on UAM input and compatibility with ERMS (in English).
http://www.arhiiv.ee/public/Digiarhiiv/UAM/UAM_input_configuration.pdf
- UAM demo
[watch demo](#)

E-ARK CMIS Browser

The tool will be available to use in November 2016: <http://cmis.eark.magenta.dk/#/repository>