Background

to the EARK Pilots

E-ARK Final Conference, Budapest

6-8 December, 2016

E-ARK Pilots

1. Full scale project pilot activities

Implementation, by consortium members, of one or more scenarios at one or more locations for a period of six months or longer.

2. Additional project pilot activities

Implementation, by consortium members of shorter ‘stretch’ pilots that extend the scenarios or apply them in different contexts.

3. External validation activities

Implementation of project results by members of DLM Forum and DPC as part of an extended ‘Beta’ program with limited involvement from consortium members.

E-ARK Full-scale Pilots - OAIS Overview

Pilot	Pre-Ingest	Ingest	Archival Storage Preservation	Data Management	Access
1. SIP creation of relational databases (Danish National Archives)	■				
2. SIP creation and ingest of records (National Archives of Norway)	■	■	■		
3. Ingest from government agencies (National Archives of Estonia)	■	■	■		■
4. Business archives (National Archives of Estonia, Estonian Business Archives)	■	■	■		
5. Preservation and access to records with geodata (National Archives of Slovenia)	■	■	■		■
6. Seamless integration between a live DMS and a long-term digital archiving and preservation service (KEEP SOLUTIONS)	■	■	■	■	
7. Access to databases (National Archives of Hungary)	■	■	■	■	■

■ Focus of the pilot
■ Elements also used/tried within the pilot

Pilot 1

SIP Creation on relational databases

Phillip Mike Tømmerholt
Danish National Archives

Pilot 1	SIP Creation on relational databases																														
Task leader	Danish National Archives																														
Scope	The scope of this Pilot is to test the E-ARK SIP Creation tool with not less than 4 databases of different sizes and complexities (one contains several million records)																														
E-ARK Specifications	SIP		AIP		DIP		SMURF (ERMS)		SMURF (SFSB)		GeoData		SIARD 2.0	X																	
E-ARK Tools	Pre-Ingest							Ingest				Access																			
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR index (E-ARK web)	Search and Dsiplay GUI		Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	Database Visualisation Toolkit	IP Viewer				QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
	X																													
Pilot Scenarios																														
Scenario 1	Extracting records from database (Data Set 1) - database with no documents																														
Scenario 2	Extracting records from database (Data Set 2) - database with no documents (large)																														
Scenario 3	Extracting records from database (Data Set 3) - database with documents																														
Scenario 4	Extracting records from database (Data Set 4) - database with documents (large)																														

The Danish Pilot site

- The Danish National Archives (DNA) has been archiving digital records from public institutions since 1973.
- Since 1997 these digital records have also included digital documents in digital document and records management systems submitted as relational databases.

The Danish Pilot site

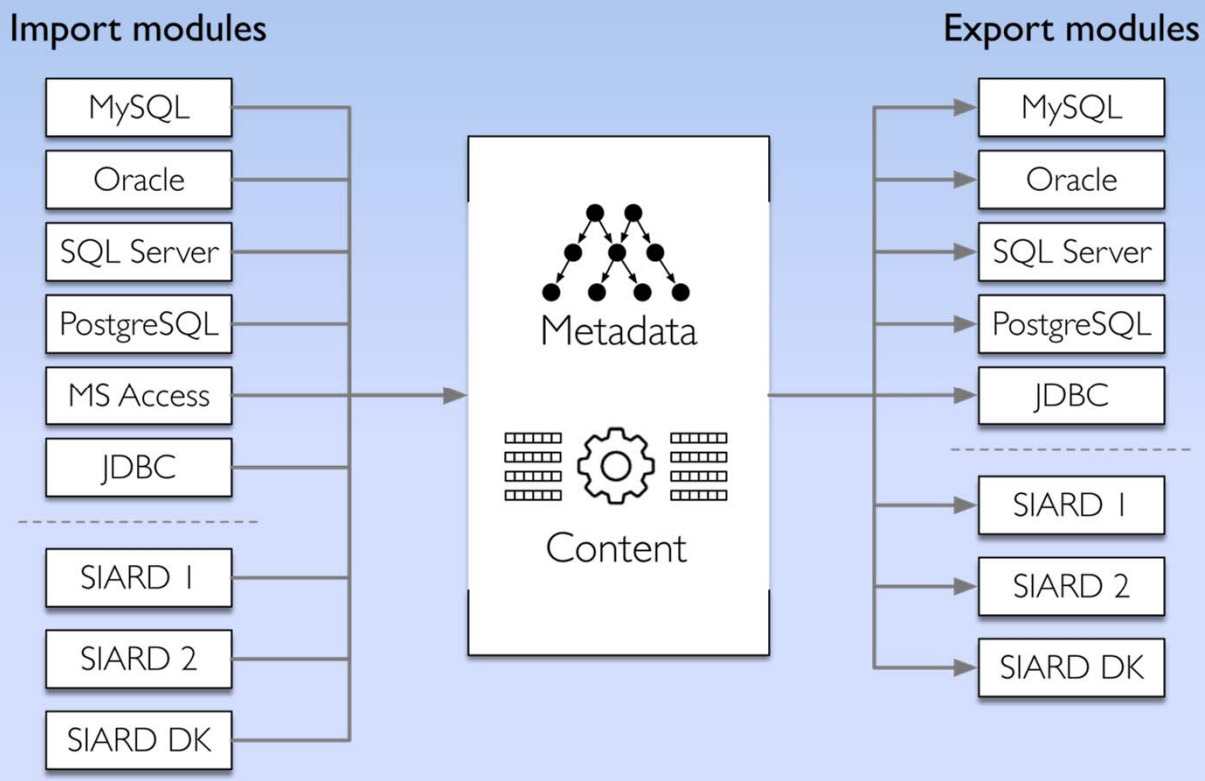
- The DNA has archived more than 4,400 databases.
- The last three years the DNA has received about 220 relational databases each year.
- This year the DNA will ingest 40 terabyte of data from databases.

The Danish Pilot site

- The current SIP format using SIARDDK has been in place since 2010.
- SIARDDK is a slight deviation from the SIARD 1.0 format (created by the Swiss Federal Archives / Enter AG). It was deviated in order to support large amounts of files, a feature now supported by SIARD 2.0.

The tool

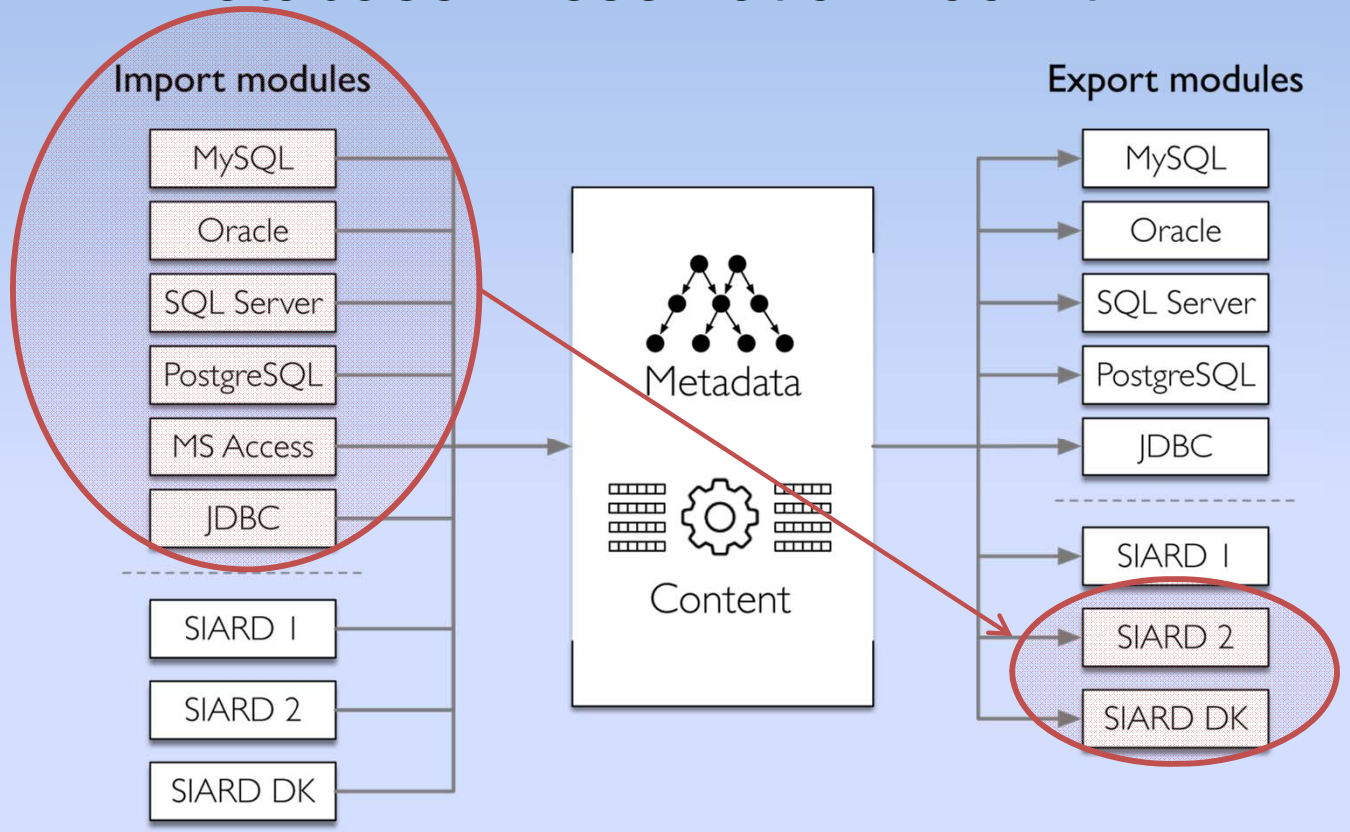
Database Preservation Toolkit



DBPTK can *(via a user logon that have certain rights)* attach to the relational database management system and export to SIARD format

The tool

Database Preservation Toolkit



DBPTK can *(via a user logon that have certain rights)* attach to the relational database management system and export to SIARD format

Scenario 1

Extracting records from database (Data Set 1) - database with no documents

Scenario 2

Extracting records from database (Data Set 2) - database with no documents (large)

Scenario 3

Extracting records from database (Data Set 3) - database with documents

Scenario 4

Extracting records from database (Data Set 4) - database with documents (large)

Technical details	System
MS SQL Server 90 tables 90,000 records	**Health system from the Danish Serum Institute** Database containing information from reported infectious diseases at a the national level. Infectious diseases for all Danish citizens.
MySQL 5 tables 33,000,000 rec.	**System from a private sector business. Kultunaut Aps** Harvesting and selling information about cultural events at national level from the smallest local gatherings to international exhibitions and events in Denmark.
MS SQL Server 289 tables 1,300,000 LOBs	**Administrative system from The Danish National Archives** Database containing information about all incoming scientific research data, and public deliveries of research data.
MS SQL Server 180 tables 100,000 LOBs	**Administrative and health records system from Ministry of Higher Education and Science** Database containing information about social, psychological, and psychiatric counseling to students in their educational situation.

The results

The results from the pilot show that the E-ARK project open source Database Preservation Toolkit (DBPTK) made by project partner KEEP Solutions can export live databases in proprietary formats to system independent archiving formats SIARD 2.0 and SIARDDK (the latter provided by E-ARK project partner Magenta).

This open source tool can ease the export of databases by decreasing the cost and time spent for producers of SIPs as well for the archives due to fewer errors being submitted.

The results

In the pilot we only validated SIARD 2.0 manually. Since SIARD DK and SIARD 2.0 specification are very close to each other, the DNA's internal validating tool was able to locate some bugs in the SIARD 2.0 products.

We still need a tool to validate SIARD 2.0

Pilot 2

SIP creation and ingest of records

Arne-Kristian Groven, Hans Fredrik Berg
National Archives of Norway

Pilot 2	SIP creation and ingest of records
Task leader	National Archives of Norway
Scope	Not less than 2 transfers of unstructured records with mixed restricted and unrestricted material, and not less than 1 transfer of structured records.

E-ARK Specifications	SIP	X	AIP	X	DIP		SMURF (ERMS)	X	SMURF (SFSB)		GeoData		SIARD 2.0		

E-ARK Tools	Pre-Ingest							Ingest				Access																		
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR index (E-ARK web)	Search and Dsiplay GUI		Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	Database Visualisation Toolkit	IP Viewer				QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
					X					X																				

Pilot Scenarios	
Scenario 1	SIP Creation and Ingest of unstructured records (Data Set 1)
Scenario 2	SIP Creation and Ingest of unstructured records (Data Set 2)
Scenario 3	SIP Creation and Ingest of structured records (Data Set 3)

Pilot 2: Goals

- Use ESSArch tools on production data sets
  - Containing a substantial number of digital records
- To evaluate
  - The ability to get a job done with a satisfactory result
- Quantitative measurements
  - Time consumption
- Preferably
  - Use the E-ARK package definitions

Data Sets

- Scenario1: The data is the standardized output from an EDRMS system(20 GB)
  - Noark-4 XML file containing (meta-) data descriptions + documents, mainly PDF/A files.
- Scenario 2: The data is the standardized output from an EDRMS system (5 GB)
  - Noark-5 XML file containing (meta-) data descriptions + 8194 documents, mainly PDF/A files.
- Scenario 3: 10 CSV files, originating from an old database from the 1980s (105 MB)
  - Containing 338.500 registrations.

ESSArch Tools

- ESSArch Tools for Producer, ETP
  - Prepare content, select SIP structure, archival descriptions/ generate metadata
  - Submit the generated SIPs
- ESSArch Tools for Archivists, ETA
  - Receive SIPs from producer, prepare for ingest
- ESSArch Preservation Platform, EPP
  - Ingest (and store) the SIPs, derive AIPs, storage management, either on tape and/or disk storage

The time used on the automated parts of these operations

- Scenario 1 (20 GB)
  - 1 h 50 min
- Scenario 2 (5 GB)
  - 1 h
- Scenario 3 (105 MB)

To summarize our experience

- The tools we evaluated were able to produce satisfactory results (output).
- But we would like to evaluate on even larger data sets to conclude about scalability/time consumption
- The evaluated tools have implemented the METS and PREMIS standards as XML structures, which is a good thing.
- Suggestions to further improvements: The ability to provide standards like DC, ISAD(G), ISAAR etc.
- And also to provide import/export on standards like EAD or DC would also be positive.

Final comments…

  - When the pilot goals for our pilot were defined at the beginning of the project,
- the direction of the E-ARK work regarding OAIS package structure implementations was still not known to us.
  - We hoped it would imply minor changes compared to OAIS package structure we had defined at the National Archives of Norway a few years earlier.
- But the resulting E-ARK package structure represented more substantial differences than anticipated.
  - ES Solutions has provided us with two sets of package structures, one for the production scenarios . Another for additional scenarios using the E-ARK package structures.

Final comments

- The resulting E-ARK package structure is for us an obstacle, preventing us from using E-ARK software
- If, in the final months of the project, the package structure could be made more generic it would have the potential to reach more European users
- … Including the National Archives of Norway

Pilot 3

Ingest from governmental agencies

Karin Oolu, Tarvo Kärberg
National Archives of Estonia

Pilot 3	Ingest from government agencies
Task leader	National Archives of Estonia
Scope	Export public records from an EDRM system of a governmental agency to the National Archives of Estonia and make these available through our own catalogue (i.e. Archival Information System, AIS) as well as provide an API for accessing the records from other systems (the original EDRMS at the agency)

E-ARK Specifications	SIP	X	AIP		DIP	X	SMURF (ERMS)	X	SMURF (SFSB)		GeoData		SIARD 2.0	X	

E-ARK Tools	Pre-Ingest							Ingest				Access																		
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository ESSArch Preservation	Platform HDFS-Storage	SOLR index (E-ARK web)	Search and Dsiplay GUI		Order Management Tool	Lily - Ingest ESSArch Preservation	Platform E-ARK Web (Search)	AIP2DIP (E-ARK Web) Database Visualisation	Toolkit IP Viewer				QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
						X																					X

Pilot Scenarios	
Scenario 1	Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica
Scenario 2	Provide access to records from governmental institution through CMIS portal/viewer

Scenario 1:

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

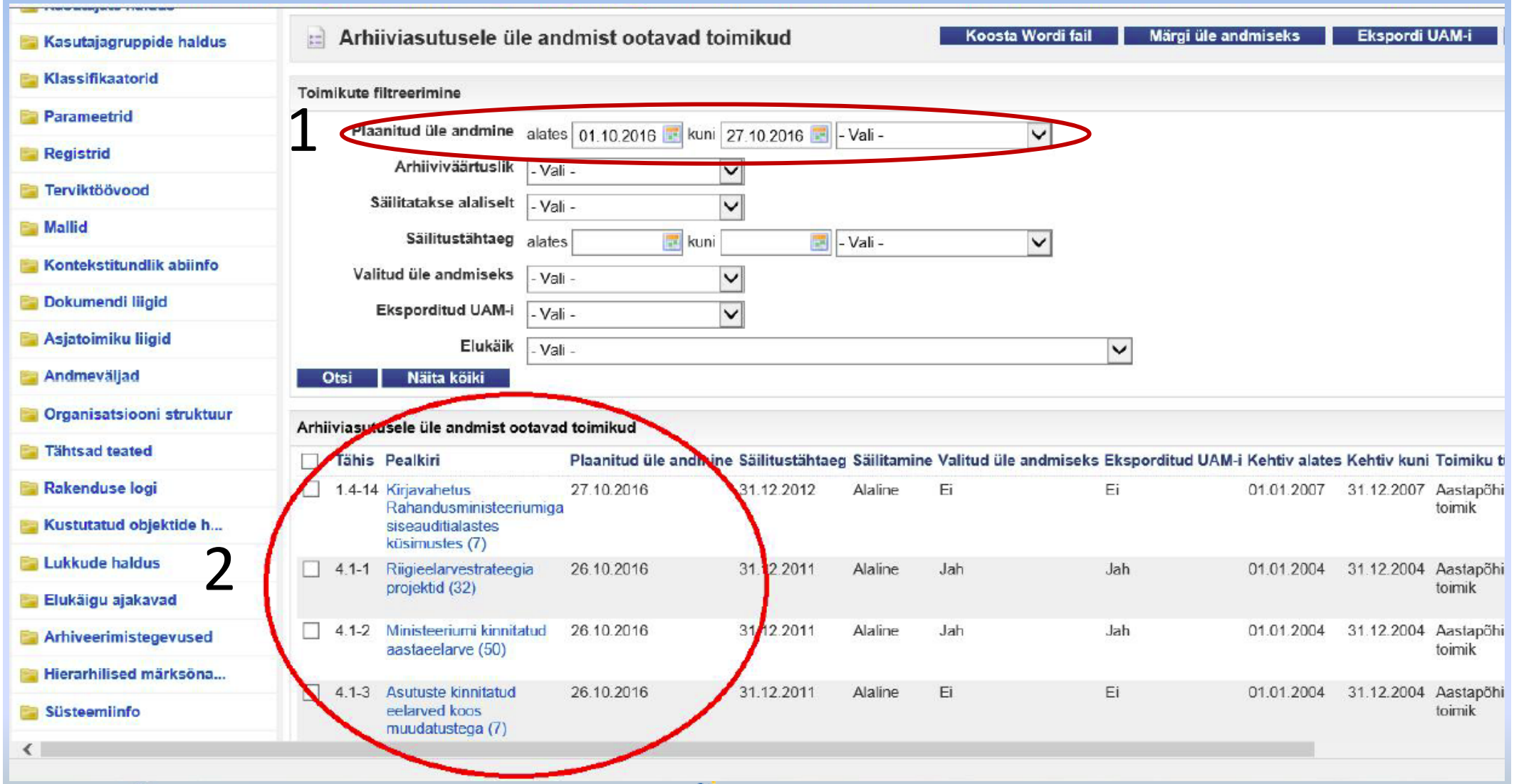
Main steps

Data **selected and extracted** from the DELTA system in the Ministry of Justice of Estonia

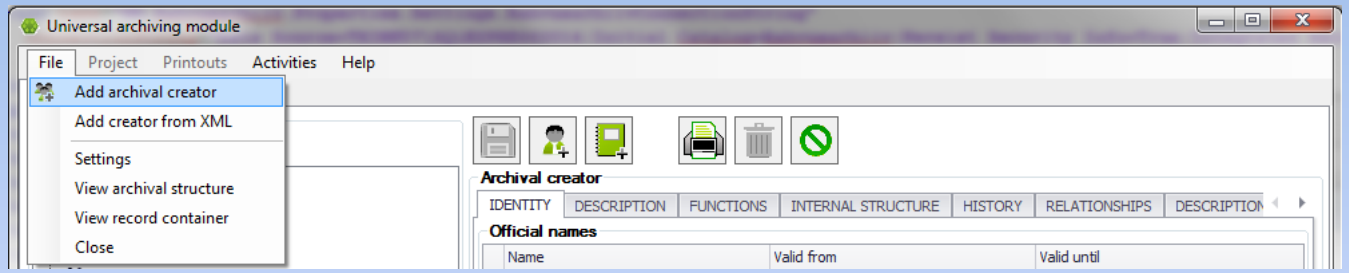
Data **imported** to the Universal Archival Module (UAM) to **create E-ARK SIPs**

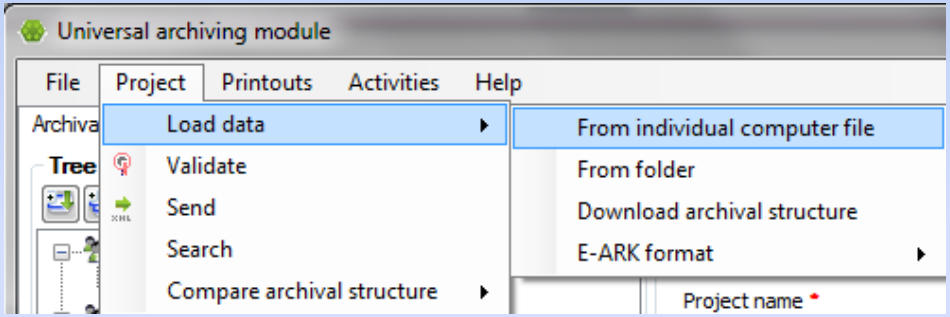
Data **ingested** to Preservica (at the National Archives of Estonia)

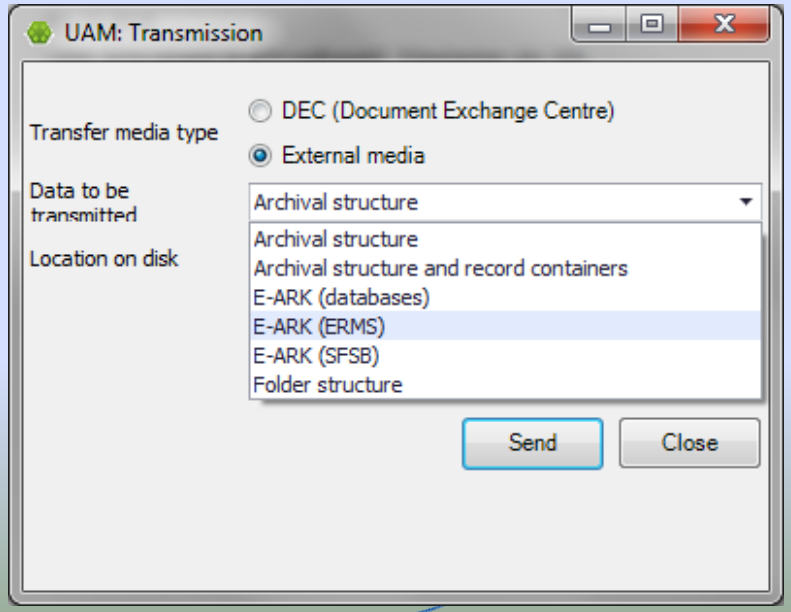
Selection and extraction



SIP Creation



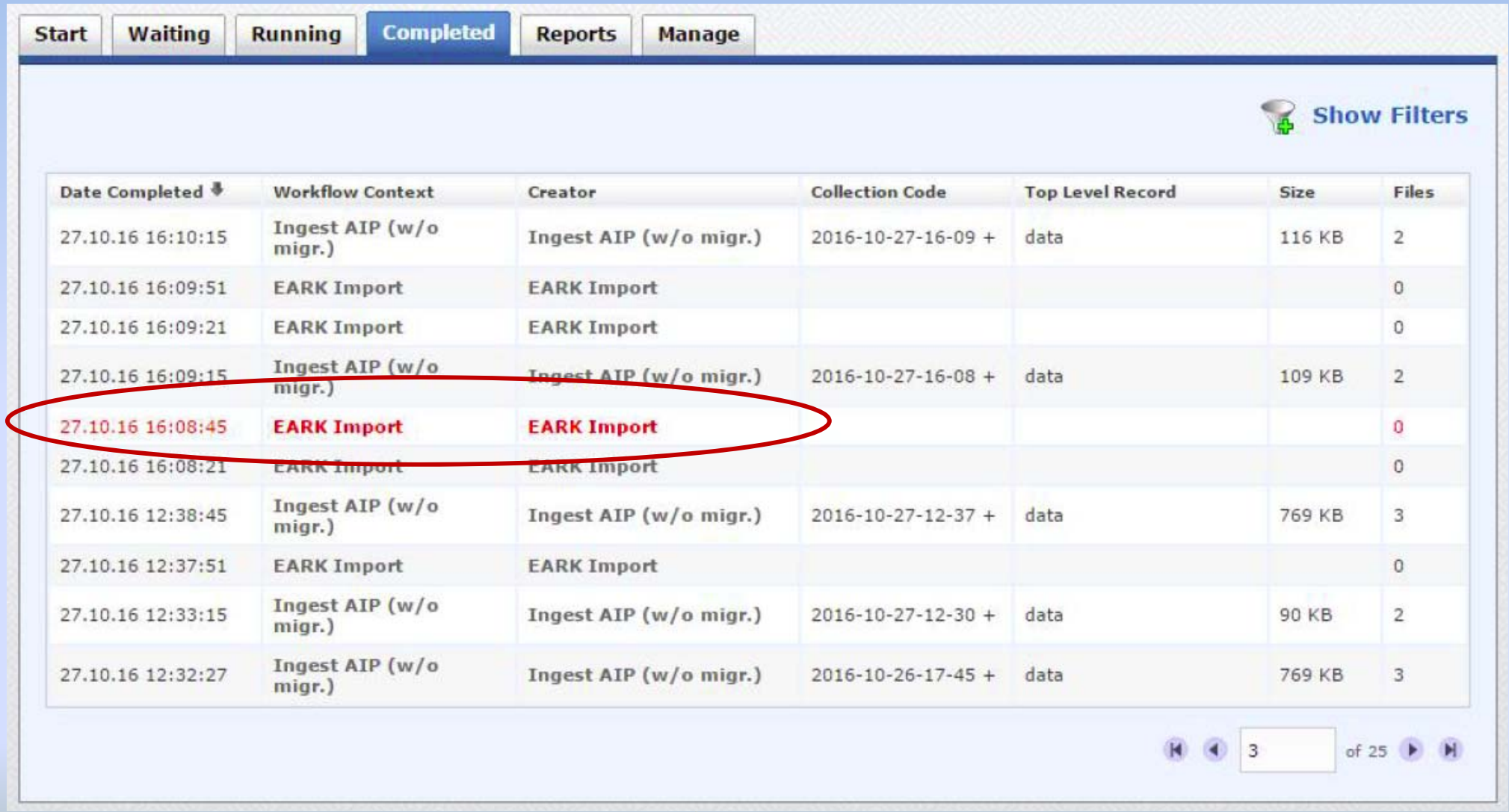




Statistics

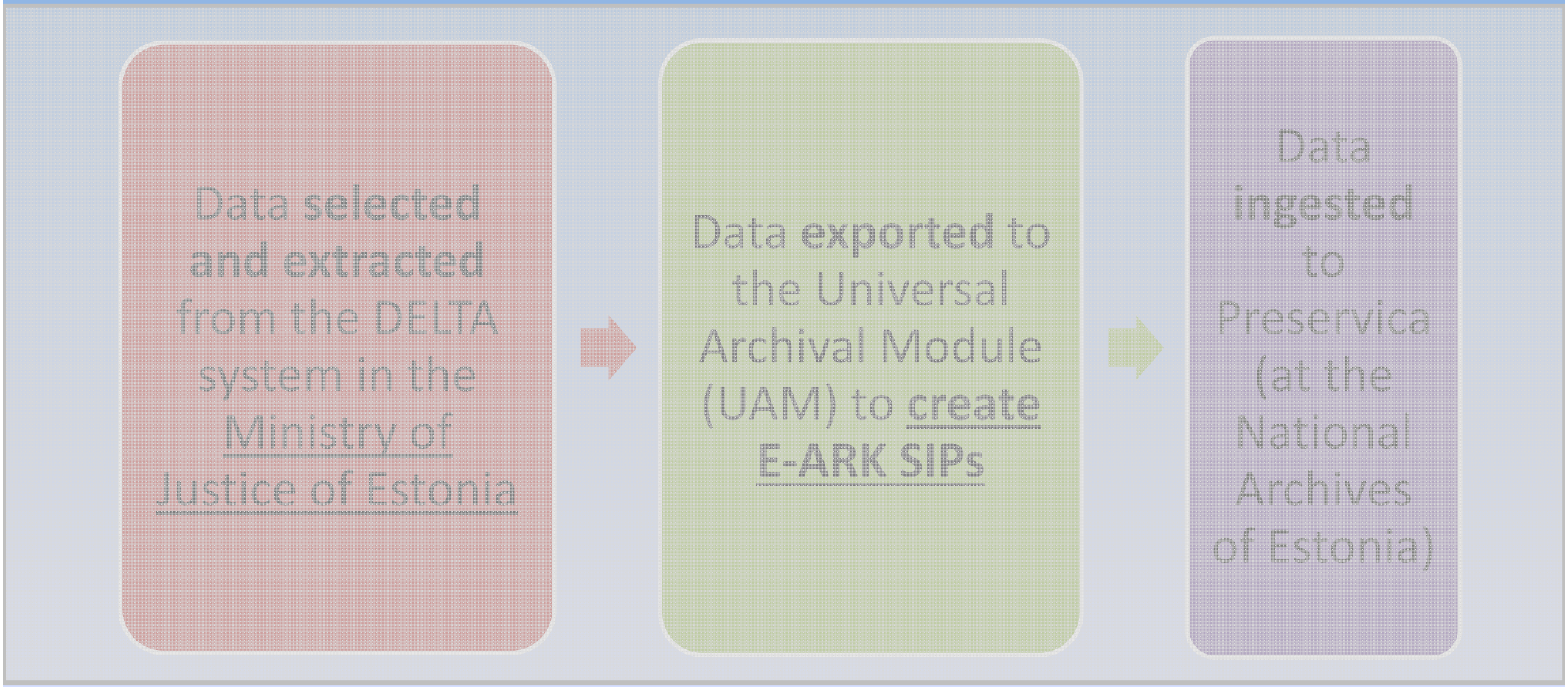
- Two different data-sets:
  - first test on Oct 21, (**15 files**)
  - second test on Nov 10, (**200 files**)
- Selection, extraction, SIP creation
  - first test, duration ca **3 hours**
  - second test, duration ca **15 minutes**

Ingest to Preservica



Scenario 2:

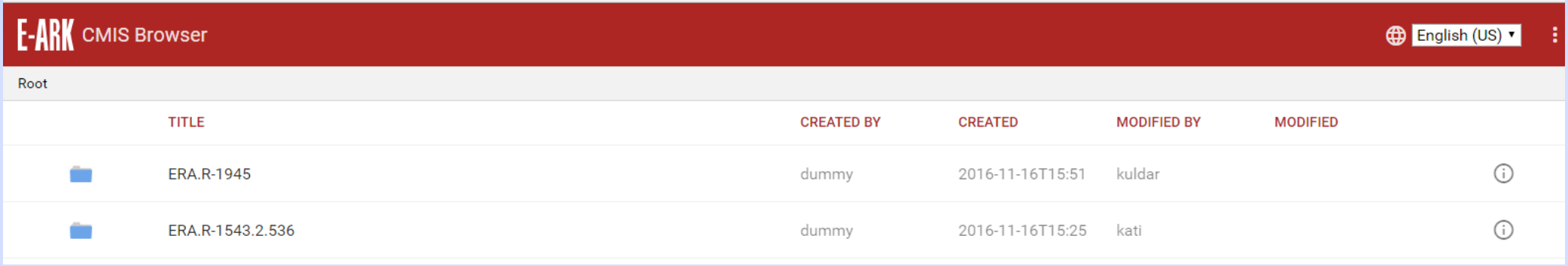
Provide access to records from governmental institution through RESTful services



Provide access to ingested records from governmental institution using **CMIS protocol**

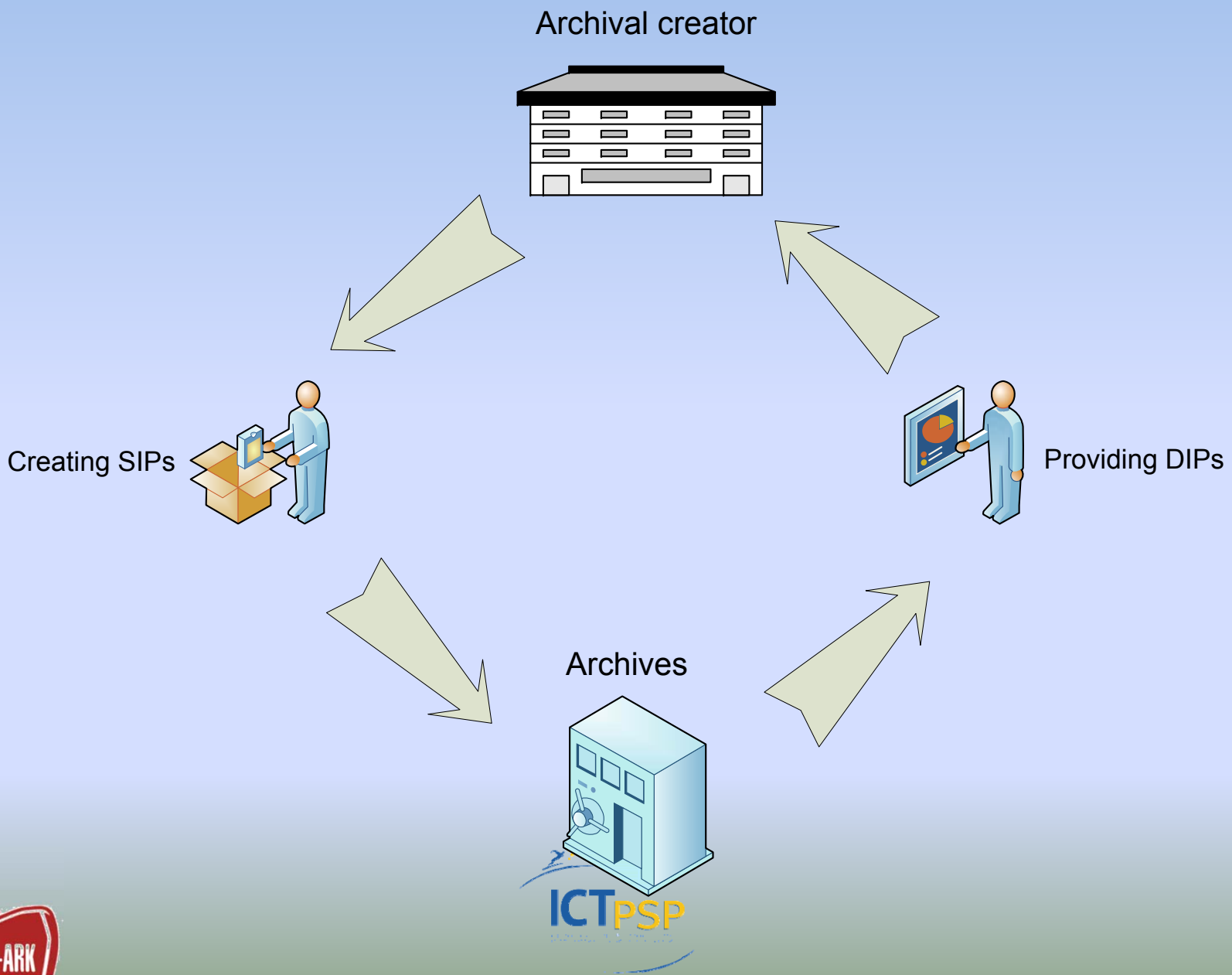
E-ARK CMIS Browser

E-ARK tool for retrieving records from repositories using CMIS protocol and a lightweight access GUI.



- configure repository
- manage users
- search and browse records

Scenario 1 + scenario 2



Eesti ÄriARHIIV
ESTONIAN BUSINESS ARCHIVES

Pilot 4

Business Archives

Ats Rand, Estonian Business Archives

Tarvo Kärberg, National Archives of Estonia

E-ARK

ICTPSP

Pilot 4	Business Archives
Task leader	Estonian Business Archives with support by National Archives of Estonia
Scope	Pre-ingest preparation and transfer of business records to a digital archive solution in a business archive

E-ARK Specifications														
SIP		AIP		DIP		SMURF (ERMS)		SMURF (SFSB)		GeoData		SIARD 2.0	X	

E-ARK Tools	Tool	
Pre-Ingest	Database Preservation Toolkit	X
	ERMS Export Module	
	RODA-In	
	ESSArch Tool Producer (ETP)	
	ESSArch Tools Archive (ETA)	
	UAM	
	SIP creator (E-ARK Web)	
Ingest	SIP2AIP (E-ARK Web)	
	RODA Repository	
	ESSArch Preservation Platform	
	HDFS-Storage	
Access	SOLR index (E-ARK web)	
	Search and Dsiplay GUI	
	Order Management Tool	
	Lily - Ingest	
	ESSArch Preservation Platform	
	E-ARK Web (Search)	
	AIP2DIP (E-ARK Web)	
	Database Visualisation Toolkit	
	IP Viewer	
	QGIS	
	Geoserver	
	Peripleo	
	Oracle (OLAP Viwer)	
	CMIS portal/viewer	

Pilot Scenarios	
Scenario 1	Migration and Ingest of business data from bespoke business system (Data Set 1)
Scenario 2	Import business records from SIARD 2.0 (Data Set 1)
Scenario 3	Migration and Ingest of business data from bespoke business system (Data Set 2)
Scenario 4	Import business records from SIARD 2.0 (Data Set 2)

The objective

To test the export functionality of business records from a bespoke business system to the digital archive system of the Estonian Business Archives (EBA) by using the Database Preservation Toolkit.

Main steps

Preparation

- Preparation phase is needed in order to ensure archiving of the correct data. For example, that there are no ongoing procedures or daily back-up routines or other actions which can be interrupted or would corrupt data turning the export process.
- Client provides access to the database - this includes making the database accessible to the Database Preservation Toolkit.

Extraction

- EBA IT staff and client's IT staff work together to set up proper configuration to start export data from database with Database Preservation Toolkit.
- The same configuration should be used for every export instance. Any modification should be documented and made public for both sides.

Transfer

Data is transferred over HTTPS using the EBA digital archive solutions' RESTful API and web user interface to submit SIARD2.0 files to the endpoint.

Id	Name
1	"test_siard.siard"
2	"ico_mssqlprod_20161020.siard"
3	"ico_mssqlprod_20161020_1.siard"
4	"ico_mssqlprod_20161020_2.siard"

Access

- Import process is started with Database Preservation Toolkit to any available database platform.
- Preferred database will always be initial database where data was exported from.
- When import is successful, the data is obtained and used through Microsoft SQL Server Management studio or other management studios PL/SQL Developer, phpMyAdmin web interface, depending on database platform.

About the client side database system

- The system generates reports and statistical information based on analyses and medical observations.
- The System contains 63 tables with approximately **200 000 records** with 33 491 files. The size of the database is approximately **160 GB**.
- Data type: MS-SQL.

Time

- The entire process for pre-ingest took roughly **6 hours**.
- Most of the time was spent on preparation and description activities.

Pilot 5

Preservation and access to records with geodata

Gregor Završnik
National Archives of Slovenia

Pilot 5	Preservation and access to records with geodata
Task leader	National Archives of Slovenia
Scope	Pilot will prove that the SIP and DIP implementations fulfill specific requirements for the records containing GIS data, test the instructions (for the producer and for the archive) regarding all phases of ingest, to prove that the archival use of GIS data is possible.

E-ARK Specifications	SIP	X	AIP	X	DIP	X	SMURF (ERMS)		SMURF (SFSB)		GeoData	X	SIARD 2.0		

E-ARK Tools	Pre-Ingest							Ingest				Access																			
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR index (E-ARK web)	Search and Dsiplay GUI		Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	Database Visualisation Toolkit	IP Viewer				QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer	
			X	X	X			X		X		X	X		X	X	X	X	X		X				X	X	X		

Pilot Scenarios	
Scenario 1	SIP Creation and Ingest of records with Geodata (2 representations)
Scenario 2	Search and Access information using Geodata (2 representations)
Scenario 3	SIP Creation and Ingest of records with Geodata (2 representations)
Scenario 4	Search and Access information using Geodata (2 representations)

What we promised?

During the e-ARK project the standardized method for ingesting geo data will be developed.

This will allow the archives to

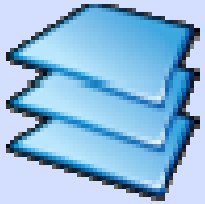
- offer geodata as a selection and display criteria of records
- by means of integration of current state of the art tools.

Presenting Geodata

GIS Application

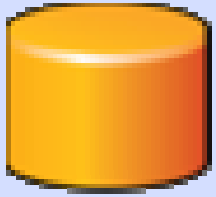
Symbology, Geoprocessing, Exports

Sorting , Querying, Cartog. Projection, transformation

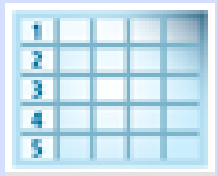


Vector layers

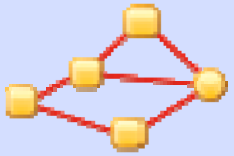
Raster layers



Databases



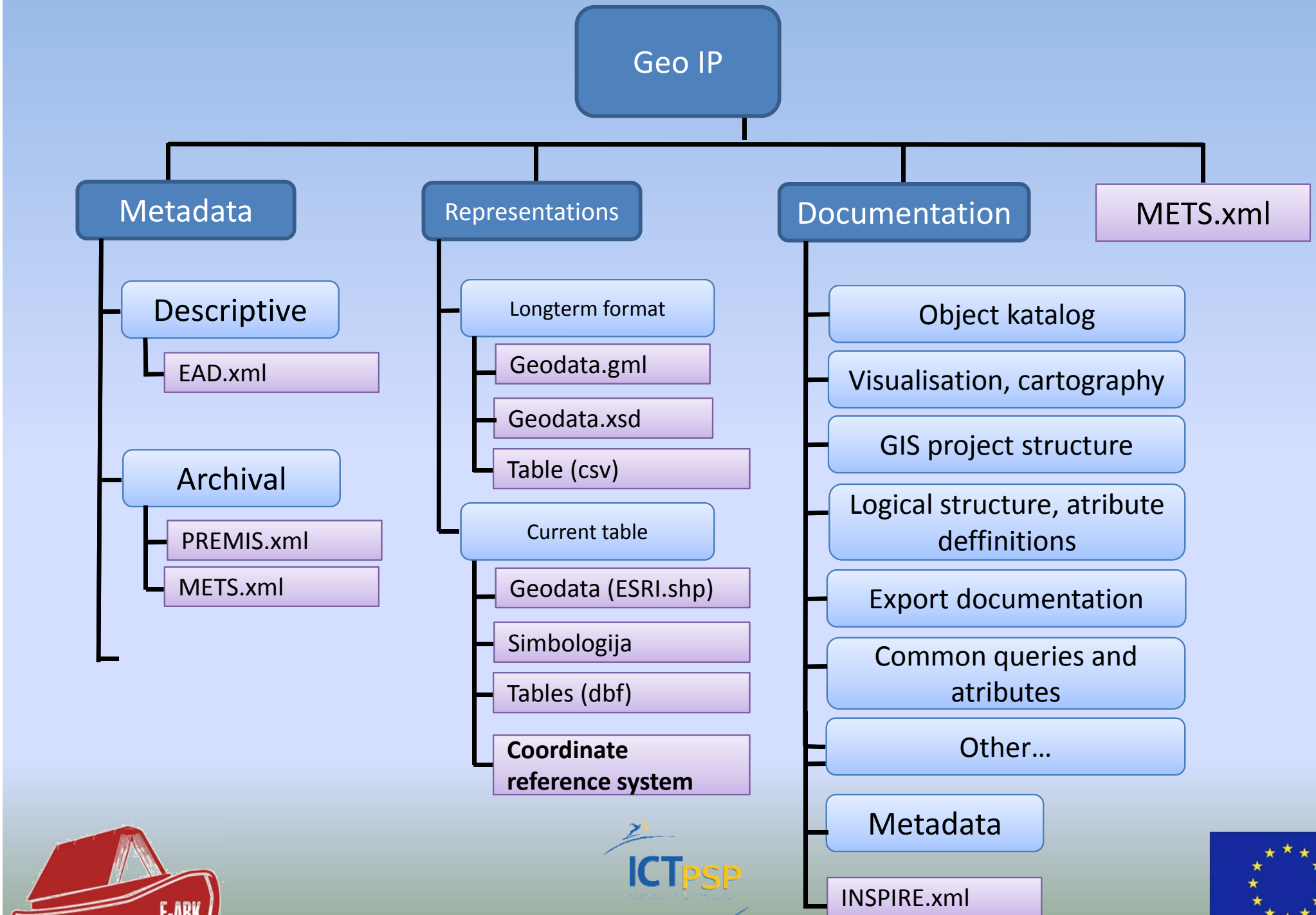
Codepages spreadsheets

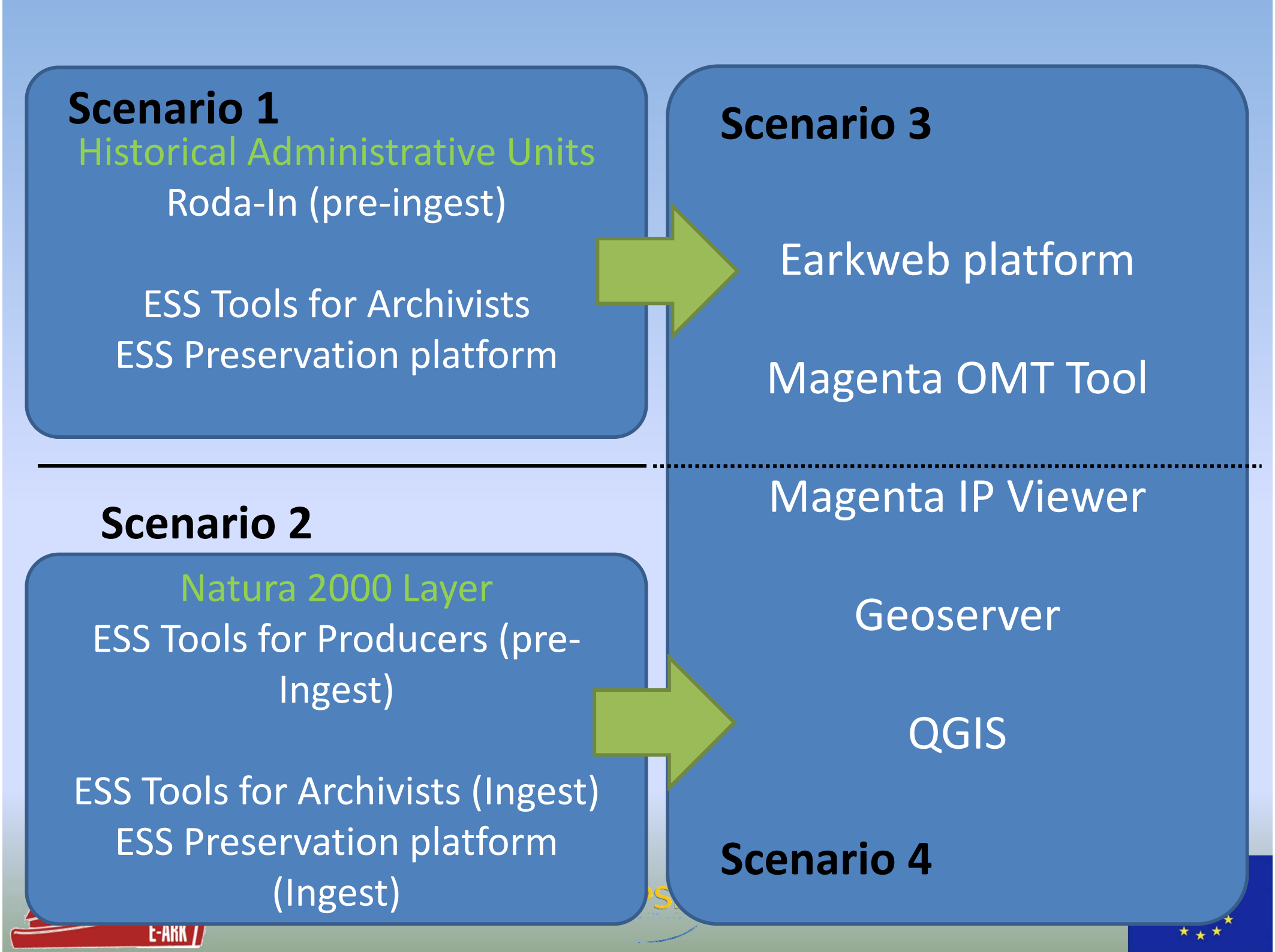


Models

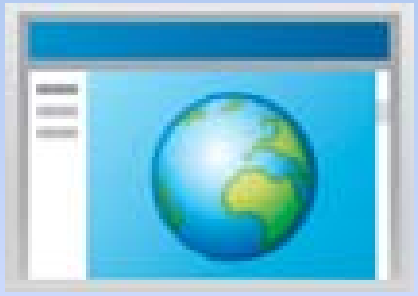
E-ARK

ICTPSP


Geo IP
Metadata
Representations
Documentation
METS.xml
Descriptive
EAD.xml
Archival
PREMIS.xml
METS.xml
Longterm format
Geodata.gml
Geodata.xsd
Table (csv)
Current table
Geodata (ESRI.shp)
Simbologija
Tables (dbf)
Coordinate reference system
Object katalog
Visualisation, cartography
GIS project structure
Logical structure, atribute deffinitions
Export documentation
Common queries and atributes
Other…
Metadata
INSPIRE.xml
E-ARK
ICTPSP

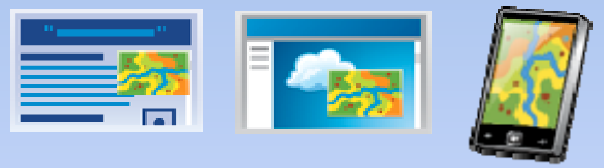

Scenario 1
Historical Administrative Units
Roda-In (pre-ingest)
ESS Tools for Archivists
ESS Preservation platform
Scenario 3
Earkweb platform
Magenta OMT Tool
Magenta IP Viewer
Scenario 2
Natura 2000 Layer
ESS Tools for Producers (pre-Ingest)
ESS Tools for Archivists (Ingest)
ESS Preservation platform (Ingest)
Geoserver
QGIS
Scenario 4

DISPLAYING DATA



Browser

GIS app (QGIS)



Web APPs

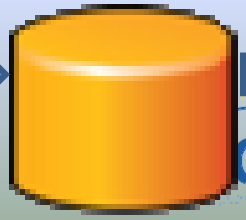
- Peripleo
- Other GIS apps
- Google Maps

DIP 0 > DIPu

Geoserver

Files

Databases



CTPSP

Pilot 6

Integration between a semi-active records management system and a long-term preservation service

Miguel Ferreira
KEEP Solutions

Pilot 6	Integration between a semi-active records management system and a long-term preservation service
Task leader	KEEP Solutions
Scope	The goal of this pilot is two-fold. On one hand, KEEP SOLUTIONS will demonstrate that the pan-European SIP structure designed in the WP3 is adequate to support the media types found in today's Electronic Records Management Systems (e.g. text documents, video, audio, images, etc) and, on the other hand, that the most adequate and scalable form of ingest is to automate the SIP creation and delivery process to the preservation service.

E-ARK Specifications	SIP		AIP		DIP		SMURF (ERMS)		SMURF (SFSB)		GeoData		SIARD 2.0		
		X		X											

E-ARK Tools	Pre-Ingest							Ingest				Access																	
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR index (E-ARK web)	Search and Dsiplay GUI		Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	Database Visualisation Toolkit	IP Viewer				QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
									X																				

Pilot Scenarios	
<u>Scenario 1</u>	Automatic ingest of records from a semi-active archival management system

Objectives

- Assess the efficacy of the E-ARK SIP specification to support the transfer of content and metadata between ERMS and a long-term preservation service
- Demonstrate that the most scalable form of transfer is to automate the SIP creation and delivery to the preservation service

How

- Develop a specially-designed tool to support the:
  - Extraction, aggregation and creation of SIPs
  - Transfer of SIPs to the preservation service
  - Start the ingest process

Success criteria

- Ingest no less than 900 records in E-ARK SIP format (90% success rate)
- Completeness of data and metadata should be 100%

Stakeholders

- Mafra Municipality
  - The data provider
- KEEP SOLUTIONS
  - The solution provider

A collection of digitised books related to the Peninsular War dating from 1778 to 1834.

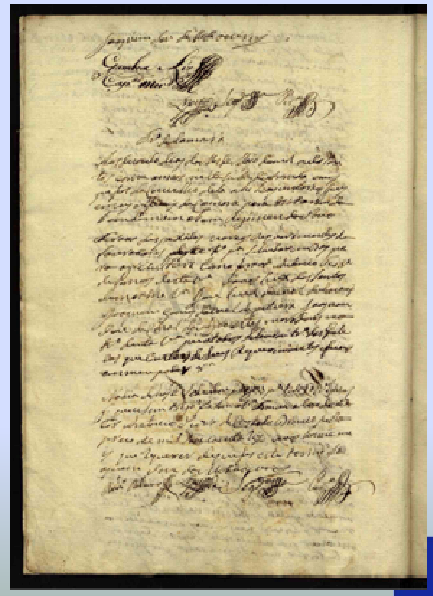
964 records

Metadata in EAD 2002

34.600 pages

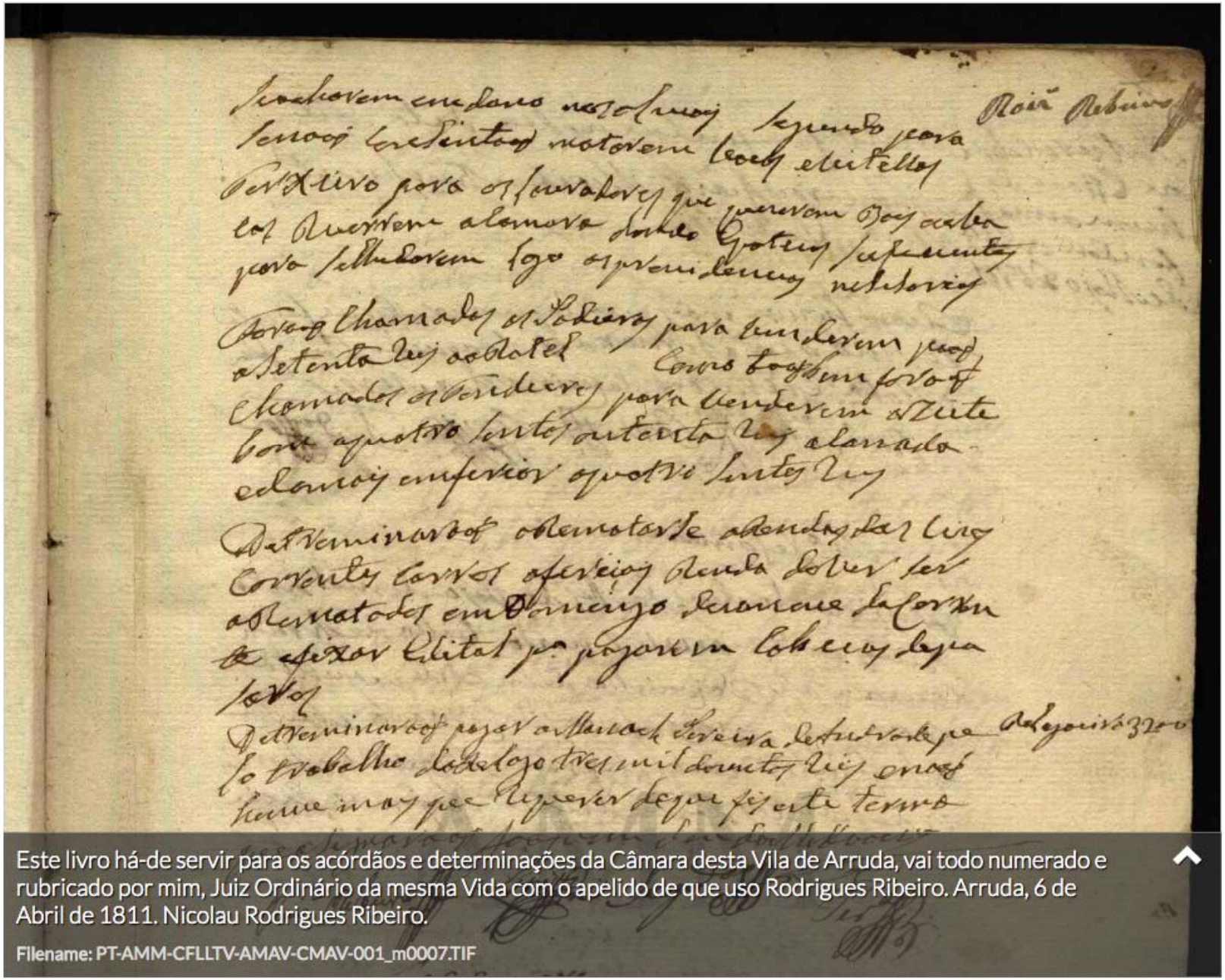
1.2 TB

300 dpi uncompressed TIFFs



E-ARK

ICTPSP



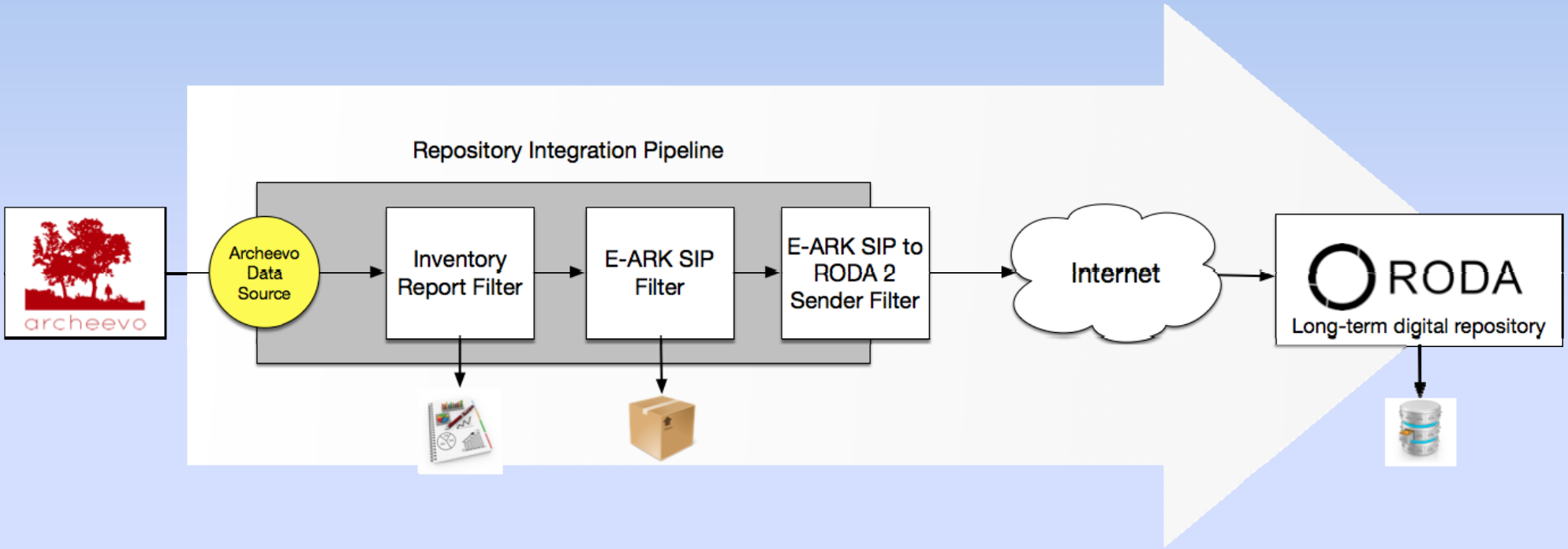
Este livro há-de servir para os acórdãos e determinações da Câmara desta Vila de Arruda, vai todo numerado e rubricado por mim, Juiz Ordinário da mesma Vida com o apelido de que uso Rodrigues Ribeiro. Arruda, 6 de Abril de 1811. Nicolau Rodrigues Ribeiro.

Filename: PT-AMM-CFLLTV-AMAV-CMAV-001_m0007.TIF

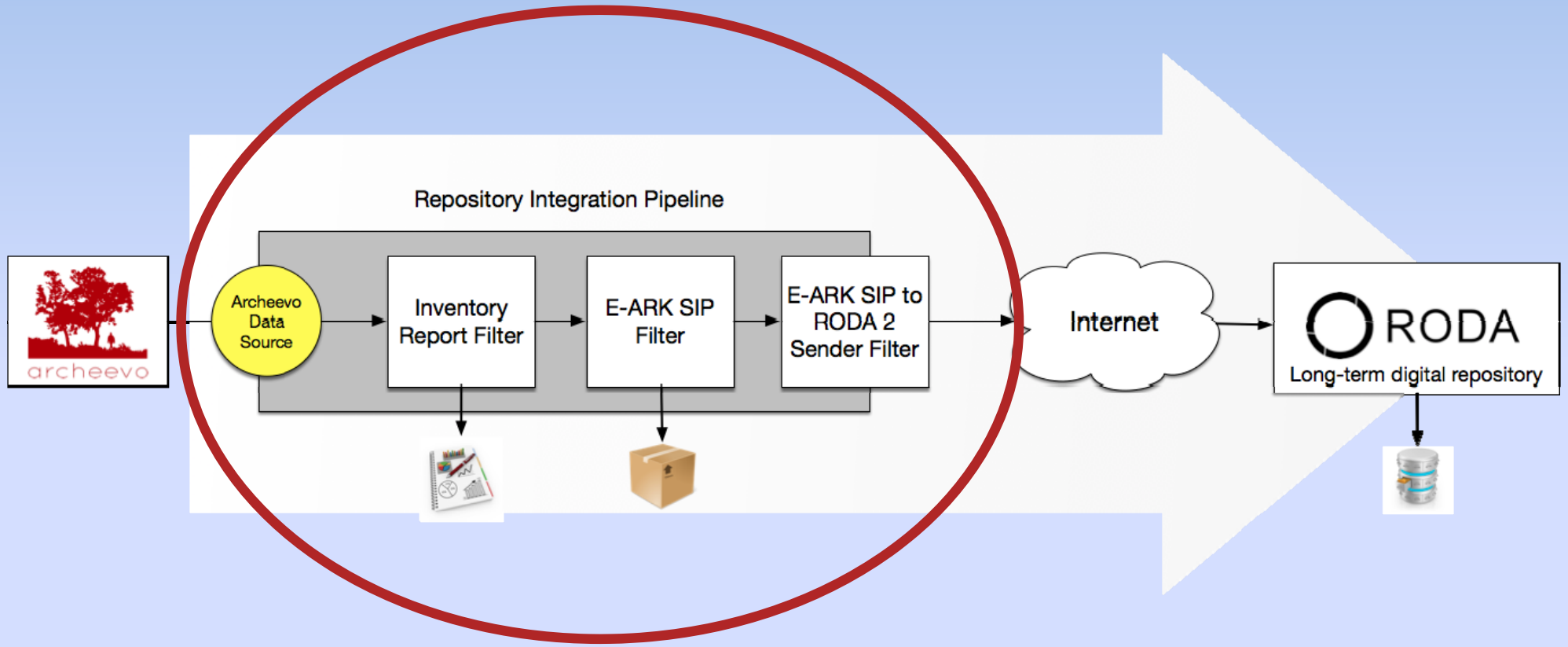
Software components

- Archeevo
  - Archival Management Software at the data provider
- RODA
  - Digital preservation service
- Repository Integration Pipeline
  - Application that runs a sequence of tasks that transform data from one repository system to another

Repository Integration Pipeline



Repository Integration Pipeline



Colecção de Fontes Locais das Linhas de Torres Vedras

65c907a8-1a82-4090-b804-839c9add2ce1
173037

/ Colecção de Fontes Locais das Linhas de Torres Vedras

Encoded Archival Description 2002

Identity

Reference code
PT/AMM/CFLLTV

Title
Colecção de Fontes Locais das Linhas de Torres Vedras

Initial date
1778

Final date
1834

Descriptive date
1778/1834

Country code
PT

Archival package
NEW
MOVE
PERMISSIONS
REMOVE

Preservation
START NEW PROCESS
EVENTS
RISKS
LOGS

Download
ARCHIVAL PACKAGE
SCHEMAS

Search
IN THIS CONTEXT

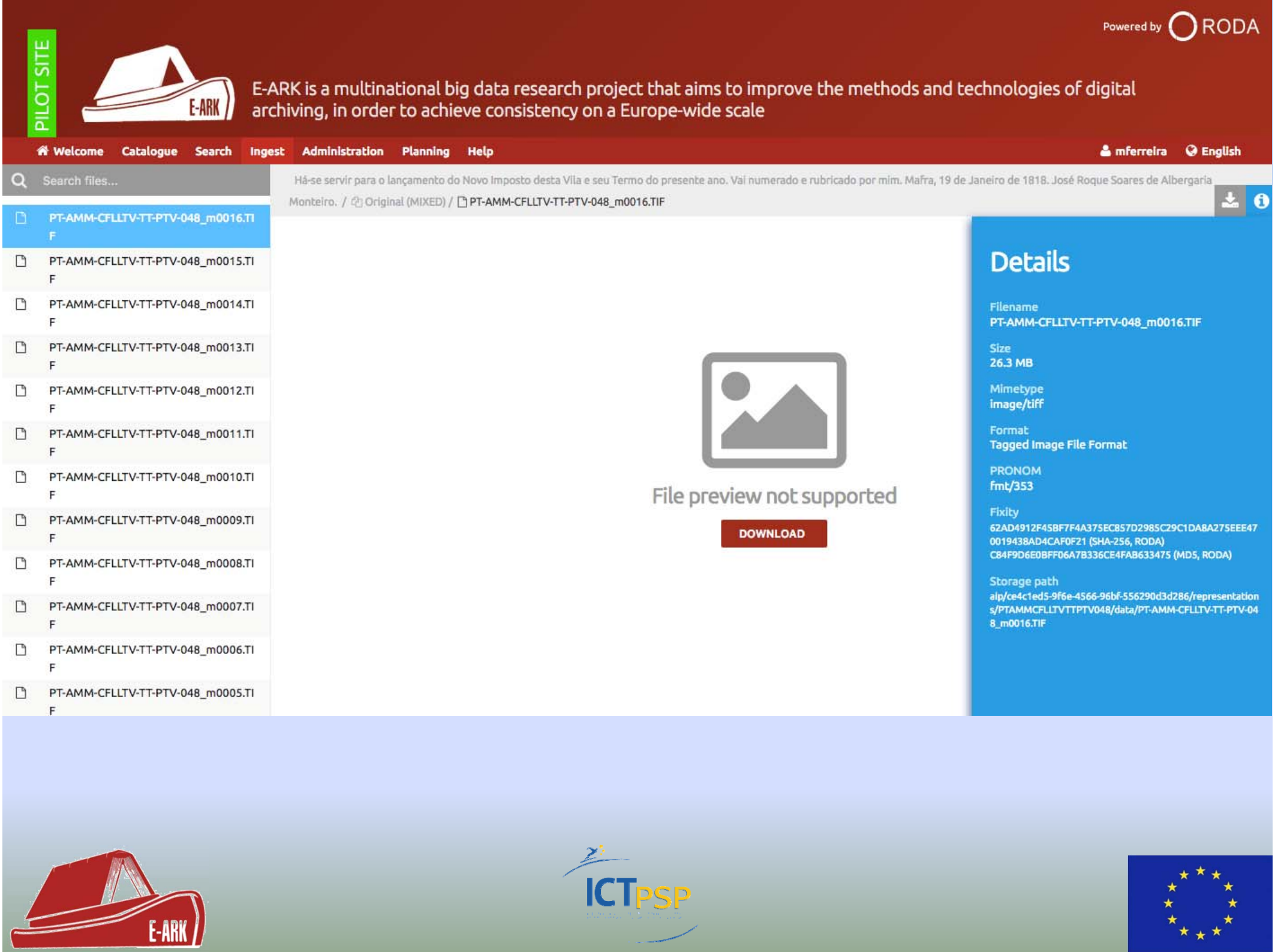
Search

In this page you can search for Intellectual Entities, Representations or Files (use the down arrow to select the search domain). For each one of these domains you can search in all its properties or in specific properties (use the down arrow to expand the advanced search). For example, if you select Intellectual Entities, you can search in a specific field of the descriptive metadata, or find files of a certain format if the Files advanced search is selected.

The search engine locates only whole words. If you want to search for partial terms you should use the '*' operator. For more information on the available search operators, take a look at the help page.

Representations | Search...

Id	Original	Type	Size	Files	Documentation	Schemas
PTAMMCFLLTVTTPTV048	original	MIXED	384 MB	16 files	0 files	0 files
PTAMMCFLLTVTTPTV047	original	MIXED	2.7 GB	96 files	0 files	0 files
PTAMMCFLLTVTTPTV046	original	MIXED	349.6 MB	12 files	0 files	0 files
PTAMMCFLLTVTTPTV045	original	MIXED	359.5 MB	12 files	0 files	0 files
PTAMMCFLLTVTTPTV044	original	MIXED	3.7 GB	131 files	0 files	0 files
PTAMMCFLLTVTTPTV043	original	MIXED	352.3 MB	14 files	0 files	0 files
PTAMMCFLLTVTTPTV042	original	MIXED	233.8 MB	8 files	0 files	0 files
PTAMMCFLLTVTTPTV041	original	MIXED	229.8 MB	8 files	0 files	0 files
PTAMMCFLLTVTTPTV040	original	MIXED	109.4 MB	4 files	0 files	0 files
PTAMMCFLLTVTTPTV039	original	MIXED	348.7 MB	12 files	0 files	0 files
PTAMMCFLLTVTTPTV038	original	MIXED	160.2 MB	6 files	0 files	0 files


PILOT SITE
E-ARK
Powered by RODA
E-ARK is a multinational big data research project that aims to improve the methods and technologies of digital archiving, in order to achieve consistency on a Europe-wide scale
Welcome
Catalogue
Search
Ingest
Administration
Planning
Help
mferreira
English
Search files...
PT-AMM-CFLLTV-TT-PTV-048_m0016.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0015.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0014.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0013.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0012.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0011.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0010.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0009.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0008.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0007.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0006.TIF
PT-AMM-CFLLTV-TT-PTV-048_m0005.TIF
Há-se servir para o lançamento do Novo Imposto desta Vila e seu Termo do presente ano. Vai numerado e rubricado por mim. Mafra, 19 de Janeiro de 1818. José Roque Soares de Albergaria Monteiro. / Original (MIXED) / PT-AMM-CFLLTV-TT-PTV-048_m0016.TIF
File preview not supported
DOWNLOAD
Details
Filename
PT-AMM-CFLLTV-TT-PTV-048_m0016.TIF
Size
26.3 MB
Mimetype
image/tiff
Format
Tagged Image File Format
PRONOM
fmt/353
Fixity
62AD4912F45BF7F4A375EC857D2985C29C1DA8A275EEE470019438AD4CAF0F21 (SHA-256, RODA)
C84F9D6E0BFF06A7B336CE4FAB633475 (MD5, RODA)
Storage path
aip/ce4c1ed5-9f6e-4566-96bf-556290d3d286/representations/PTAMMCFLLTVTTPTV048/data/PT-AMM-CFLLTV-TT-PTV-048_m0016.TIF
E-ARK
ICTPSP

Results

- 100% of the records were successfully transferred and ingested
- Extraction, SIP creation and transfer took 1,5 weeks (including retries)
- Ingest took 27,5 hours
- Completeness was checked by comparing Inventory Reports produced at both ends of the pipeline

Results

- Users can now search records on preservation service
  - Descriptive metadata
  - Technical metadata
- Download representations
- Perform preservation actions
  - With PREMIS events
- Manage risks
- Create dissemination derivatives
  - for special designated communities

Pilot 7

Access to Databases

József Mizsei, Zoltán Lux
National Archives of Hungary

Pilot 7	Access to databases
Task leader	National Archives of Hungary
Scope	Pre-ingest and ingest of databases and unstructured document files and access them via 3 defferent access methods

E-ARK Specifications	SIP		AIP		DIP		SMURF (ERMS)		SMURF (SFSB)		GeoData		SIARD 2.0		
		X		X		X								X	

E-ARK Tools	Pre-Ingest							Ingest				Access																	
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR index (E-ARK web)	Search and Dsiplay GUI		Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	Database Visualisation Toolkit	IP Viewer				QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
	X		X					X			X					X		X	X	X	X							X	

Pilot Scenarios	
Scenario 1	SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format
Scenario 2	SIP Creation and Ingest of unstructured files
Scenario 3	Extract SIARD Package from Preservica/E-ARK AIP (APEX/Oracle BI access)
Scenario 4	Search and present SIARD based information with E-ARK access tools (HADOOP based search and access with HIVE Presentation in local environment)
Scenario 5	Access information from unstructured files

Dataset 1	Hungarian Prosecution Office database
Description	Old (not normalized) database in CSV exports of DBASE files.
Data type	CSV files
Metadata format	none
Quantity	more then 300.000 cases and 500.000 name. (1,6 GB)
Dataset 2	**Scanned meeting minutes of the Central Committee of the Hungarian Socialist Party**
Description	Scanned documents in file systems in PDF file and corresponding metadata (EAD)
Data type	PDF/JPG files (representations)
Metadata format	EAD
Quantity	123.225 files. (101 GB)

5 Scenarios

2 ingest

3 access

Scenario 1

BÜRKE

RODA-in

DBPTK

Scenario 1 - BÜRKE

- Data from Hungarian Prosecution Office - from DBASE database
- Load data to Oracle server
- Create SIARD files
- SIARD to SIP
- SIP to Preservica & E-ARK WEB

Scenario 2

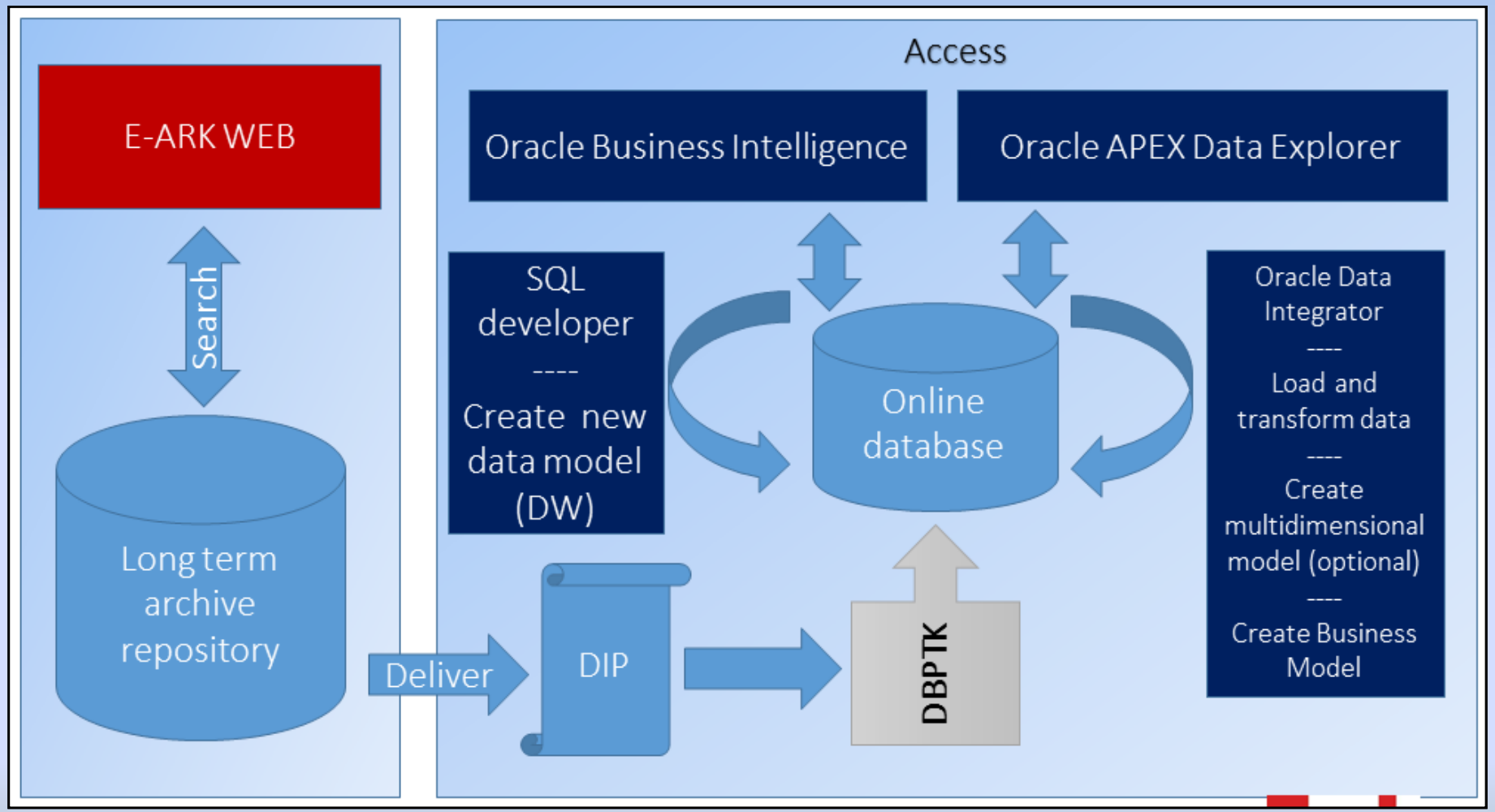
PDF to E-ARK SIP by RODA-in

PDF to NAH SIPs by Elev SIP készítő or XIP Creator by Poliphon

Scenario 2

- SIP from PDF files with E-ARK SIP creator (RODA-in)
  - Load into E-ARK WEB
- Create SIPs from PDF files with our SIP or XIP creator
  - Load into our SDB

Scenario 3-4


E-ARK WEB
Search
Long term archive repository
Deliver
DIP
Access
Oracle Business Intelligence
Oracle APEX Data Explorer
SQL developer

Create new data model (DW)
Online database
DBPTK
Oracle Data Integrator

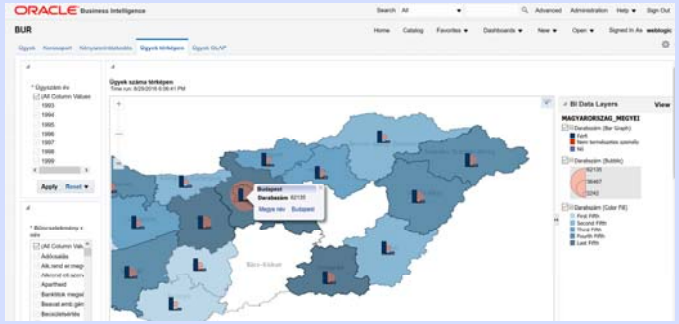
Load and transform data

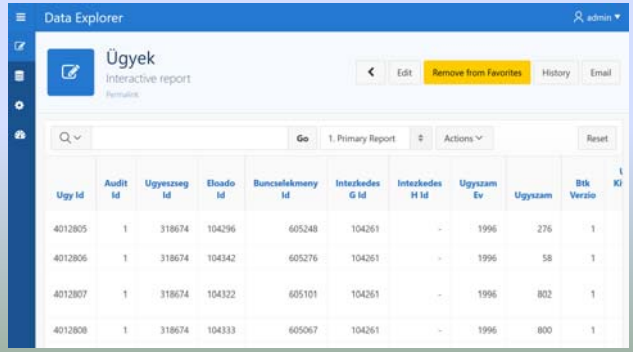
Create multidimensional model (optional)

Create Business Model

Scenario 3 - access

- Access database infromation of the Hungarian Prosecution Office in SIARD format using Oracle APEX and Oracle Business Inteligence.
  - Relational database archived as SIARD package in SDB (SIARD 1.0) and in E-ARK Webe (SIARD 2.0)
  - Load SIARD package into a live database by DPTK
  - Transform data (ODI, OWB)
  - Create multidimensional model
  - Create business model
  - Present it by OBI
  - Present it by Oracle APEX (Data Exlplorer Interacitive Report)





Scenario 4 - access

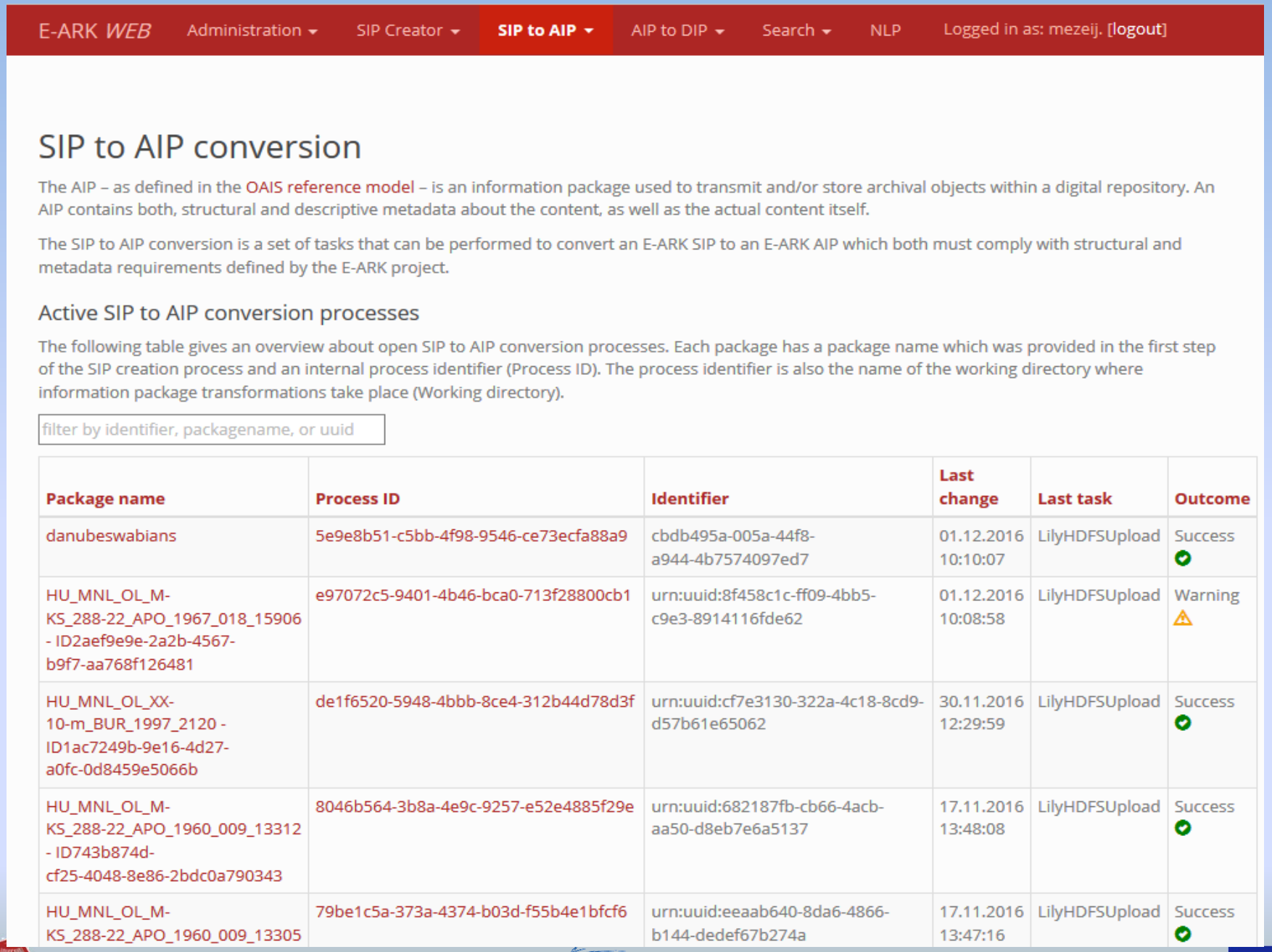
- Access database information of the Hungarian Prosecution Office in SIARD format using HADOOP based search and present it by E-ARK access tools.
  - IP Viewer
  - Data Visualization Toolkit

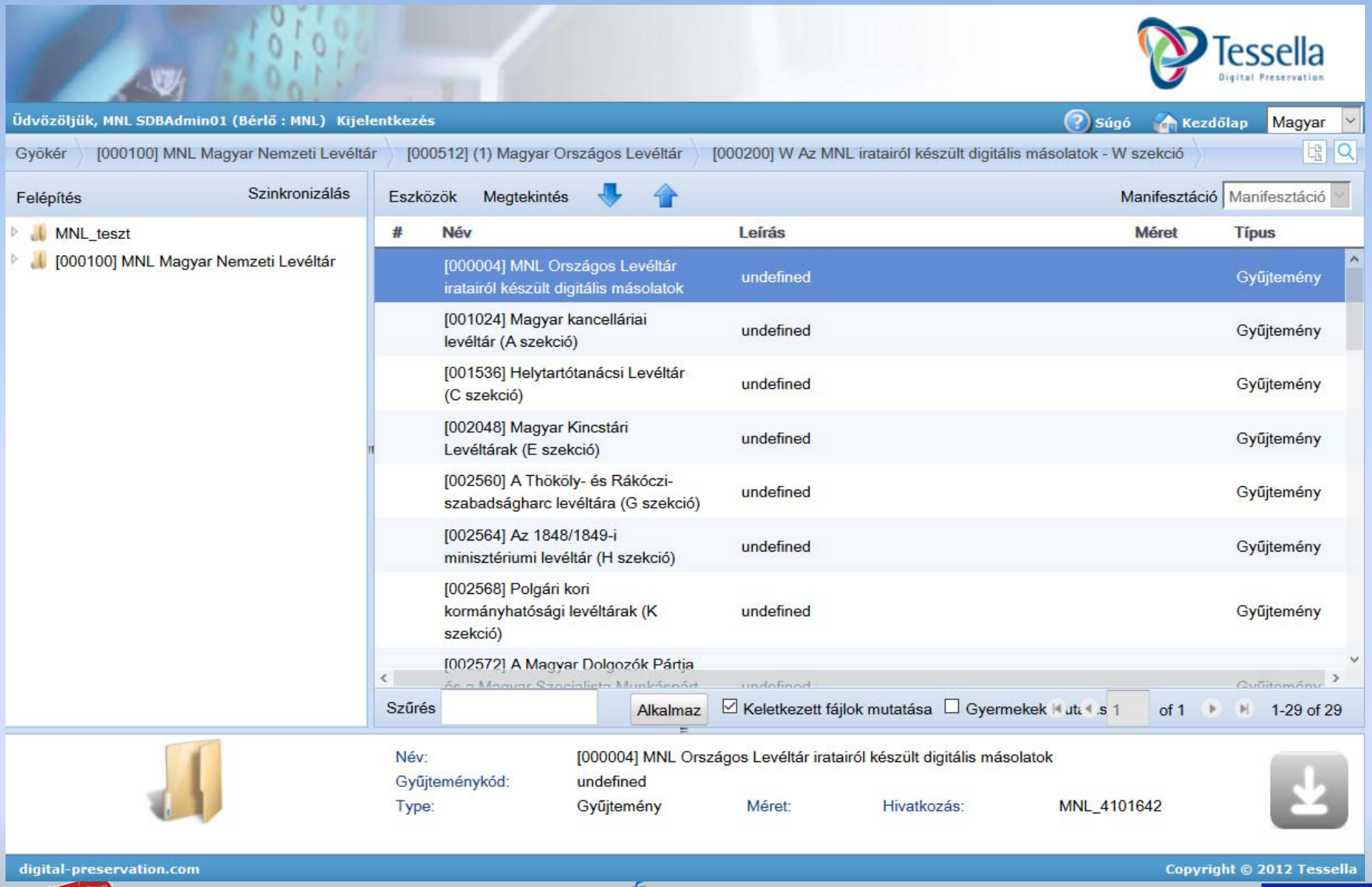
Scenario 5

Access information from unstructured files - access

Scenario 5 - access

- DIPs can be created
- AIPs can be found by OCR text


E-ARK WEB
Administration
SIP Creator
SIP to AIP
AIP to DIP
Search
NLP
Logged in as: mezeij. [logout]
SIP to AIP conversion
The AIP – as defined in the OAIS reference model – is an information package used to transmit and/or store archival objects within a digital repository. An AIP contains both, structural and descriptive metadata about the content, as well as the actual content itself.
The SIP to AIP conversion is a set of tasks that can be performed to convert an E-ARK SIP to an E-ARK AIP which both must comply with structural and metadata requirements defined by the E-ARK project.
Active SIP to AIP conversion processes
The following table gives an overview about open SIP to AIP conversion processes. Each package has a package name which was provided in the first step of the SIP creation process and an internal process identifier (Process ID). The process identifier is also the name of the working directory where information package transformations take place (Working directory).
filter by identifier, packagename, or uuid
Package name
Process ID
Identifier
Last change
Last task
Outcome
danubeswabians
5e9e8b51-c5bb-4f98-9546-ce73ecfa88a9
cbdb495a-005a-44f8-a944-4b7574097ed7
01.12.2016 10:10:07
LilyHDFSUpload
Success
HU_MNL_OL_M-KS_288-22_APO_1967_018_15906 - ID2aef9e9e-2a2b-4567-b9f7-aa768f126481
e97072c5-9401-4b46-bca0-713f28800cb1
urn:uuid:8f458c1c-ff09-4bb5-c9e3-8914116fde62
01.12.2016 10:08:58
LilyHDFSUpload
Warning
HU_MNL_OL_XX-10-m_BUR_1997_2120 - ID1ac7249b-9e16-4d27-a0fc-0d8459e5066b
de1f6520-5948-4bbb-8ce4-312b44d78d3f
urn:uuid:cf7e3130-322a-4c18-8cd9-d57b61e65062
30.11.2016 12:29:59
LilyHDFSUpload
Success
HU_MNL_OL_M-KS_288-22_APO_1960_009_13312 - ID743b874d-cf25-4048-8e86-2bdc0a790343
8046b564-3b8a-4e9c-9257-e52e4885f29e
urn:uuid:682187fb-cb66-4acb-aa50-d8eb7e6a5137
17.11.2016 13:48:08
LilyHDFSUpload
Success
HU_MNL_OL_M-KS_288-22_APO_1960_009_13305
79be1c5a-373a-4374-b03d-f55b4e1bfcf6
urn:uuid:eeaab640-8da6-4866-b144-dedef67b274a
17.11.2016 13:47:16
LilyHDFSUpload
Success


Tessella
Digital Preservation
Üdvözöljük, MNL SDBAdmin01 (Bérlő : MNL) Kijelentkezés
Súgó
Kezdőlap
Magyar
Gyökér
[000100] MNL Magyar Nemzeti Levéltár
[000512] (1) Magyar Országos Levéltár
[000200] W Az MNL iratairól készült digitális másolatok - W szekció
Felépítés
Szinkronizálás
MNL_teszt
[000100] MNL Magyar Nemzeti Levéltár
Eszközök
Megtekintés
Manifesztáció
Manifesztáció
Név Leírás Méret Típus
[000004] MNL Országos Levéltár iratairól készült digitális másolatok undefined Gyűjtemény
[001024] Magyar kancelláriai levéltár (A szekció) undefined Gyűjtemény
[001536] Helytartótanácsi Levéltár (C szekció) undefined Gyűjtemény
[002048] Magyar Kincstári Levéltárak (E szekció) undefined Gyűjtemény
[002560] A Thököly- és Rákóczi-szabadságharc levéltára (G szekció) undefined Gyűjtemény
[002564] Az 1848/1849-i minisztériumi levéltár (H szekció) undefined Gyűjtemény
[002568] Polgári kori kormányhatósági levéltárak (K szekció) undefined Gyűjtemény
[002572] A Magyar Dolgozók Pártja
Szűrés
Alkalmaz
Keletkezett fájlok mutatása
Gyermekek
of 1
1-29 of 29
Név: [000004] MNL Országos Levéltár iratairól készült digitális másolatok
Gyűjteménykód: undefined
Type: Gyűjtemény
Méret:
Hivatkozás: MNL_4101642
digital-preservation.com
Copyright © 2012 Tessella

E-ARK

ICTPSP

QUESTIONS?