



D.2.5 Recommended Practices and Final Public Report on Pilots

DOI: 10.5281/zenodo.1172058

Grant Agreement Number:	620998
Project Title:	European Archival Records and Knowledge Preservation
Release Date:	12 th February 2018
Conti	ributors
Name	Affiliation
lstván Alföldi	National Archives of Hungary
István Réthy	National Archives of Hungary
Andrew Wilson	University of Brighton
Clive Billenness	University of Brighton
Anders Bo Nielsen	Danish National Archive
Phillip Mike Tømmerholt	Danish National Archive
Alex Thirifays	Danish National Archive
Hans Fredrik Berg	National Archives of Norway
Terje Pettersen-Dahl	National Archives of Norway
Arne-Kristian Groven	National Archives of Norway
Tarvo Kärberg	National Archives of Estonia
Karin Oolu	National Archives of Estonia
Raivo Ruusalepp	Estonian Business Archive
Ats Rand	Estonian Business Archive
Gregor Završnik	National Archives of Slovenia
Boris Domajnko	National Archives of Slovenia
Joze Skofljanec	National Archives of Slovenia
Miguel Ferreira	Keep Solutions
Zoltán Lux	National Archives of Hungary
Mezei József	National Archives of Hungary
David Anderson	University of Brighton
Janet Anderson	University of Brighton

Table of Contents	
EXECUTIVE SUMMARY	
PLANNING AND EXECUTING THE E-ARK PILOTS	
PILOT PLANNING IN THE DESCRIPTION OF WORK (DOW)	4
PILOT PLANNING DURING THE PROJECT	4
PILOT PREPARATION	5
PILOT EXECUTION	11
PILOT EVALUATION	
OVERVIEW OF THE E-ARK PILOTS	
Full-scale pilots and OAIS process	20
Full-scale pilots and E-ARK uses-cases	21
Pilots using E-ARK tools and format specifications	22
PILOT REPORT	
PILOTS 1 - SIP CREATION ON RELATIONAL DATABASES	23
Scenarios	25
Execution report	28
Changes to the original plans	
Feedback report	
Recommended practices and further recommendations	
PILOTS 2 - SIP CREATION AND INGEST OF RECORDS	
Scenarios	
Execution report	
Changes to the original plans	
Feedback report	
Recommended practices and further recommendations	
PILOTS 3 - SIP CREATION AND INGEST OF RECORDS	40
Scenarios	42
Execution report	45
Changes to the original plans	47
Feedback report	47

Recommended practices and further recommendations	48
Pilots 4 - Business Archives	49
Scenarios	51
Execution report	53
Changes to the original plans	53
Feedback report	54
Recommended practices and further recommendations	54
PILOTS 5 - PRESERVATION AND ACCESS TO RECORDS WITH GEODATA	55
Scenarios	57
Execution report	60
Changes to the original plans	61
Feedback report	61
Recommended practices and further recommendations	62
PILOTS 6 - INTEGRATION BETWEEN A LIVE DOCUMENT MANAGEMENT SYSTEM AND DIGITAL ARCHIVING AND PRESERVATION	N SERVICE64
Scenarios	66
Execution report	68
Changes to the original plans	69
Feedback report	70
Recommended practices and further recommendations	70
Pilots 7 – Access to Databases	72
Scenarios	73
Execution report	78
Changes to the original plans	78
Feedback report	79
Recommended practices and further recommendations	80
EXTERNAL EVALUATIONS	82
PILOT EVALUATION	
PROJECT LEVEL PILOT SUCCESS EVALUATION	84
PILOT AND SCENARIO LEVEL SUCCESS EVALUATION	88
REFERENCED DOCUMENTS	

APPENDIX 1 – EXTRACT FROM E-ARK DOW

Executive Summary

E-ARK project

The goal of the European Archival Records and Knowledge Preservation (E-ARK) Project is to pilot archival services to keep records authentic and usable based on current best-practices. These will address the three main endeavours of an archive – acquiring, preserving and enabling re-use of information. E-ARK will demonstrate the potential benefits for public administrations, public agencies, public services, citizens and business by providing easy and efficient access to the archived records.

The project brings together a core group of European national archives, four leading research institutions, three providers of archiving software solutions and services, two government agencies, and two international membership organisations that represent the communities who stand to benefit from the project: data owners/providers, archives, software vendors and solution providers.

E-ARK will, over a three year period, harmonise archival processes at a pan-European level supported by guidelines and recommended practices that will cater for a range of data from different types of source including record management systems and databases.

Work Package 2 (description from DoW)

The E-ARK General Model definition is a public deliverable of Work Package 2.

The overall objective of this work package is to ensure that the scenarios implemented at 7 identified pilot sites are both realistic and relevant, that they bring together a meaningful subset at each site of the use cases in order to establish a general model of the E-ARK service.

WP2 will

- Identify specific use cases that will each be implemented in at least one pilot scenario, covering:
 - Export from business systems
 - o Creation of SIPs from unstructured and structured data
 - Execution of the complete SIP -> AIP -> DIP data-flow to support migration and submission/access scenarios
 - Existing use cases for access to content in physical and virtual reading rooms (with appropriate access controls) and as web-applications
 - Additional use cases that augment the main pilot programme including short "stretch tests" and 3rd party validation
- Identify and mitigate legal and regulatory constraints.
- Provide support and advice about the operational environment of the pilot sites to the teams in WP3-6 during the planning phase (which corresponds to their main cycles of iterative (agile) design and development.

- Support the teams working at the pilot site in the planning and deployment phase
- Ensure smooth execution of the pilots.
- Document the recommended practices and lessons learned in the project knowledge base.

T2.4 Future pilot deployment (M25-M27)

The objective of this task is to finalize the pilots in harmony with D2.1.

The Electronic Archiving Service consists of a series of activities covered by software tools and manual workflow steps. These tools are currently partly in existence, some are being developed by E-ARK project, many more are to be added by developments of the digital preservation community in the future. The role of this task is to identify the most relevant scenarios for the E-ARK Service, define for each scenario which level of activity is needed in order to bridge the gaps of the currently existing solutions (e.g. integration, software development, interface definition).

In order for the E-ARK service to demonstrate the functionality of the service built on D2.1 as fully as possible, the pilot will be finalized around the 7 pilot sites. In order to plan ahead for the pilots, the project previously identified three activity levels:

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. Members of DLM forum and DPC will receive details of the pilot implementation and be invited to participate as observers. There are seven full scale pilots.

2. Additional project pilot activities – implementation, by consortium members of shorter 'stretch' pilots that extend the scenarios or apply them in different contexts. This may include the participation of members of DLM Forum and DPC who are not directly members of the E-ARK consortium

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. Outcome of this task is the high-level requirement specification of the full scale pilots and also scenarios, sites and requirements of the 2nd and 3rd level pilots.

T2.5 Support and execution of pilots. (M7-M33)

The task is concerned with the implementation of the pilots defined in D2.3. The Task Leader contributes to providing an appropriate methodological framework for all pilots for specifying the input/output points and the uniform principles applied in the different areas, such as source data management, user training, user documentation, interim reports and the final reports. In this way the results of the pilot sites are comparable and can be reliably proven in this E-ARK-service pilot. There are seven 6-month pilot sites running concurrently and these are defined in Section B3.2a, Approach. This document corresponds to the deliverable:

D2.5 Recommended practices and Final public report on Pilots

Arising from the experiences acquired during the 7 pilot deployments, this report describes the achievements and results of the pilot activities over the entire three-year period with emphasis on the final year of the project. The report lists the resources used and provides an evaluation of progress and final result against the project objectives and milestones and documents the remaining problems. It summarises the recommendations and lessons learned from each pilot and provides input for the overall final report of the project. This report will also be included in the final, publishable project report [month 36]

Structure of this deliverable

This document summarizes pilot activities, achievements and best practice recommendations using the following chapter structure:

- Chapter 1 This introductory chapter.
- Chapter 2 Planning and executing the E-ARK pilots Summary of all pilot related activities in the 3 years of the pilot, from planning to evaluation.
- Chapter 3 Pilot overview A brief overview of the full-scale and additional pilots.
- Chapter 4 Pilot report

Summary of the pilot execution and results with recommended practices and further development recommendations. The chapter consists of the following sections for each full-scale pilot:

- Pilot scenario details
- Execution report
- Changes to previous plans
- Feedback report, and
- Recommended practices and lessons learnt.

Chapter 4 ends with an overview of the external evaluations performed by non-EARK member organizations.

- Chapter 5 Pilot evaluation Evaluation of the full-scale pilot against project objectives and success criteria.
- Chapter 6 Referenced documents and web pages
- Appendix 1 Extract from E-ARK Description of Work

Planning and executing the E-ARK pilots

This chapter summarizes all the pilot related activities of the E-ARK project. The seven full-scale pilots were already quite well planned in the Description of Work (DoW) document when we started the real project work at the beginning of 2014. From that point until the very end, Work Package 2 (WP2) was focusing on pilot planning and, later on, on execution and evaluation.

Phases of pilot related activities coordinated by WP2:

- Pilot planning in the Description of Work (DoW) document The starting point of our work was the pilot descriptions in the DoW.
- Pilot planning during the project

In the first year our main goal was to define the use-cases and processes to serve as the basis of tool development and format specification. The first version of the E-ARK General Model defined the use-cases and processes along with cross-reference tables between E-ARK processes, tools, work packages, and pilots. After the publishing of E-ARK General Model, colleagues at the pilot sites were developing part of the requirement specification of the E-ARK tools.

- Pilot preparation
- Pilot execution
- Pilot evaluation

This chapter is organized according to the above phases.

Pilot planning in the Description of Work (DoW)

The starting point of our work was the pilot descriptions in the DoW. The Description of Work (DoW) document defines the pilot related tasks and the role of Work Package 2. Appendix 1 is an extract of the relevant part of the E-ARK DoW.

Pilot planning during the project

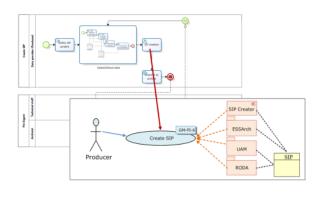
Pilots were planned to take place in the third year of the project when all tools and format specifications were ready to be tested, but pilot related activities started at very beginning and accompanied the tool development and format specification work throughout the project.

General Model 1.0

One of the first deliverables was the D2.1 E-ARK General Model of Use-cases and Processes. In the General Model we defined the use-cases and processes which were the basis for further project activities like planning and development of the E-ARK tools, and specification of E-ARK information package and content types.

The General Model was a joint work by the tool developers of the partner IT companies, and archivists from the pilot sites. Along with the use-case definition we tried to reach a common understanding of the project. At that point – at the very beginning of the work – every partner had some ideas about their own goals and tasks but hardly anyone could see what the other partners would provide to the project. We found that some overall birds-eye approach would help people better see their place among the various activities planned so we have included some cross-

reference tables in the General Model as well. The cross-reference tables present relations between the different project activities and products like work packages, tools, formats, and pilots.



Use Ca	se View												
E-ARK	General Model												
Use Case					Pilot		_		w	ork F	Packa	ge	Tools
		Pilot 1 (DNA)	Pilot 2 (NAN)	Pilot 3 (NAE)	Pilot 4 (EBA)	Pilot 5 (NAS)	Pilot 6 (KEEP)	Pilot 7 (NAH)	EdM	WP4	WPS	WP6	
Pre-Inges	st												
GM-PI-1	Define SIP content	x	х	х	х	х	х	х					
GM-PI-2	Select data (with rules)	×		х	х		х	х	х				DBExport tool
GM-PI-3	Select data (manual)		х	х	х	х			х				DBExport tool
GM-PI-4	Extract data from DB	×				х		х	х				ESSArch Tools
GM-PI-5	Extract data from DMS/RMS		х	x	х		х		х				ESSArch Tools, Noark, Alfresco, RODA
GM-PI-6	Create SIP	×	x	x	x	x	x	x	x				DBExport tool, ESSArch Tools, SIP creation tools, RODA-in, UAM
GM-PI-7	Start transfer to archive	×	х	х	х	х	х	х					
GM-PI-8	SIP reception	×	х	х	х	х	х	х					
GM-PI-9	Validate SIP	×	х	х	х	х	х	х	х				SIP to AIP conversion tools
GM-PI-10	Manipulate SIP	?	?	?	?	?	?	?	х				
GM-PI-11	Create fond(s)	x	х	х	х	х	х	x					SDB, EPP, RODA, AIS
GM-PI-12	Start generating E-ARK SIP	x	x	x	х	х	x	x	х				SIP to AIP conversion tools
Ingest													
GM-I-1	Upload SIP	x	х	х	х		х	х					Preservica, EPP, RODA, AIS
GM-I-2	Start AIP generation workflow	×	x	x	х		х	х	х	х			SIP to AIP conversion tools, Preservica, EPP, RODA, AI
GM-I-3	Validate AIP	x	х	х	х		х	х		x			SIP to AIP conversion tools, Preservica, EPP, RODA, AI
GM-I-4	Start AIP finalization workflow	×	x	x	x		x	x		x			SIP to AIP conversion tools, Preservica, EPP, RODA, AI

The General Model helped us better understand every partner's planned contribution to the overall objectives and gave us a better picture of the whole project. As a result of this common approach the pilot representatives at the meetings tried to think ahead about what they really need and wanted to try out later in the third year, and tried to gently lead tool developers towards solutions which better suited their demands.

Requirement specification

After completing the General Model the Pilot site members took part in the next project phase, the requirement specification work. On the basis of the General Model (and the discussions about it) they could articulate their requirements better at the technical work package (WP3-6) requirement specification meetings. The results of this work were the requirement specifications of the pre-ingest, ingest and access tools, along with E-ARK information package (SIP, AIP, DIP) and content type (SIARD 2.0, SMURF) specifications.

Tool development and format specification

Cooperation between archivist of the pilot sites and tool/specification developers continued during the development and specification phase, keeping the pilots in mind.

Changes to the planned pilot activities

At this phase there were no major differences identified compared to the plans written in the Description of Work.

Pilot preparation

Actual pilot preparation work started in the second year. WP2 and the pilot sites wanted to make sure that the tools being developed and format specifications being defined were in line with their planned piloting activities. Therefore we started to define the pilots very early.

Early pilot preparation works

At the 2015 Portsmouth and Lisbon meetings we held pilot preparation sessions. We agreed on the organization of preparation activities and a schedule. In the summer of 2015 the structure of the pilot definition document was also approved by project members.

Pilot Cards

In order to promote early understanding of the pilot activities and requirements and to provide a quick overview at a central point of information we developed the Pilot Cards. Pilot Cards were the first formalized appearance of the pilot activities in the E-ARK community.

The Pilot Cards provide an overview of the pilot including scope and objective, contact info of the pilot leader and contributors, OAIS relevance, usages of E-ARK tool and information package as well as status information about the definition, installation and execution. Pilot Cards can also serve as a central information point for the EARK community to find detailed pilot information descriptions and corresponding documents.

Access

CMIS portal/viewe

Pilot #1 SIP Creation on relational databases Status OAIS relevance ٧ Pilot defined Management Preservation E-ARK AIP ٧ Installation started **Pre-Ingest** E-ARK SIP E-ARK DIP Ingest • Installation ready -Data Pilot execution started Pilot execution completed х Task leader **Danish National Archives** Supported by Magenta Name (Title) Contacts e-mail Skype **Contact Person** philliptommerholt_rigsarkivet Phillip Mike Tømmerholt pmt@sa.dk **Contact Person** Anders Bo Nielsen abn@sa.dk **Pilot staff members** The scope of this Pilot is to test the E-ARK SIP Creation tool with not less than 4 databases of different sizes and Scope complexities (one contains several million records) Object Creating SIPs for relational databases using the tool created in WP3, T3.3: SIP Creation Tools, for further evaluation The goal of the pilot is to create SIPs in EARK-SIP format of each selected database with the DBextract tool. After Short description quality assurance on each SIP, a feedback will be given to WP3 M28-M33 Timeframe M03.3, M03.4 (DoW) Preconditions E-ARK tools T00 SSArch Preservation Platform SSArch Preservation Platform Database Preservation Toolkit **DMT - Search and Dsiplay GUI** ESSArch Tools Archive (ETA) ESSArch Tool Producer (ETP) **DMT - Order Management Order Submission Service** SIP creator (E-ARK Web) RMS Viewer (Alfresco) Catalogue (E-ARK web) Alfrsco Export Module (IP2AIP (E-ARK Web) IP2DIP (E-ARK Web) **Dracle (OLAP Viwer)** ARK Web (Search) **tODA Repository DB Viewer (Sofia)** Single file Viewr DFS-Storage ily - Ingest IP Viewer Geoserver Peripleo RODA-In DBPTK ğ Ā х 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) Links Process and use case information Pilot definition

Fest data specification

Pilot Card example

Pilot documentation

Pilots Definition

At the fall of 2015 we had the first draft of the document D2.3 Detailed pilot requirements. The most important part of this document was the "Pilot Definition".

Pilot definitions came in the form of Excel files and defined the pilot scenarios in detail. The sheets of the excel file are:

- Overview
- Scenario description
- Data description
- Pilot preparation checklist
- Step-by-step process description sheets for Pre-Ingest, Ingest and Access processes

The logical structure of the Pilot Definition description:

Pilot

Scenario

- Business use-case (from General Model)
- Used Information package types
- ➢ Used E-ARK tools
- Data Set description
 - > Content description
 - Metadata description
- Pilot preparation description and status information
- Process description
 - Process step and low-level use-case (from General Model)
 - Used E-ARK and local tools
 - Preliminaries and start condition
 - Input/output description
 - > E-ARK (and local) tools usage details

The scenarios, data and tool usage along with pilot preparation and step-by-step process activities are defined in detail in the Pilot Definition excel documents. The final version of the Pilot Definition excel file of each pilot is part of the deliverable D2.4 Pilot Documentation.

Detailed Pilot Requirements

Beside the pilot definition excel files, the D2.3 Detailed Pilot Requirements document defined the following requirement types:

- Schedule
- Success criteria
- Support requirements Requirements for tool developers in regard to supporting pilot preparation and execution activities
- Feedback requirements
 Requirements for pilot staff members about how to provide feedback on tools and format specifications
- Documentation requirements

Here are some example pages of the pilot definition from the deliverable D2.3 Detailed Pilot Requirements:

Pilot 5	Pre	ese	rvat	tior	n ar	nd a	icce	ess	to r	eco	rds	wi	thg	geo	dat	а												
Task leader	Na	National Archives of Slovenia																										
Supported by	Da	Danish National Archives																										
Scope	data arch	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 (via open data method, direct access in the archives and use GIS data as search criteria in the DIP contents).																										
Object	WP	Pilot report with recommendations about urgent improvements and possible future improvements support for WP6 & WP7 setting up the work environment of selected E-ARK archival tools provide real life examples how the project deliverables can be used																										
Short description	arch	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.																										
Timeframe	M25	M25-M27: setting up the pilot sites; M28-M31: running the pilots; M32-M33: testing and reporting																										
Preconditions	MO	M03.3, M03.4, M04.2, M05.4, M05.6 (DoW)																										
Contacts	Nan	ne (1	litle)							E-m	E-mail Skype																
Contact Person	Gre	gor Z	Zavr	šnik	()						gregor.zavrsnik@gov.si gregor.zavrsnik																	
Pilot staff member	Ale	nka S	Starr	man	()						alenka.starman@gov.si																	
Pilot staff member	Joze	e Sko	oflja	nec	()						joze.skofljanec@gov.si																	
OAIS Relevance			Pre	-Ing	est -ARK	SID	x	-	est - E-ARK		orage Storage - Access								סוח	x								
E-ARK Tools	Database Preservation Toolkit	Alfrsco 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 Plat form	HDFS-Storage	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Plat form	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Ge oser ver	Peripleo	Oracle (OLAP Viwer)
				х	х					х		х	х	х	х	х	х	х			х				х	х	х	

Г

Scenario 2	Sea	irch	and	l Ac	cess	inf	orm	atio	on u	sing	Gea	dot	ta															
Description	Cre con A d req to c des (QC	reate DIP from AIP containing record with Geodata. Present Geodata information with QGIS along with ontent and metadata from DIP. data object containing geodata can be identified by using search criteria as specified by E-ARK Tool equirement specification. Selected data objects are selected and order is issued. DIP is prepared according order specification and end user credentials. DIP file structure with file descriptions (mime type, short escription) is presented to the enduser. Geodata from the order can be accessed in the designated viewer QGIS). The user checks authenticity of the DIP by accessing PREMIS documentation. Access to DIP is ocumented and captured metadata can be exported.																										
OAIS Relevance		anne		-Ing		aptu	ircu			Stor			pon	.cu.				St	orag	ge - I	Acce	ss						
		E-ARK SIP E-ARK AIP X E-ARK DIP								x																		
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	NAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	× ICA-AtoM Catalogue	× OMT - Search and Dsiplay GUI	× Order Submission Service	× OMT - Order Management Tool	× Lily - Ingest	× ESSArch Preservation Platform	× E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	× IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	× QGIS	× Geoserver	× Peripleo	Oracle (OLAP Viwer)
Use-case	Acc	ess	geo	data	a via	a QO	SIS	-	•		-		•		•		•				•			•	•			
Note	Acc	Access records with Geodata and present geodata with QGIS																										

Pilot 5	Pilot Data	İ							
Information Packages (IP)	IP		Note						
	E-ARK SIP	х	Focusing on Geodata preservation						
	non E-ARK SIP								
	E-ARK AIP	х	Focusing on Geodata preservation						
	non E-ARK AIP								
	E-ARK DIP	х	Focusing on Geodata access						
	non E-ARK DIP								
Pilot data description									
Data Set 1	Records and metadata of ac (The Surveying and Mappin		nits until 1994 exported from GURS the Republic of Slovenia)						
Description	Records and metadata of m	aps with Geod	ata						
Data type	GML document with metad	ata in XML forr	nat, ESRI Shapefile, csv						
Metadata format	ISO 19115 (INSPIRE)								
less	62 records (cca. 3MB)								
Data Set 2	Records and metadata of Na	Records and metadata of Natura 2000 areas, exported from ARSO							
Description	Records and metadata of m	Records and metadata of maps with Geodata							
Data type	GML document with metad	GML document with metadata in XML format							
Metadata format	ISO 19115 (or INSPIRE)								
less	1209 records (cca. 10 MB)	.209 records (cca. 10 MB)							

D2.5 Recommended Practices and Final Public Report on Pilots

OAIS Process									
					Pre-Ingest				
Main Process Stepps	Content definition Technical feasibility Legal issues etc.	Create/Review transfer agreement	Select data	Data Extraction	Manual compilation of non ERMS content	Metadata mapping	Create SIP	Post-packaging quality control	Submit SIP
Scenario 1	SIP Creation and Ingest of	f geodata in GML format							
Used E-ARK tool		ľ		QGIS			ESS Arch ETP		ESS Arch ETP
					Producer tools, open	MS Excel, Inspire Metadata			
Used local tools			Existing archival system	Producers tools	convesion tools	Creator			
		Producer + Archivist +	. .						
Perfomer (actor)		Technical Specialist	Producer	Producer	Producer	Producer	Producer		Producer
		Official archival records							
Prelemineries and Start condition		definition							
lanut		Official archival records definition and technical documentation	Cubmission Agroement	Submission Arrowmant	Eulissian Arronment	INSPIRE.xml, Submission Agreement, MS Excel template for EAD conversion	Extracted data Additional Data and documentation Inspire.xml, MS excel w. metadata		Cubicsion Associate CID
Input	1	documentation	Submission Agreement	Submission Agreement	Subission Agreement Additional Data and	Inspire.xml, MS excel w.	metadata		Subission Agreement, SIP
Output		Submission Agreement	Data selection list	Extracted data	documentation	metadata	E-ARK SIP		Submited SIP
Output		Submission Agreement	Data selection list	Extracted data	documentation	metadata	E-ARK SIP		Submitted SIP
	SIP Creation and Ingest of	f geodata in GML format							
Used E-ARK tool							ESS Arch ETP		ESS Arch ETP
Used local tools			Existing system	Producers tools	Producer tools	Producer GIS system, MS Excel			
		Producer + Archivist +							
Perfomer (actor)		Technical Specialist	Producer	Producer	Producer	Producer	Producer		Producer
		Official archival records							
Prelemineries and Start condition		definition							
		Official archival records definition and technical				INSPIRE.xml, Submission Agreement, MS Excel template for EAD	Extracted data Additional Data and documentation Inspire.xml, MS excel w.		
Input		documentation	Submission Agreement	Submission Agreement	Subission Agreement	conversion	metadata		Subission Agreement, SIP
					Additional Data and	Inspire.xml, MS excel w.			
Output		Submission Agreement	Data selection list	Extracted data	documentation	metadata	E-ARK SIP		Submited SIP

Pilot 5	Pilot Preparation								
						Preparat	ion status		
Software component	Tool / Version number	Scenario	Process	Tool selected	Tool available for Pilot	Tool/Version installation	Tool configuration	Knowledge overtaken	Tool ready for Pilot
Preparation tasks related to the software	from Software Component Matrix				Yes /		No needed / Configured /		
components	(for EARK tools)	from Scenarios sheet	from Processes sheets	Yes /No / (issue)	(planned date of availability)	Installed / (issues)	(issues)		Ready/(issues)
Component 1.	ESSArchive ETP	Scenario 1, 3	Pre-ingest	Yes	Yes	Not installed	Need support form ESS	Basic training completed	No, local installation needed
Component 2.	ESSArchive ETA	Scenario 1, 3	Ingest	Yes	Yes	Not installed	Need support form ESS	Basic training completed	No, local installation needed
Component 3.	ESSArchive EPP	Scenario 1, 3	Ingest (Access?)	Yes	Yes	Not installed	Need support form ESS	Training in progres	EAD Support, Some validation feat
Component 4.	Integrated Platform (EARK WEB)	Scenario 1, 2, 3, 4	Ingest, Access	Yes	No	Not installed	Need support form AIT	Training required	???
Component 5.	QGIS	Scenario 1, 2, 3, 4	Pre-Ingest, Ingest, Access	Yes	Yes	Installed	None needed	Yes	Yes
Component 6.	Inspire metadata editor	Scenario 1	Pre-ingest	Yes	Yes	Online	None needed	Yes	Yes
Component 7.	EAD metadata editor	Scenario 1, 3	Ingest	No	No	Not installed	Need support form ESS	Further knowladge transfer	???
Component 8.	Search and display GUI	Scenario 2, 4	Access	No	No	Not installed	Need support form AIT	Further knowladge transfer	NO
Component 9.	Peripleo	Scenario 1, 2, 3, 4	Ingest, Access	Yes	Yes / in 2/2 April	Not installed	Need support from AIT	Further knowladge transfer	Yes
Component 10.	OMT	Scenario 2.4	Access	No	No	Not installed	Need support form Magenta	Further knowladge transfer	NO
Component 11.	Archival Catalogue (EAD based)	Scenario 1, 2, 3, 4	Ingest, Access	No	No	Not installed		Further knowladge transfer	
Component 12.	Lilly	Scenario 1, 2, 3, 4	Ingest, Access	Yes	Yes / in 2/2 April	Not installed		Further knowladge transfer	
Component 13.	Geoserver	Scenario 2, 4	Access	Yes	Yes	Installed	None needed	Yes	Yes
				Pren	ration status		1		
Pilot dataset	Dataset #	Scenario	Data selected	Legal issues	Data available	Dataset ready for Pilot	-		
i not uataset	Dataset #	Scenario	Data selected	Legarissues	Data available	Dataset ready for Phot			
Preparation tasks related to pilot data	from Pilot Data sheet	from Scenarios sheet	Yes / (issues)	None / (issue)	Yes / (planned date) / (issue)	Ready/(issue)			
Slovenian Register of spatial units sele	ecte Data set 1	1,	2 Yes	None	Yes	Yes			
Natura 2000 dataset	Data set 2	3,	4 Yes	None	Yes	Yes			
				Prepration statu	s				
Infrastructure	Scenario	Process	Element selected	Issues	Element ready for Pilot				
Preparation tasks related to pilot infrastructure	from Scenarios sheet	from Processes sheets	Yes / (issues)	None / (issue)	Ready / (issue)				
Virtual server - Linux									

For details please examine the complete D.2.3 Detailed Pilot Requirements document here:

http://eark-project.com/resources/project-deliverables/60-23pilotsspec

Weekly pilots meeting

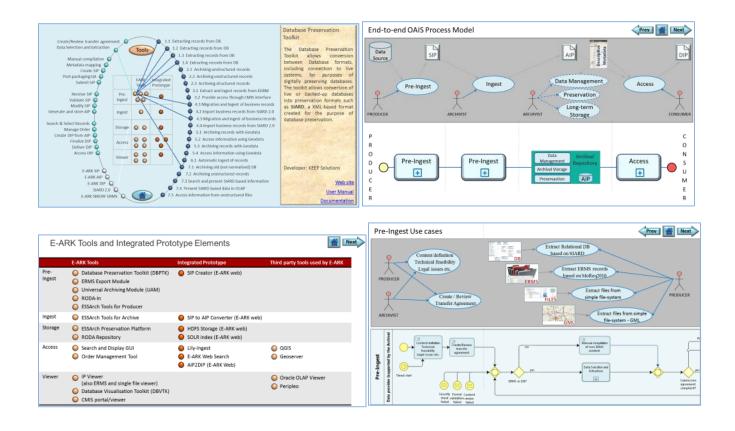
From the beginning of 2016 weekly progress meetings were held via a Webex teleconference service. The pilot representatives and staff members along with technical work package leads and some of the tool developers were regular members of these meetings.

Changes to the planned pilot activities

Only smaller changes were necessary at this phase. Some of the data providers were not ready with the planned input data so the archives needed to arrange different data sets. Some tools were not completed in accordance with the original timetable so we rescheduled some of the scenarios, but fundamentally nothing threatened the successful pilot execution.

General Model 2.0

The creation of the General Model was originally planned to be a one-time activity in order to be the foundation of tool development and format specification. No goals or requirements in the DoW corresponded to any further developmental work. But after seeing how important a role it played in the common understanding of the various goals and approaches of the E-ARK community, we have decided to update the General Model in order to keep the model alive as a reference for the most important E-ARK elements such as tools, formats, use-cases and pilots. The 2.0 version of the model was an online PowerPoint presentation, but we soon discovered that an HTML version would be more suitable both for project members and the wider public. The Power Point version was soon followed by an online presentation in HTML format.



The General Model in its present form is a perfect starting point to get acquainted with the E-ARK project. It includes a complete general reference to present the relationship among tools, use-cases, formats and pilots along with thematic overview chapters with links to more detailed documents and corresponding web pages.

The latest version of the General Model can be found in the E-ARK Knowledge Base and is also accessible from the E-ARK project web site: <u>http://eark-project.com/resources/general-model</u>

Pilot execution

The execution of the full-scale pilots was planned for a 6 month period between month 28 and 33 (from May to October 2016.). All technical and organizational arrangements were in place in April 2016. The full-scale pilots started on 1 May 2016 as planned. Not every scenario was planned to start in May, but every pilot site started with some scenarios in that month.

Software deployment

The software components for the first scenarios were all deployed and configured. Pilot staff members got preliminary knowledge of the tools from the user manuals and on-demand consultations with the developers. The interrelationships among tools were not clear enough so those pilots using many tools (Pilot 5 and 7) tried to create the appropriate tool portfolio to cover all the steps and transitions being tried.

Feedback about tools and format specifications

The pilots were required to give feedback about the deployment, installation, execution and documentation of the E-ARK tools and about format specifications. The developers managed the issues, wishlists and comments on the GitHub sites of the product, while feedback to format specification providers and information on recommended practices was collected respectively in excel files provided by WP2 on the project's Google drive.

Feedback lists

Feedback list	Description	Provided by
For tool developers		
- Bug list	Bugs (issues) found during product execution	Developer on GitHub
- Wish list	Tool extension or modification demands	Developer on GitHub
- Comments list	Comments on tool functioning (anything worth to inform developer about)	Developer on GitHub
- Installation recommendations	Comments or recommendation about the installation process, install kits or installation documentation	WP2 on Google drive
- Feedback on documentation	Comments or suggestions to tool documentation	WP2 on Google drive
- Recommended practices	Experiences with tool execution and recommended practices	WP2 on Google drive
For specification providers	SIP: WP3, AIP: WP4, DIP: WP5	
- General comments and wishes	Issues, comments or wishes related the specific IP	WP2 on Google drive
- Recommended practices	Experiences with IP implementation (structure, mapping, etc.) and recommended practices	WP2 on Google drive

Early progress

As with all large scale projects, at the beginning progress was very slow. We had to accept that only a part (and probably the smaller part) of the archives' work is the actual technical ingest or dissemination of the information. The creation and approval of the formal submission agreements with the data providers took months in some cases. Also some tools (like export modules, and some interfaces) needed adjustments according to the specific data types they were to process. This was a normal procedure which could only be started after the formal agreement with the provider of the data. In some cases (Estonian and Portuguese pilots) this activity required input from a local developer who was not part of the E-ARK project. And we have to confess that the first versions of the new or modified tools had bugs or incompatibility issues with each other and the format specifications. Newly recognized requirements appeared, too, because despite all the discussions and consultations the archivists' knowledge of the tools and the developers' knowledge of the archival work were initially incomplete.

It was originally intended before the execution started that many scenarios would be ready by mid-summer but found that at the end of July there was only one completed scenario.

Weekly pilots meeting

At the weekly pilots meetings every pilot representative reported on progress. We were able to discuss the issues with the tool developers, find solutions to problems, or formulate questions to other project members who were not present. The weekly pilots meeting continued until the end of the project.

Half-time report

At the end of the third month of the pilot WP2 created a (project internal) Half-time Report. The Half-time Report summarized the progress of each scenario with status, and progress overview information and gave a list of the most important issues.

Completing the scenarios

Then things speeded up. The tool developers' response time was very quick. Right after an issue had been recorded at GitHub it was possible to tell when the bug had been corrected or the new requirement could be implemented. Archivists got better understanding of the tools. All legal issues with the submission agreements were solved and at end of August and in September work normalized. Pre-ingest and ingest scenarios were close to reaching their goals and almost all access scenario were able to be started. Only two permanent issues slowed the two scenarios at the Estonian and the Slovenian National Archives. These were due to the late development of the required versions of the ERMS Export Module and the Order Management Tool.

By the end of October – except for the two scenarios – all the full-scale pilots were completed according to the workplans. These two scenarios were also completed later in 2016.

Monthly reports

The pilot progress was tracked in Monthly Pilot Reports produced at the end of each month by the pilot sites. The report summarizes the activities of the last month, any issues and possible solutions, other comments and recommended practices.

The monthly pilot report contained:

- Scenario overview
- Tools overview
- IP feedback overview
- Scenario details per scenario

Scenario Overview

Scenario	Started Completed	Status	Comment
Number and Title of pilot scenario	date date	0 % Not	
		started	
Number and Title of pilot scenario	date	0-100 %	reason for delayed status or any

	date	Delayed	important comments at scenario level
Number and Title of pilot scenario	date	0-100 %	
	date	Started	
Number and Title of pilot scenario	date	100 %	
	date	Completed	
Number and Title of pilot scenario	date	0-100 %	reason for pending status or any
	date	Pending	important comments at scenario level
Number and Title of pilot scenario	date	0-100 %	reason for pending / aborted / delayed
	date	Aborted	status or any important comments at
			process step level

Tools Overview

E-ARK Tool – Version	Issues (bugs, wishes, comments) Experiences / Recommended practices
Tool name – version	
Used in tasks	list of process steps (or tasks)
Data (input / output)	Input: summary of input data
	Output: summary of input data
Performance	Excellent / OK / Pure
Issues	issues that were entered to the bug list provided by the tool developers
Wishes	wishes that were entered to the wish list provided by the tool developers
Comments	comments that were entered in the comment list provided by the tool developer
Experiences and recommended practices	any info on tool execution that could be important to tool developers

Scenario execution

Scenario	1. SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format
Started	date
Completed	date
Status	Not started, Started, Delayed, Pending, Aborted, Completed
Comment	reason for Pending / Aborted / Delayed status or any important comments at process step level

Pre-Ingest / Ingest / Access steps

Process step*	name of the process step from Pilot Definition excel
Started [*]	date
Completed [*]	date

Status [*]	status at the end of the reporting period (Not started, Delayed, Started, Pending, Aborted, Completed)
Duration [*]	duration of the process (only for Completed tasks)
Comment*	reason for Pending / Aborted / Delayed status or any important comments at process step level
Task [*]	name of the task within the process step (each task must have a separate process step table, see sample on Pilot 7)
Used tools [*]	empty if detail fields are filled or summary of tools if detail fields are empty (Manual, Local tool name)
Tool	tool name (Indicates if a tool is not developed by using E-ARK \rightarrow "local")
Version	(mandatory for E-ARK tools)
Input	input summary
Output	output summary
Performed by	task actor (e.g. Archivist, IT specialist, Technical administrator, etc.)
Performance	any performance related info
lssues	all bugs, wishes, comments (that were entered in any of the lists provided by the tool developer)
Experiences / Recommended practices	any important info on tool execution
Data	empty or "None" or "Not relevant"
Input data [*]	empty if detail fields are filled or summary of input data if detail fields are deleted
Description	input data description
Content type	type of content
Metadata format	format of the metadata
Volume	volume of input data
Data manipulation tasks	further data manipulation activities (if any)
Output data [*]	empty if detail fields are filled or summary of input data if detail fields are deleted
Description	output data description
Content type	type of content
Metadata format	format of the metadata
Volume	volume of output data
Data manipulation tasks	further data manipulation activities (if any)
Internal data manipulation tasks	further task-internal data manipulation (if any)
Task description	description of the data manipulation activities
Input	internal input

Output	internal output
IP usage [*]	empty if detail fields are filled
	or summary of IPs implemented if detail fields are deleted
IP type	SIP, AIP, DIP
	(indicate if not E-ARK specification compliant $ ightarrow$ "local")
Description	IP description (structure, content)
Mapping concerns	any important metadata mapping related info
Content concerns	any important content related info
IP related issues, comments	important information for WPs responsible for the IP specification
Data related issues, comments	issues/comments worth mentioning (but not tool or IP related)
Data management experiences and	any important info on data handling
best practices	
Used resources [*]	empty or "None"
Human resource	number of Archivists, IT specialists, Technical administrators, etc.
IT resource	IT environment, hardware and base software (any resources
(PCs, servers, architecture, OS, DB,	important to reproduce the pilot)
)	

Pilot documentation

At the end of October 2016 we had published deliverable D2.4 Pilot Documentation. This document had two parallel goals. On one hand it is the latest version of the documentation followed by the pilots. It contains an updated version of the pilot definition excel spreadsheet, the latest version of the actions to be performed with the latest tool versions within the pilot period (month 28-33). It also provides the latest snapshot with the most up-to-date information on pilot execution as we have performed it. On the other hand this documentation is the most comprehensive set of instructions and information that could be provided to archives outside the project. It is useful for archives and archivists who would like to use our outputs and repeat, in whole or in part, the pilot activities. The documentation includes an overview document by WP2, the updated pilot definition files and detailed description of the scenario execution by each of the pilot sites. These documents, created by the pilot representatives, lead the user through the pilot process via a step-by-step explanation with user screen examples.

An updated version of the documentation has been delivered in January 2017 along with updated documentation for Pilot 3.

For details, please read the complete D.2.4 Pilot Documentation here:

Part 1: http://eark-project.com/resources/project-deliverables/87-d24docs-p1-1

Part 2: http://eark-project.com/resources/project-deliverables/88-d24docs-p2-1

Changes to the planned pilot activities

At the execution phase there were some changes compared to the original workplans. These mainly extended the scope of the pilots and are shown below:

Pilot 1 – No changes

- Pilot 2 The National Archives of Norway (NAN) wanted to test the full spectrum of the ESSArch tool set. The ESSArch Tool for Producers (ETP) is a component to help producers create SIP packages. The producer partners of NAN on the other hand use a previous version of this tool which creates NOARK (the Norwegian standard) output. NAN has therefore performed an additional scenario to test ETP. The ETP tool has also been tested in Pilot 5.
- Pilot 3 Pilot 3 was supposed to perform pre-ingest scenarios with the ERMS Export Module but used the native export functionality of their DELTA ERMS system because of the late deployment of the appropriate ERMS Export Module version corresponding to the local producer's requirements. The ERMS Export Module was tested in 2 additional scenarios.
- Pilot 4 Pilot 4 had planned only 1 scenario with DBPTK but actually performed 3 more scenarios and all 4 were extended by a DBVTK restore database step as well.
 RODA-In was not used in this pilot because the native SIP creation tool was required to ingest into the preservation system of the Business Archives. RODA-In, on the other hand, was tested in Pilot 5 and 7.
- Pilot 5 No changes
- Pilot 6 At the pilot planning phase the Porto Municipality in Portugal also showed great interest in participating in an automatic ingest scenario. So a second scenario was planned with the same E-ARK component and infrastructure. Subsequently, there were some resource planning problems with their local developer who was needed to implement the producer-side infrastructure. The discussions and preparations continued until August 2016, when the Porto Municipality finally decided to delay the project. It is still possible that in the near future this scenario can be executed, but this will be beyond the timescales of this project.

Pilot 7 – No changes

Additional scenarios and External evaluation

Beside the 25 scenarios of the 7 full-scale pilots we have performed several additional scenarios. Additional scenarios, according to the Description of Work, are other, simpler scenarios also performed by the E-ARK members. Additional scenarios are either parts of the planned full-scale scenarios that, for some various (timing, not enough support from producer, late development), could not be performed within scope of the full-scale pilots or additional steps the pilot team wanted to try.

An external evaluation or validation, according to the Description of Work, is an evaluation or implementation of E-ARK products by members of DLM Forum and DPC or third parties outside the project with limited involvement from consortium members. We have supported 5 external evaluations by 5 different institutions from around the world. Some scenarios are completed and highly successful, some are still in progress or in preparation phase.

Additional scenarios and external evaluations, because they were outside the scope of the Description of Work, could not be planned in the same manner and in the same detail as the full-scale pilots were. They were prepared according to the results of other project activities and according to the needs and resources of the external partners.

Additional scenarios are presented along with the full-scale scenarios in this document because they were performed by the same pilot team. External evaluations are detailed in a separate chapter (Chapter 4.8).

Pilot evaluation

Evaluating success criteria

In the D2.3 Detailed Pilot Requirements document we have defined several success criteria at project, pilot and scenario level for the 25 scenarios of the 7 full-scale pilots. The evaluation of the pilots against these criteria can be found in Chapter 5. of this document.

E-ARK Final conference

At the E-ARK Final conference we had a session related to the experiences with the pilots. After an overview of the piloting activities each full-scale pilot representative gave a presentation on pilot execution, results and lessons learnt. The session ended with a panel discussion with all the pilot staff at the table and the audience could provide their opinion and ask questions about the pilots.

Recommended practices and lessons learned

Collecting and publishing recommended practices along with other pilot results is one of the most important objectives of the E-ARK project. Recommended practices and lessons learned are the essence of the all the pilot planning and execution activities.

With this in focus we have been collecting our experiences in the form of recommended practices and other comments during both the planning and execution phase of the pilots. During (and) after the execution period of the pilots recommended practices and comment have been registered at different levels.

- Tool related notes at the GitHub page of the tool developers
- Format specification related notes in a Google Drive Excel table
- Other recommended practices in a Google Drive Excel table
- All kinds of comments on pilot experience in the Monthly pilot report

Pilot level recommendations about the usage of the tools and specifications are presented as separate chapters in the main chapter for each pilot in the Pilot report part of this document.

D2.5 Final public report (this deliverable)

This deliverable summarizes the pilot planning and execution activities of the project. It provides details on the pilot execution and recommended practices when using E-ARK tools or format specifications.

Overview of the E-ARK Pilots

In the scope of the E-ARK project the format specification and tool development have been performed by the 4 technical work packages:

WP3

- Supplier Information Package (SIP) information package format specification
- SIARD 2.0 content type standard for archiving databases,
- SMURF (ERMS) and SMURF (SFSB) content type defined by E-ARK to archive ERMS system or simple file system based records,
- Content type specification to store Geodata information during the archival and dissemination processes,
- Data export and SIP creation tools supporting pre-ingest processes.

WP4

- Archival Information Package (AIP) information package format specification,
- SIP validation and SIP to AIP conversion tools supporting ingest processes.

WP5

- Dissemination Information Package (DIP) information package format specification,
- DIP creation and content viewers tools supporting access processes.

WP6

• Integrated Prototype (E-ARK Web) – a complete reference implementation consisting of several stand-alone tools supporting the full spectrum of OAIS processes.

In order to test the format specifications and tools developed by the project several pilot scenarios have been planned and performed during project. The pilots have been organized in seven full-scale pilots, each performed by one of the archival institution partners in E-ARK. (And one performed by an archival solution provider KEEP Solutions).

In the scope of the seven full-scale pilots we have defined 25 scenarios testing all the tools and formats developed and specified by E-ARK in different combinations, different business and IT environments, according to different archival strategies.

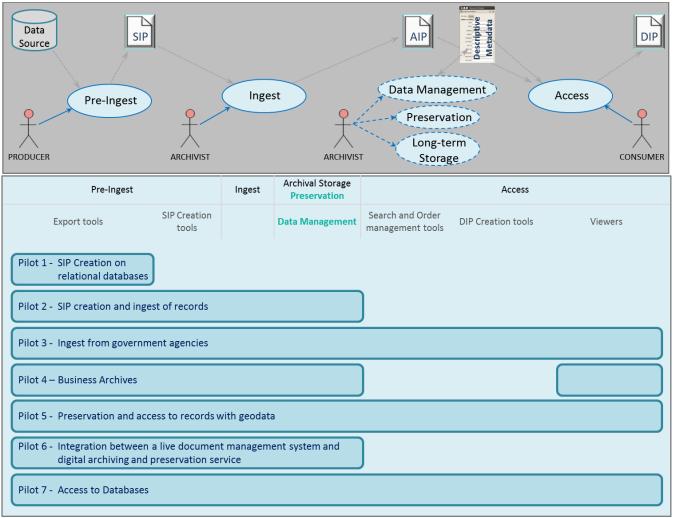
Some pilots were focusing on specific tools or processes of the OAIS models (1, 2, 4, 5, 6), others on archival and access of specific content types (4, 5, 7), one on automated ingest (6), and two pilots had scenarios to test the full spectrum of the OAIS processes along with the reference implementation: E-ARK Web (5,7). Some pilots followed a business-as-usual strategy (1, 2, 4, 6), some piloted the tools in a combination of a test and the production environment (3, 5, 7). We have tested both deployment versions of the E-ARK Web toolset, the virtual (5), and the full deployment (7).

Beside the 25 full-scale pilot scenarios the project has performed some smaller-scope additional scenarios and external evaluation scenarios, too. Additional scenarios are prepared and executed by the same pilot teams as the full-scale pilots. External evaluations are performed by non-E-ARK member organizations.

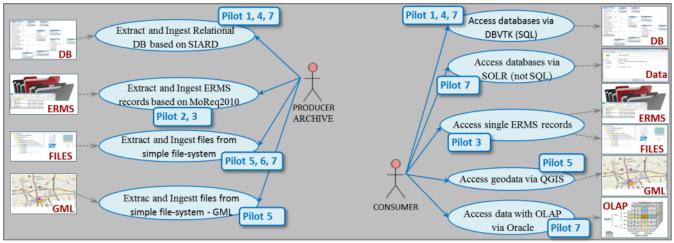
The following tables and graphs present the pilots and their relationships to other E-ARK elements. They help positioning the pilot scenarios on the OAIS map and among the various E-ARK tools and format specifications.

(The figures are from the E-ARK General Model version 2.2.)



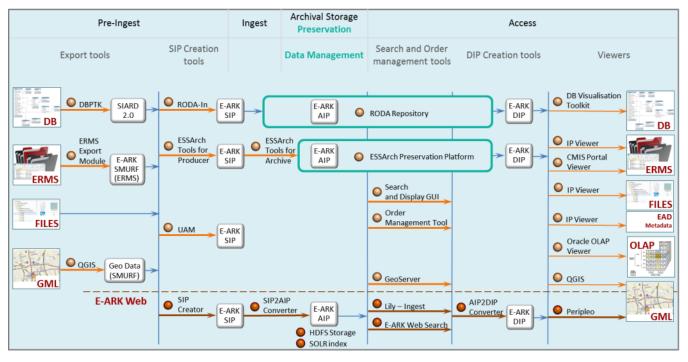


Full-scale pilots and E-ARK uses-cases

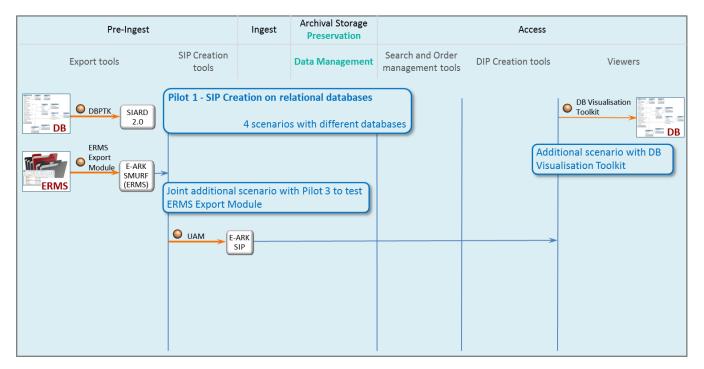


Pilots using E-ARK tools and format specifications

E-ARK Tools and Format Specifications



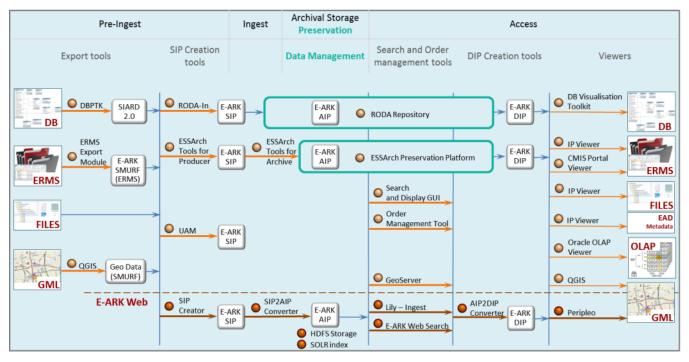
Pilot 1 – Danish National Archives

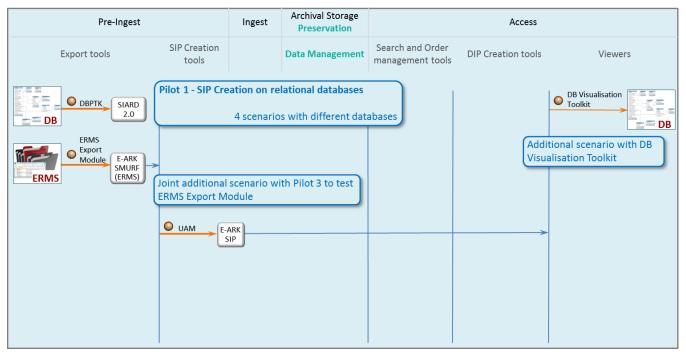


Pilot report

This section gives detailed information about the pilot scenarios performed in the scope of the E-ARK project.

Pilots 1 - SIP Creation on relational databases





Pilot 1	SIP	Creat	tion	on re	elatio	nal c	latab	ases	5															
Task leader	Dan	ish N	atio	nal A	rchiv	es																		
Supported by	Mag	genta	I																					
Scope		•									Crea n reco		ool v	with	not l	ess tl	nan 4	4 dat	abas	es of	diffe	erent	sizes	i
Object		ating uatic		for r	elatio	onal	datak	oases	s usir	ng th	e toc	l cre	ated	in W	/P3, 1	3.3:	SIP (Creat	ion T	ools	for	furth	er	
Short description		goal form		ne pil	ot is	to m	ake f	ours	succe	essfu	l dat	a ext	ractio	ons f	rom	live a	iuthe	entic	data	base	s into	o the	SIAR	D
Contacts	Nam	ne (T	itle)							E-m	ail								Skyp	be				
Contact Person	And	ers B	o Ni	elsen	1					<u>abn</u>	@sa.	<u>dk</u>												
Pilot staff member	Phill	ip M	ike T	ømn	nerho	olt				pmt	:@sa	dk							phill vet	liptoi	nme	rholt	_rigs	arki
OAIS Relevance		I	Pre-l	nges	t			Ing	est -	Stor	age						Sto	rage	– Ac	cess				
E-ARK Formats		E-ARK SIP E-ARK AIP E-ARK DIP																						
			S	IARD	2.0	Х		9	SMU	RF E	RMS			1	SMU	JRF S	FSB				1	Geo	data	
E-ARK Tools	× Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
Cooreria 1		+ :			£		_	- (D-		-+ 1)	ا مام	_	: .											·
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			<u> </u>									apas	e wit	in do	cum	ents	liarg	e)						
Additional scenario	· ·				Data																			
Additional scenario	Extr	act re	ecor	ds wi	th EF	RMS	Expo	rt Mo	odule	e and	d inge	est in	to Pr	eser	vica (Joint	: scei	nario	with	n NAE)			

Scenarios

Extrac	ctin	g recc	ords f	fron	n dat	abas	e (D	ata S	et 1)														
Extrac	ctin	g reco	ords f	rom	ı data	abase	e cor	ntaini	ng no	o doc	ume	nts.											
Pre-In	nges	,t																					
Extrac	ct ar	nd Ing	gest r	elat	iona	l data	abas	e bas	ed or	n SIA	RD 2	.0											
SIARD	2.0)																					
Datab	base	Prese	ervat	ion	Tool	kit																	
Health	h sy	stem	from	ו The	e Dar	nish I	Vatio	nal S	serun	n Inst	citute	į											
				-				n rep	orte	d inf	ectio	us di	seas	es at	a nat	iona	l leve	el. 50	-60 t	able	s and	abo	ut
Micro	osoft	t SQL	Serve	er 20	008																		
Not re	elev	ant																					
small				-																			
	P	re-In،	gest				Ing	est -	Stora	age						Sto	rage	- Acc	ess				
		E-,	ARK	SIP				E	-ARK	AIP										E	-ARK	DIP	
		SI/	ARD 2	2.0	Х			SMU	RF EF	۲MS				SMU	JRF S	FSB					Geod	data	
Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	Leservation Toolkit	Extracting Pre-Inges Extract ar SIARD 2.0 Database Health sy Database 90.000 re Microsoff Not relev small P	Extracting reco Pre-Ingest Extract and Ing SIARD 2.0 Database Press Health system Database cont 90.000 records Microsoft SQL Not relevant small Pre-Ing E- SI	Extracting records f Pre-Ingest Extract and Ingest r SIARD 2.0 Database Preservat Health system from Database containin 90.000 records in th Microsoft SQL Serve Not relevant small Pre-Ingest E-ARK SIARD 2.0 Not relevant	Extracting records from Pre-Ingest Extract and Ingest relat SIARD 2.0 Database Preservation Health system from The Database containing int 90.000 records in the m Microsoft SQL Server 20 Not relevant Small Pre-Ingest E-ARK SIP SIARD 2.0 I an power SIARD 2.0 Database containing int 90.000 records in the m Signal Pre-Ingest E-ARK SIP SIARD 2.0 Pre-Ingest E-ARK SIP	Extracting records from data Pre-Ingest Extract and Ingest relational SIARD 2.0 Database Preservation Tool Health system from The Dar Database containing inform 90.000 records in the main to Microsoft SQL Server 2008 Not relevant small Pre-Ingest SIARD 2.0 X Yathining Module SIARD 2.0 X Image: Simil stress SIARD 2.0	Extracting records from database Pre-Ingest Extract and Ingest relational data SIARD 2.0 Database Preservation Toolkit Health system from The Danish I Database containing information 90.000 records in the main table Microsoft SQL Server 2008 Not relevant small Pre-Ingest E-ARK SIP SIARD 2.0 X (L' YK Me SIARD 2.0 X (L' YK Me SIARD 2.0 (L' YK ME SIARD 2	Extracting records from database cor Pre-Ingest Extract and Ingest relational database SIARD 2.0 Database Preservation Toolkit Health system from The Danish Nation Database containing information from 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ing E-ARK SIP I SIARD 2.0 SIARD 2.0	Extracting records from database containi Pre-Ingest Extract and Ingest relational database bas SIARD 2.0 Database Preservation Toolkit Health system from The Danish National S Database containing information from rep 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - E-ARK SIP E SIARD 2.0 X SMU SIARD 2.0 X SMU Ingest - E-ARK SIP E SIARD 2.0 X SMU Ingest - E-ARK SIP E SIARD 2.0 X SMU Ingest - E-ARK SIP E SIARD 2.0 X SMU Ingest - E-ARK SIP I SIARD 2.0 X SMU Ingest - Ingest -	Pre-Ingest Extract and Ingest relational database based or SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serun Database containing information from reporter 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Stora E-ARK SIP E-ARK SIARD 2.0 X SMURF EF Ingest - Stora Ingest - Stora	Extracting records from database containing no doc Pre-Ingest Extract and Ingest relational database based on SIA SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Inst Database containing information from reported infe 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage E-ARK SIP X Mep SIARD 2.0 X SMURF ERMS SIARD 2.0 X SMURF ERMS Ingest - Storage E-ARK AIP SIARD 2.0 X SMURF ERMS Ingest - Storage E-ARK AIP	Extracting records from database containing no docume Pre-Ingest Extract and Ingest relational database based on SIARD 2 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectio 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP SIARD 2.0 X SMURF ERMS is a small Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP SIARD 2.0 X SMURF ERMS is a small Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP SIARD 2.0 X SMURF ERMS is a small Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP SIARD 2.0 X SMURF ERMS Ingest - Storage Ingest - Storage Ing	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious di 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage I E-ARK SIP I SIARD 2.0 SIARD 2.0 Not relevant small I Pre-Ingest Ingest - Storage I E-ARK AIP I SIARD 2.0 SIARD 2.0 S	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious disease 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage F-ARK AIP SIARD 2.0 X SMURF ERMS I Ingest - Storage I I I Ingest - Storage I I I I I I I I I I I I I I I I I I I	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage E-ARK SIP E-ARK SIP E-ARK SIP E-ARK SIP SIARD 2.0 X SMURF ERMS SMU in District in Distri	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at a nat 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage E-ARK SIP E-ARK SIP SIARD 2.0 X SMURF ERMS SMURF S	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at a nationa 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage Sto E-ARK SIP E-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB ig or in the server is the	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at a national leve 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage E-ARK SIP E-ARK SIP E-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at a national level. 50 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage Storage - Acce E-ARK SIP Ingest - Storage Storage - Acce Ingest - Storage Ingest	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at a national level. 50-60 t 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage Storage - Access E-ARK SIP E-ARK AIP I SIARD 2.0 X SMURF ERMS SMURF SFSB I Ingest - Storage Storage - Access	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at a national level. 50-60 tables 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage Storage - Access E-ARK SIP E-ARK AIP A-ACCESS E-ARCESS E-ARCESS E-ARCESS E-ARCESS E-ARCESS E-ARK AIP A-ACCESS E-ARK AIP E-ARK	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at a national level. 50-60 tables and 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage Storage - Access E-ARK SIP E-ARK AIP E-ARK AIP E-ARK AIP E-ARK SIP E-ARK SIMURF ERMS SMURF SFSB Geoc	Extracting records from database containing no documents. Pre-Ingest Extract and Ingest relational database based on SIARD 2.0 SIARD 2.0 Database Preservation Toolkit Health system from The Danish National Serum Institute Database containing information from reported infectious diseases at a national level. 50-60 tables and abo 90.000 records in the main table. Microsoft SQL Server 2008 Not relevant small Pre-Ingest Ingest - Storage E-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB Geodata Ingest - Storage C - ARK DIP SIARD 2.0 X SMURF ERMS SMURF SFSB Geodata Ingest - Storage Ingest

Scenario 2	Extracting records from	n dat	tabase (Data Set 2)									
Description	Extracting records from	dat	abase containing no docun	nent	ts.							
OIAS relevance	Pre-Ingest	Ingest										
Use-case	Extract and Ingest relat	tract and Ingest relational database based on SIARD 2.0										
E-ARK specifications	SIARD 2.0	RD 2.0										
E-ARK Tools	Database Preservation	Tool	kit									
Data	Registry of Cultural Eve	nts f	rom Kultunaut Aps									
Description	Database from the commercial company Kultunat Aps, which holds information about cultural events at a											
	national level, from eve	ents a	arranged by local communi	ities	s to cultural events fro	m th	e Danish cultural					
	institutions. The databa	ise c	ontains more than 5 millio	n re	cords.							
Data type	MySQL											
Metadata format	Not relevant											
Quantity	large											
OAIS Relevance	Pre-Ingest Ingest - Storage Storage - Access											
E-ARK Format	E-ARK SIP		E-ARK AIP				E-ARK DIP					
specifications	SIARD 2.0	SIARD 2.0 X SMURF ERMS SMURF SFSB Geodata										

|--|

Scenario 3	Extr	actir	ng re	cords	fror	n dat	abas	ie (D	ata S	et 3)														
Description	Extr	actin	ng rea	cords	fron	n dat	abas	e cor	ntaini	ing d	ocum	nents												
	The	DNA	will	go to	the	prod	ucer	's site	e wit	h the	e tool	on a	USB	. The	DNA	will	toge	ther	with	the	prod	ucer	use t	he
	tool	and	mak	e ext	ractio	ons ii	nto t	wo fo	orma	ts: SI	ARDI	DK ar	nd SI/	ARD2	.0.									
OIAS relevance	Pre-	Inge	st																					
Use-case	Extr	act a	nd Ir	ngest	relat	iona	l data	abas	e bas	ed o	n SIA	RD 2	.0											
E-ARK specifications	SIAF	RD 2.	0																					
E-ARK Tools	Data	abase	e Pre	serva	tion	Tool	kit																	
Data	Adm	ninist	trativ	e sys	tem	from	The	Dani	ish N	ation	al Ar	chive	es											
Description				ntainii se coi	-							-			earch	data	a, and	d pul	olic d	elive	ries d	of res	searc	h
Data type	Mic	rosof	ft SQ	L Serv	ver 2	008																		
Metadata format	Not	relev	vant																					
Quantity	sma	II																						
OAIS Relevance			Pre-l	ngest	:			Ing	est -	Stora	age						Sto	rage	- Acc	cess				
E-ARK Format			E	E-ARK	SIP				E	-ARK	AIP										E	-ARK	DIP	
specifications			S	SIARD	2.0	х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	data	
E-ARK Tools	× Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	Х																							

Scenario 4	Extracting records from database (Data Set 4)
Description	Extracting records from database containing documents.
	The DNA will go to the producer's site with the tool on a USB. The DNA will together with the producer use the
	tool and make extractions into two formats: SIARDDK and SIARD2.0.
OIAS relevance	Pre-Ingest
Use-case	Extract and Ingest relational database based on SIARD 2.0
E-ARK specifications	SIARD 2.0
E-ARK Tools	Database Preservation Toolkit
Data	Administrative and health records system from Ministry of Higher Education and Science.
Description	Studenterrådgivningen is an institution under Ministry of Higher Education and Science, whose purpose is to

Data type	situa MS :	ation SQL S	. The Serve	l, psy e data er 200	abase	-						-				nt to	stud	ents	in th	eir e	duca	tiona	al	
Metadata format		relev	/ant																					
Quantity	large	e																						
OAIS Relevance		I	Pre-l	ngest				Ing	est -	Stora	age						Sto	rage	- Acc	cess				
E-ARK Format			E	-ARk	(SIP				E	-ARK	AIP										E	-ARK	DIP	
specifications			S	IARD	2.0	Х			SMU	RF EF	RMS				SMU	JRF S	SFSB					Geod	data	
E-ARK Tools	× Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer

Please note that you can find more details with screenshots on scenario execution in the previous deliverable <u>D2.4</u> <u>Pilot Documentation</u>.

Additional scenarios

Additional scenario	Expe	xperiments with Database Visualization Toolkit																						
Description	The	he users search the database for information with real-life search scenarios.																						
OIAS relevance	Part	of a	ccess	5																				
Use-case																								
E-ARK specifications	non	е																						
E-ARK Tools	Data	abase	e Visi	Jaliza	tion	Tool	kit																	
Data																								
Description	Data	abase	e con	taini	ng fil	m ar	nd rel	lated	data	1														
Data type	Mic	rosof	ft SQ	L Serv	ver 2	800																		
Metadata format	Not	relev	/ant																					
Quantity	sma	II																						
OAIS Relevance			Pre-l	ngest	:			Ing	est -	Stora	age						Sto	rage	- Acc	cess				
E-ARK Format			E	-ARK	SIP				E	-ARK	AIP										E	-ARK	DIP	
specifications			S	IARD	2.0	Х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	data	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	× Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer

Additional scenario	Extr	act r	ecor	ds wi	ith El	RMS	Expo	rt M	odul	e an	ding	est ir	nto P	resei	rvica	(Join	t sce	nario	o wit	h NA	E)			
Description	NAE	NAE was supposed to use the ERMS Export Module to export records from ERMS but because of the late																						
	dep	deployment of the tool NAE had to use a local export tool to complete the full-scale pilot. To test the ERMS																						
	Expo	xport Module a joint additional scenario has been executed. DNA exported the records from Alfresco ERMS vith the newly deployed ERMS Export Module and sent the SMURF ERMS file to NAE where a SIP was created,																						
	with	n the	new	ly de	ploye	ed ER	IMS E	хро	rt Mo	odule	and	sent	the	SMU	RF EF	RMS	file to) NAI	Ewh	ere a	SIP v	was c	reate	ed,
	and	inge	sted	to Pr	eser	vica.	With	this	addi	tiona	al sce	naric	o eve	ry ste	ep th	at wa	as ori	gina	lly pl	anne	d to l	be te	sted	in
	Pilo	t 3 h	as be	en sı	ucces	sfully	y test	ted.																
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd Ir	ngest	ERM	lS red	cords	base	ed or	n Mo	Req2	010												
E-ARK specifications	SMU	JRF E	RMS	;																				
E-ARK Tools	ERM	/IS Ex	port	Mod	ule																			
Data	ERN	/IS sy	stem	of Tl	he Da	anish	Scho	o loc	f Me	dia a	nd Jo	urna	lism	(Dan	mark	s Me	edie-	og Jo	ourna	alisth	øjsko	ole) (DMJ)	()
Description	Diffe	Different kinds of letters and documents																						
Data type	Rec	Records from Alfresco ERMS																						
Metadata format	EAD)																						
Quantity	121	files	, 17 M	ИВ																				
OAIS Relevance			Pre-l	ngest	:			Ing	est -	Stor	age						Sto	rage	- Acc	cess				
E-ARK Format			E	-ARk	(SIP				E	-ARK	AIP										E	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EI	RMS	Х		SMURF SFSB Geodata										
E-ARK Tools	t			(d.			٩)			2										t.				
	olki			Ē	e		(ET			forn										olki				
	ιTo			ncer	npo	(q	live			Plat			=	8						10 10				
	atior	ule		rod	ß٨	We	Arch	eb)		ion			ל פר	lt T				_	eb)	atio			/er)	2
	erva	Podia Policia																						
	res	Database Preservation Toolkit ERMS Export Module RODA-In ESSArch Tool for Producer (ET Universal Archiving Module SIP creator (E-ARK Web) SIP creator Preservation Platform HDFS-Storage SOLR Index SSArch Preservation Platform HDFS-Storage SOLR Index SSArch Preservation Platform Broch Platform COLR Index Search and Display GUI Order Management Tool Lify - Ingest Search and Display GUI Order Management Tool Lify - Ingest Search and Display GUI Order Management Tool Database Visualization Toolkit Pripleo Database Visualization Toolkit Pripleo Oracle (OLAP Viewer)																						
	ase F	Database Preservation Toolkit ERMS Export Module RODA-In ESSArch Tool for Producer (ETP) Universal Archiving Module SIP creator (E-ARK Web) ESSArch Tools for Archive (ETA) ESSArch Tools for Archive (ETA) SIP2AIP (E-ARK Web) SIP2AIP (E-ARK Web) SIP2AIP (E-ARK Web) BAF5-Storage SOLR Index SOLR Index SOLR Index E-ARK Web Search AIP5DIP (E-ARK Web) Crder Management Tool Lily - Ingest Geoserver Geoserver BAR Web Search AIP2DIP (E-ARK Web) Database Visualization Toolkit IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer																						
	taba	MS F	RODA-In	SArc	iver		SArc	SIP2AIP	DA	SArc	FS-S	LR II	arch	der I	/ - In	oser	IIS	RK	2DI	taba	IP Viewer	Peripleo	acle	
	Da	ER	RO	ES	٩	SIP	ES	SIP	RO	ES	ЯH	so	Se	ō	Lil	Ge	QGIS	E-4	AIF		Ъ	Pel	Ori	S
																				Х				

Execution report

Please note that SIARD DK is a standard database preservation format in Denmark. This is the reason for creating (non-E-ARK) SIARD DK packages besides the SIARD 2.0 packages in Pilot 1. 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

Scenario	Started	Completed	Summary
1. Extracting records from database	May	September	<u>SIARD2.0:</u>
(Data Set 1) - database with no documents	2016	2016	 100% extraction of all tables and their data. The DNA has manually validated the SIARD-package up against the <i>"eCH-0165 SIARD Format Specification 2.0"</i>. There is no automatic tool for this yet. SIARDDK: 100% extraction of all tables and their data. The DNA has validated against <i>"Executive Order on Submission Information Packages"</i> and found no errors in the product.

			644992.0
2. Extracting records from database	June	September	<u>SIARD2.0:</u>
(Data Set 2) - database with no documents (large)	2016	2016	 100% extraction of all tables and their data. The DNA has manually validated the SIARD-package up against the <i>"eCH-0165 SIARD Format Specification 2.0"</i>. There is no automatic tool for this yet. <u>SQL Server:</u> SIARD-file was successfully uploaded to a MS SQL Server. First attempt failed due to differences in primary key names from PostgreSQL. Key names were manually altered and created new SIARD-file and successfully exported to MS SQL Server. <u>SIARDDK:</u> 100% extraction of all tables and their data. The DNA has validated against <i>"Executive Order on Submission Information Packages"</i> and found no errors in the product.
3. Extracting records from database	July	September	<u>SIARD2.0:</u>
(Data Set 3) - database with documents	2016	2016	 100% extraction of all tables and their data in one single SIARD-file. The DNA still has to export with a split to a SIARD-file and an external LOB-folder. The DNA also needs to validate the SIARD-package up against the <i>"eCH-0165 SIARD Format Specification 2.0"</i> SIARDDK: 100% extraction of all tables and their data. The DNA has validated against <i>"Executive Order on Submission Information Packages"</i> and found no errors in the end product.
4. Extracting records from database	August	September	<u>SIARD2.0:</u>
(Data Set 4) - database with documents (large)	2016	2016	 <u>SIARD2.0:</u> 100% extraction of all tables and their data. The DNA has manually validated the SIARD-package up against the <i>"eCH-0165 SIARD Format Specification 2.0"</i>. There is no automatic tool for this yet. <u>SIARDDK:</u> 100% extraction of all tables and their data.

Additional scenarios

Scenario	Started	Completed	Summary
Extract records with ERMS Export Module	December	December	Successful extraction of 120 files. The SMURF ERMS file
and ingest into Preservica (Joint scenario with NAE)	2016	2016	was sent to NAE for SIP creation and ingest. (for more details see the documentation of Pilot 3)
Experiments with Database Visualization	November	December	4 archivists tested the DBVTK application with real life

Toolkit	2016	2016	scenarios on a movie database looking for answers to questions like "What langue is used in this film?" or "Which stars plays in the movie?" They compered DBVTK to the local search capabilities and screens of the database.
			The users were absolutely satisfied with the logic and design of the tool and mentioned several clever ideas compared to the search and display functions of Sofia. They had many recommendations for the tool developer. (see Recommended practices later in this chapter)

Changes to the original plans

There were no changes. The scenarios have been performed according to plans in DoW and D2.3 Detailed Pilot Requirements.

Feedback report

The following table summarizes the feedback communication between the pilot staff and tool developers or format specification providers.

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
Database Preservation Toolkit	For the complete issue history, please refer to the GitHub page:
(version2.0.0-beta4.2)	https://github.com/keeps/db-preservation-toolkit
Used in tasks	Data extraction – all scenarios
Data (input / output)	Input: 4 databases from different producers
	Output: 1 SIARD2.0 package + 1 SIARDDK package.
Performance	Excellent with SIARD 2.0 (OK with SIARD DK)
Issues	There have been several issues with DBPTK related SIARD 2.0 output. KEEP Systems has corrected all the bugs and the response time was excellent. After the completion of the scenarios no known issues remained.
Wishes	A tool or function for automatic validation of SIARD 2.0 would be nice to have.
Comments	None
Experiences and recommended practices	After correcting the early bugs the tool functioned properly.

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
Database Visualization Toolkit	
Used in Additional scenario	Experiments with Database Visualization Toolkit
Data (input / output)	Movie database
Performance	Good
lssues	No issues found
Wishes	Users recommend showing technical information about the database on a separate page.
Comments	
Experiences and recommended	

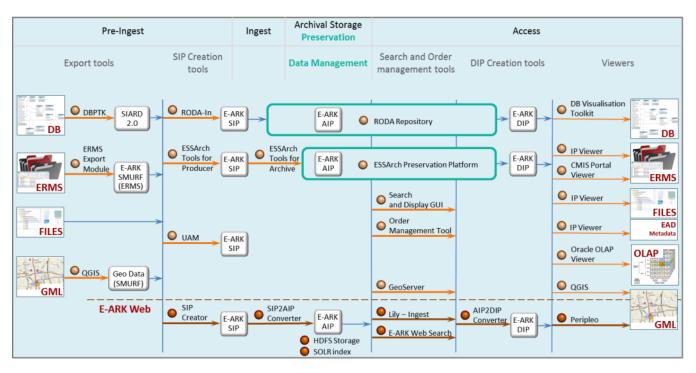
practices	

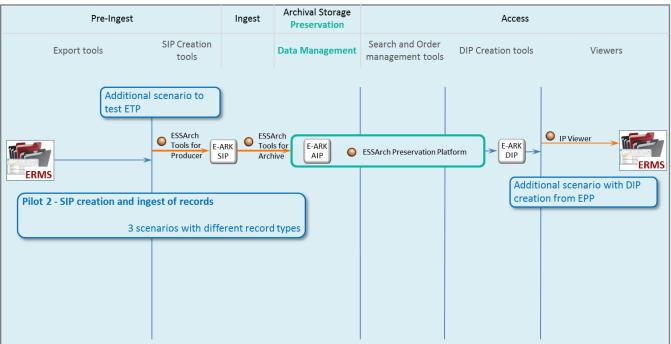
E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
ERMS Export Module	
Used in Additional scenario	Extract records with ERMS Export Module and ingest into Preservica
	(Joint scenario with NAE)
Data (input / output)	ERMS system of The Danish School of Media and Journalism (Danmarks Medie- og
	Journalisthøjskole) (DMJX)
Performance	Good
Issues	No issues found
Wishes	
Comments	
Experiences and recommended	
practices	

Recommended practices and further recommendations The following table contains the recommended practices and further development suggestions collected during pilot execution and evaluation.

Category	Relates to	Recommended practices / Further developments
Further requirement	SIARD 2.0	A tool or function for automatic validation of SIARD 2.0 would be required
Further recommendation	DBPTK documentation	It would be nice if there were more documentation on which user roles and privileges the tool works best under
Further recommendation	DBVTK	Users made a very detailed analysis of the tool and have a lot of smaller recommendations and wishes. (for details see documentation of the additional scenario)

Pilots 2 - SIP Creation and ingest of records





Pilot 2	SIP	creat	tion	and i	nges	t of r	ecor	ds																
Task leader	Nati	ional	Arch	nives	of N	orwa	у																	
Supported by	ESS	Solut	tions																					
Scope		less ⁻ n 1 tr								reco	rds w	ith n	nixed	l rest	ricte	d an	d uni	restri	icted	mat	erial,	and	not l	ess
Object		act d ls, inរ្																				-	SArcł	า
Short description			•							•													stem: osito	
Contacts	Nam	ne (T	itle)							E-m	ail								Skyp	be				
Contact Person	Arne	e-Kris	stian	Gro	ven					arng	ro@	arkiv	verk	et.nc	2									
Pilot staff member	Terj	e Pet	tters	en-D	ahl					geih	au@	arkiv	verk	et.no	<u>)</u>									
Pilot staff member	Geir	^r Hau	-																					
Pilot staff member	Jørg	;en Ø	Ø. Vik-Strandli jorvik@arkivverket.no																					
OAIS Relevance		ſ	Pre-Ingest Ingest - Storage Storage - Access																					
E-ARK Formats			E	E-ARI	< SIP	Х			E	-ARK	AIP	Х									E	-ARK	DIP	Х
			S	SIARD	0 2.0				SMU	RF EF	RMS	Х			SMI	JRF S	FSB					Geo	data	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	× ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	× ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	× ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	× IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
Cooporio 1	SIP (Creat	tion	and I	nges	tofu	instru	uctur	ed re	ecord	ls (Da	ata S	et 1)	-			-					•		
Scenario 1	5																							
Scenario 2		Creat			-			uctur	ed re	ecord		ata S	et 2)											
	SIP (tion	and I	nges	t of u	Instru				ls (Da		-											
Scenario 2	SIP (SIP (Creat	tion a	and I and I	nges nges	t of u t of s	instru truct	ured	reco	ords (ls (Da		-											

Scenario 1	SIP	Crea	tion	and l	nges	t of ι	unstr	uctu	red r	ecor	ds (D	ata S	et 1)										
Description	Extr	act u	Instru	uctur	ed re	ecord	ls fro	m ED	DRMS	base	ed or	the	Nor	wegia	n NC)ARK	4 sta	anda	rd. C	reate	e SIP	using	5	
	ESSA	Arch	Tool	s. Ing	est tl	he SI	P to	the r	epos	itory	using	g ESS	Arch	Pres	ervat	tion I	Platfo	orm,	for f	urthe	er eva	aluati	on.	
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd Ir	ngest	ERM	lS rec	cords	s (sim	nilar t	o Mo	Req	2010)											
E-ARK specifications	E-AF	rk-Si	Р																					
E-ARK Tools	ESSA	Arch	Tool	Prod	ucer	(ETP	P), ES	SArcl	h Toc	l Arc	hive	(ETA), ES	SArch	n Pres	serva	tion	Platf	orm					
Data	Noa	rk 4	outp	ut fro	om El	DRM	S																	
Description	EDR	MS c	data f	from	publ	ic pro	oduc	er co	nver	ted ir	nto N	oark	4 οι	itput	(real	proc	ductio	on da	ata)					
Data type	Noa	rk 5	XML	file, c	docu	ment	ts in l	PDF/	A (or	a fev	v oth	ier sp	pecif	ied fo	ormat	ts), ir	n TAR	file						
Metadata format	XMI	.: ME	METS, PREMIS, ADDML (local)																					
Quantity	20G																							
OAIS Relevance			Pre-l	ngest	:			Ing	gest -	Stora	age						Stor	rage	– Ace	cess				
E-ARK Format			E	-ARK	SIP	Х			E	-ARK	AIP	Х									E	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geod	data	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
				х			Х			Х														

Scenario 2	SIP Creation and Ingest	t of ι	Instructured records (D	ata S	iet 2)								
Description	Extract unstructured re	cord	s from EDRMS based or	the	Norwegian NOARK 5 stand	lard. Create SIP using							
	ESSArch Tools. Ingest th	ne Sl	P to the repository using	g ESS	Arch Preservation Platforn	n, for further evaluation							
OIAS relevance	Pre-Ingest, Ingest												
Use-case	Extract and Ingest ERM	S red	cords (similar to MoReq	2010)								
E-ARK specifications	E-ARK-SIP												
E-ARK Tools	ESSArch Tool Producer (ETP), ESSArch Tool Archive (ETA), ESSArch Preservation Platform (EPP)												
Data	Noark 5 output from E	DRM	S										
Description	EDRMS data public pro	duce	er converted into Noark	5 out	put (real production data)								
Data type	Noark 5 XML file, docur	nent	ts in PDF/A (or a few oth	er sp	pecified formats), in TAR fil	e							
Metadata format	XML: METS, PREMIS, A	DDM	IL (local)										
Quantity	5 GB												
OAIS Relevance	Pre-Ingest		Ingest - Storage		Storag	je - Access							
E-ARK Format	E-ARK SIP	Х	E-ARK AIP	Х		E-ARK DIP							
specifications	SIARD 2.0		SMURF ERMS	Х	SMURF SFSB	Geodata							

E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
				Х			Х			Х														

Scenario 3	SIP	Crea	tion a	and l	nges	t of s	truc	tured	d rec	ords	(Data	ı Set	3)											
Description	Extr	act d	lata f	rom	old d	atab	ase c	outpu	ut, cr	eate	SIPs	or st	ruct	ured	recoi	rds u	sing l	ESSA	rch T	ools,	inge	st th	e SIP	s to
	the	repo	sitor	y usir	ng ES	SArc	h Pre	eserv	ation	Plat	form	, for	furth	ner ev	/alua [.]	tion.								
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd Ir	ngest	ERM	S red	cords	(sim	ilar t	o Mo	Req	2010)											
E-ARK specifications	E-AF	RK-SI	Р																					
E-ARK Tools	ESSA	Arch	Tool	Prod	ucer	(ETP), ES	SArch	n Toc	l Arc	hive	(ETA), ES	SArch	n Pres	serva	tion	Platf	orm					
Data	Old	data	base	(CSV)																			
Description	The	data	set l	nere i	is the	e nati	ional	regi	stry c	of lice	encec	l hun	ters	cont	ainin	g dat	a fro	m th	e pei	riod 1	L985	-1999	Э.	
Data type	CSV	form	format (input), tar file																					
Metadata format	XML	: METS, PREMIS, ADDML (local)																						
Quantity	Con	ntaining 338.500 registered persons. 105 MB																						
OAIS Relevance			Pre-li	ngest	:			Ing	est –	Stor	age						Sto	rage	- Acc	ess				
E-ARK Format			E	-ARK	SIP	Х			E	-ARK	AIP	Х									E	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geod	lata	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
				Х			Х			Х														

Please note that more details with screenshots on scenario execution are available in the deliverable <u>D2.4 Pilot</u> <u>Documentation</u>.

Additional scenarios

Additional scenario	Creating SIP with ESSArch Tool for Producer
Description	NAN wanted to test the EssArch Tool for Producer (ETP) in the full-scale pilot scenarios but because of the
	"business as usual" full-scale pilot strategy they had to use the previous version of this tool. NAN therefore
	tested ETP in an additional SIP creation scenario in a virtual environment. The SIP then was ingested to EPP (as

	with	n full-	scale	e scei	nario	s) in	the v	/irtua	al env	/ironi	ment													
OIAS relevance	Pre-	Inge	st																					
Use-case	Extr	act a	nd Ir	ngest	ERM	lS red	cords	s (sim	nilar t	o Mo	Req	2010)											
E-ARK specifications	E-AF	RK-SI	Р																					
E-ARK Tools	ESSA	Arch	Tool	Prod	lucer	(ETP)																	
Data																								
Description	Loca	al tes	t dat	а																				
Data type	Mic	rosof	ft and	d pdf	docu	ımer	nts																	
Metadata format	Not	relev	/ant																					
Quantity	sma																							
OAIS Relevance		Pre-Ingest Ingest – Storage Storage – Access																						
E-ARK Format			E	-ARk	(SIP	Х			E	-ARK	AIP	Х									E	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geod	lata	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
				х			Х			Х														

Additional scenario	Gen	erat	ing E	-ARK	DIP	from	ESS/	Arch	Pres	ervat	tion I	Platfo	orm											
Description	The	EssA	rch F	Prese	rvati	on Pl	atfor	rm (E	PP) i	s full	y E-A	RK co	ompa	atible	e. In t	his a	dditio	onal	scen	ario a	an E-	ARK I	DIP is	
	gene	erate	ed fro	om El	PP. TI	he sc	enari	io co	uld n	ot be	e yet	com	plete	ed be	cause	e of t	he st	rict N	lorw	egia	n dat	a har	ndling	3
	regu	latic	ons m	nake	it ver	y dif	ficult	to u	se ar	chive	ed da	ta.												
OIAS relevance	Acce	ess																						
Use-case	Acce	ess E	RMS	reco	rds																			
E-ARK specifications	SMU	JRF E	RMS	5																				
E-ARK Tools	ESSA	Arch	Pres	ervat	ion P	latfo	orm (I	EPP)																
Data	Sele	cted	arch	ived	data																			
Description	Diffe	erent	t kind	ls of	lette	rs an	d do	cume	ents															
Data type	Mic	rosof	ft and	d pdf	docu	umer	nts																	
Metadata format	Not	t relevant																						
Quantity	sma																							
OAIS Relevance			Pre-l	ngest	t			Ing	est –	Stor	age						Stor	rage	– Ace	cess				
E-ARK Format			E	E-ARk	(SIP				E	-ARK	AIP	Х									E	-ARK	DIP	Х
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geod	lata	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	× ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	× IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
										~											~			

Execution report

Scenario	Started	Completed	Summary
1. SIP Creation and Ingest of unstructured records (Data Set 1)	May 2016	September 2016	After a longer testing period the scenario has been performed as planned.
2. SIP Creation and Ingest of unstructured records (Data Set 2)	June 2016	October 2016	After a longer testing period the scenario has been performed as planned.
3. SIP Creation and Ingest of structured records (Data Set 3)	May 2016	October 2016	After a longer testing period the scenario has been performed as planned.

Additional scenarios

Scenario	Started	Completed	Summary
Creating SIP with ESSArch Tool for Producer	November 2016	January 2017	The scenario has been performed successfully. The overall impression is that the tool is useful for data. providers/agencies.
Generating E-ARK DIP from ESSArch Preservation Platform	December 2016	Not yet finished	The scenario could not be yet completed because of the strict Norwegian data handling regulations make it very difficult to use archived data.

Changes to the original plans

The E-ARK compatible version of ESSArch Tool for Provider (ETP) could not be tested in the "business as usual" fullscale pilot because of data provider's IT infrastructure. The tool has been tested in an additional scenario by NAN. The ETP tool has also been tested in Pilot 5.

Feedback report

The following table summarizes the feedback communication between the pilot staff and tool developers or format specification providers.

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
ESSArch Tool for Producer (ETP)	For the complete issue history, please refer to the GitHub page:
v0.95	https://github.com/ESSolutions/ESSArch_Tools_Producer
Used in tasks	SIP Creation
Data (input / output)	3 different input sources at 3 data providers
Performance	Good
Issues	No issues left at scenario completion
Wishes	
Comments	NAN would like to evaluate on even larger data sets to conclude about scalability.
Experiences and recommended	The tool worked well
practices	

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
ESSArch Tools Archive (ETA)	For the complete issue history, please refer to the GitHub page:
v0.93.1	https://github.com/ESSolutions/ESSArch_Tools_Archive
Used in tasks	Ingest preparations
Data (input / output)	SIPs from 3 different input sources
Performance	Good
Issues	No issues left at scenario completion
Wishes	
Comments	NAN would like to evaluate on even larger data sets to conclude about scalability.
Experiences and recommended	To tools has been tested very thoroughly and all the bugs issues been solved before
practices	deployed in production environment. The tool was able to produce satisfactory results.

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
ESS Preservation Platform	For the complete issue history, please refer to the GitHub page:
v2.7.3	https://github.com/ESSolutions/ESSArch EPP
Used in tasks	Ingest, Long-term preservation
Data (input / output)	SIPs from 3 different input sources
Performance	Good
Issues	No issues left at scenario completion
Wishes	
Comments	NAN would like to evaluate on even larger data sets to conclude about scalability.
Experiences and recommended	To tools has been tested very thoroughly and all the bugs issues been solved before
practices	deployed in production environment. The tool was able to produce satisfactory result.

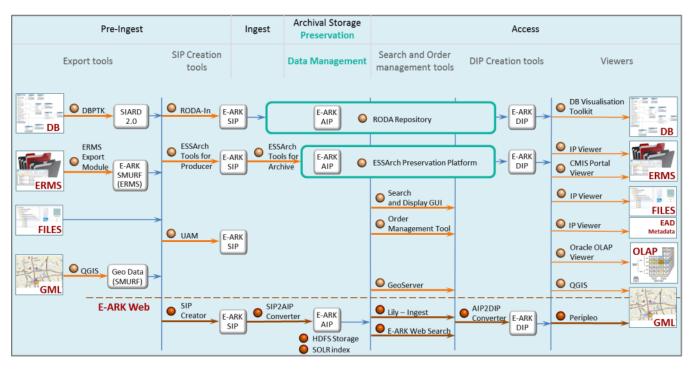
Recommended practices and further recommendations

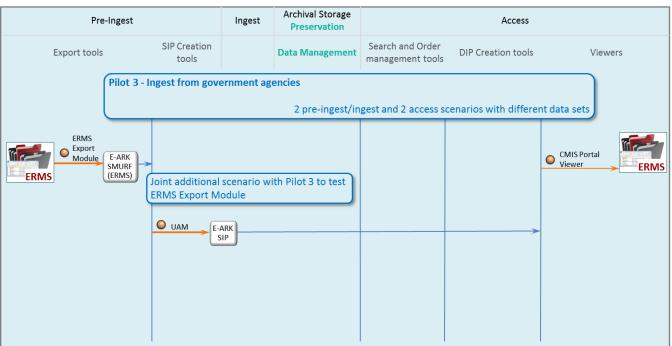
The following table contains the recommended practices and further development suggestions collected during pilot execution and evaluation.

Category	Relates to	Recommended practices / Further developments
Recommended practices	ETP	Submission Agreement (SA) profiles are configured in ETP, based on selecting sub-profiles of various categories such as "SIP profiles", "Submit description profiles", "Transfer project profiles" and more. The data providers/agencies using ETP should predefine their own sub-profiles according to their specific needs using the tool Profile maker, also developed by ES Solutions. Profiles must be locked before processing further, Therefore metadata must be edited before locking the profiles.
		Various degree of automation in ETP can be defined through definition of profiles. EAD and EAC-CPF schemas have to be provided with the content.
Recommended practices	ΕΤΑ	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.
Recommended practices	EPP	In EPP, AIPs are generated in an automatic manner using a queue-handling system. The AIPs can be stored on either tapes or disks.
Recommended practices	ЕТР, ЕТА, ЕРР	For installing the ESSArch ETP, ETA and EPP tools we recommend to get support from ES Solutions for installation and configuration of the application.
Further	Testing	Content size should also be tested a bit further, since the largest content of the original pilots

recommendation		were 20 GB
Further	SIP Format	A more flexible format specification would perhaps be more suitable in the future.
recommendation		

Pilots 3 - SIP Creation and ingest of records





Pilot 3	Inge	est fr	om g	govei	rnme	nt ag	genci	ies																
Task leader	Nat	ional	Arch	nives	of Es	toni	а																	
Supported by																								
Scope	Exp	ort p	ublic	reco	ords f	rom	an E	DRM	syst	em o	fago	overi	nmer	ntal a	igeno	y to	the l	Natio	nal A	Archiv	ves o	f Este	onia	and
	mał	ke the	ese a	vaila	ble t	hrou	gh o	ur ov	vn ca	talog	gue (i	.e. A	rchiv	al In	form	atior	n Sys	tem,	AIS)	as w	ell as	prov	vide a	an
	API	for a	cces	sing	the re	ecord	ls fro	om of	ther	syste	ms (t	he o	rigin	al EC	RMS	at tl	he ag	gency	/); Th	ie wł	ole s	et w	ill	
	incl	ude a	abou	t 500	00 red	cords	(but	: dep	ends	on t	he e>	kact a	agen	cy of	cou	ˈse).								
Object	Nat	ive E	DRM	S at	a gov	ernn	nent	al ag	ency	(Alfr	esco	DELT	ΓA), r	ecor	ds pı	epar	atio	n too	I (UA	M), (digita	al		
	pres	serva	ition	and	acces	ss sys	stem	s (Pre	eserv	vica, I	AIS)													
Short description	The	mair	n par	t of t	the p	ropo	sed p	oilot	inclu	des t	he e	(port	t of e	lectr	onic	reco	rds a	ind th	neir r	neta	data	from	EDR	М
					an pu									-				-			-			
					an ag							-												
					ble o																-	stand	lardiz	ed
	_	-		s me	thod	s tha	t are	imp	leme	1		ie ag	encie	es' di	gital	infra	stru	cture			e			
Contacts	Nan	ne (T	ïtle)							E-m	ail								Skyp	be				
Contact Person	Kar	in O	olu							kari	n.oc	olu@	ttu.	ee					kari	inoo	lu			
Pilot staff member	Tar	vo K	ärbe	erg						tarv	vo.ka	rbei	ˈɡ@ı	a.ee	2				tarv	/o.ka	rbei	g		
OAIS Relevance			Pre-l	nges	st			Ing	est -	Stor	age						Sto	rage	– Ac	cess				
E-ARK Formats			E	E-ARI	k SIP	Х			E	-ARK	AIP										E	-ARK	DIP	Х
			S	SIARE	0 2.0				SMU	RF EI	RMS	Х			SM	JRF S	SFSB					Geo	data	
E-ARK Tools				((
	lkit			ETP			ETA			r m										lkit				
	Database Preservation Toolkit			ESSArch Tool for Producer (ETP)	Universal Archiving Module	_	ESSArch Tools for Archive (ETA)			ESSArch Preservation Platform				-						Database Visualization Toolkit				
	- uo	a		quo	δ	SIP creator (E-ARK Web)	chiv	<u> </u>		n Pl			IJ.	Order Management Tool					-	- uo			÷	
	/ati	ERMS Export Module		Pro	ng I	×	An	(E-ARK Web)	_	atio			Search and Display GUI	ent				Ę	AIP2DIP (E-ARK Web)	zati			Oracle (OLAP Viewer)	/er
	serv	β		for	hivi	-AR	for	XX V	RODA Repository	e va			splå	e m				E-ARK Web Search	×Χ	illa			Vie	CMIS portal/viewer
	Pres	t		Į	Arcl	r (E	sloc	-AF	osit	ese	age	×	i Di	Jage	4			b Se	-AR	Visu			AP	 ∧ E
	se	, dx	2	μŢ	sal ,	atoi	μT		Rep	ЧЧ	tor	apu	anc	Mar	ges	ver		Wel	P (E	Ise 1	/er	0	0	ort
	aba	AS I	RODA-In	Arc	ver	cre	Arc	SIP2AIP	DA	Arc	HDFS-Storage	SOLR Index	rch	ler I	Lily - Ingest	Geoserver	S	RK	2DI	aba	IP Viewer	Peripleo	cle	IS p
	Dat	ER	ß	ESS	Ū n	SIP	ESS	SIP	RO	ESS	HD	sol	Sea	ő	Lily	Ge	QGIS	E-A	AIP	Dat	<u>ا</u>	Per	Ora	S
		Х			Х																			Х
Scenario 1	Extr	act r	ecor	ds fro	om E	DRM	(of a	a gov	ernn	nenta	al inst	tituti	on),	creat	te SIF	o and	inge	est to	Pres	servi	ca (D	ata s	et 1)	
Scenario 2	Prov	vide	acces	ss to	reco	rds fr	om (gove	rnme	ental	instit	tutio	n thr	ough	n RES	Tful	servi	ces (Data	set 1)			
Scenario 3	Extr	act r	ecor	ds fro	om E	DRM	(of a	a gov	ernn	nenta	al inst	tituti	on),	creat	te SIF	and	inge	est to	Pres	servio	ca (D	ata s	et 2)	
Scenario 4	Prov	vide	acces	ss to	reco	rds fr	om į	gove	rnme	ental	instit	utio	n thr	ough	RES	Tful	servi	ces (Data	set 2	2)			
Additional scenario	Extr	act r	ecor	ds w	ith EF	RMS	Ехро	rt M	odul	e and	l inge	est in	to Pr	eser	vica	(Join	t sce	nario	with	ו NA	=)			
Additional scenario	ERN	/IS Ex	port	Mod	dule s	cena	rio v	vith l	ocal	ERM	S sys	tem	DELT	A										

Scenario 1	Extr	act r	ecor	ds fro	om E	DRM	l (of a	a gov	ernn	nenta	al ins	titut	ion),	crea	te SI	P and	d inge	est to	Pre	servi	са			
Description	Expo	ort p	ublic	reco	rds f	roma	an EC	DRM	syste	em of	^f a go	vern	men	tal ag	gency	, cre	ate S	IP, a	nd in	igest	to th	ne Pre	eserv	ica
	syst	em a	it the	Nati	onal	Arch	ives	of Es	tonia	i .														
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd Ir	ngest	ERM	lS rec	cords	base	ed or	n Mol	Req2	010												
	(Alfr	resco	o is no	ot Mo	oreq-	com	plian	t sys	tem)															
E-ARK specifications	E-AF	RK-SI	P, SN	/URF																				
E-ARK Tools	Univ	versa	l Arc	hivin	g Mo	dule	(UA	M)																
Data	Reco	ords	and	meta	data	expc	orted	fron	n nat	ive E	RMS	(DEL	TA) E	xpor	t Mo	dule	at M	linist	ry of	Justi	ce of	^E Sto	nia	
Description	Data	a set	cons	ists o	of dif	feren	nt do	cume	ents d	of Mi	nistry	/ of J	ustic	e fro	m 6 s	series	s with	n diff	eren	t rete	entio	n pe	riod.	
Data type	ddo	c, do	ocx, P	DF, T	IFF																			
Metadata format	SMU	JRF E	RMS	5																				
Quantity	15 f	iles																						
OAIS Relevance			Pre-l	ngest				Ing	est -	Stora	age						Stor	rage	– Ace	cess				
E-ARK Format			E	-ARK	SIP	Х			E	-ARK	AIP										E	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geod	data	
E-ARK Tools				(d			G			5														
	Toolkit			ET (ET	e		(ET/			Platform										Toolkit				
	Toc			rcer	npo	(q	ive			Plat			_	ō						To				
	Preservation	ər		lodu	ž	We	Arch	eb)		- uo			GU	IT TO					(qə	Visualization			er)	<u>۔</u>
	irva	lodi		r Pi	ving	RK	or /	Š	≥	vati			olay	nen				rch	Ň	liza			iew	e ve
	rese	≥ 12		ol fc	rchi	(E-⊅	ols f	(E-ARK Web)	sito	ser	ge		Dis	ager				Sea	ARK	isua			ΡΛ	/vie
	ie Pl	podx	-	Tot	al A	tor	To	Ľ.	epo	Pre	ora	dex	pue	lané	gest	'er		Veb	Ľ.	ie V	Ŀ		OLA	portal/viewer
	abas	IS E	A-Ir	Arch	ers	crea	Vrch	AIP	AR	Vrch	S-St	R In	ch	er ⊘	- Ing	serv	s	× X	DIP	bas	ewe	plec	cle (S pc
	Database	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP	RODA Repository	ESSArch Preservation	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS
					X											-	-							

Scenario 2	Provide access to records f	rom governmental insti	tutio	n through RESTful serv	ices		
Description	Estonian agencies have the	responsibility to make p	ublic	electronic records with	n no a	access restrictions availa	ble
	on their web sites, which m	eans that the pilot will a	lso e	nable this through stan	dard	ized linking/access	
	methods that are implemer	nted in the agencies' dig	ital ir	nfrastructure / web site.			
OIAS relevance	Access						
Use-case	Access single ERMS records	via CMIS Browser					
	(To be consolidated with a (CMIS interface access so	lutio	n)			
E-ARK specifications	SMURF						
E-ARK Tools	CMIS Browser						
Data	Records and metadata expo	orted from native ERMS	(DEL	TA) Export Module at M	linist	ry of Justice of Estonia	
Description	Data set consists of differer	t documents of Ministr	/ of J	ustice from 6 series wit	h diff	erent retention period.	
Data type	ddoc, docx, PDF, TIFF						
Metadata format	SMURF ERMS						
Quantity	15 files						
OAIS Relevance	Pre-Ingest	Ingest - Storage		Sto	rage	- Access	
E-ARK Format	E-ARK SIP	E-ARK AIP				E-ARK DIP	Х
specifications	SIARD 2.0	SMURF ERMS	Х	SMURF SFSB		Geodata	

|--|

Scenario 3	Extr	act r	ecor	ds fro	om E	DRM	(of a	a gov	ernn	nenta	al ins	titut	ion),	crea	te SII	P and	l inge	est to	o Pre	servi	ica			
Description	Expo	ort p	ublic	reco	rds f	roma	an ED	DRM	syste	em of	[:] a go	vern	men	tal ag	gency	, cre	ate S	iP, a	nd in	gest	to th	e Pre	eserv	ica
	syst	em a	it the	e Nati	onal	Arch	ives	of Es	tonia	۱.														
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd Ir	ngest	ERM	lS rec	cords	base	ed or	Mol	Req2	010												
	(Alfı	resco	o is no	ot Mo	oreq-	com	plian	t sys	tem)															
E-ARK specifications	E-AF	rk-Si	P, SN	/URF																				
E-ARK Tools	Univ	versa	l Arc	hivin	g Mo	dule	(UAI	M)																
Data	Reco	ords	and	meta	data	expc	orted	fron	n nat	ive E	RMS	(via l	DELT	A) at	Mini	stry o	of Ju	stice	of Es	stonia	Э			
Description	Data	a set	cons	sists c	of dif	feren	t doo	cume	ents o	of Mi	nistry	/ of J	ustic	e fro	m dif	ferer	nt sei	ries.						
Data type	DDC	DC (a	file f	forma	at ho	lding	Esto	nian	digit	al sig	natu	re in	form	atior	ı), DC	DCX,	PDF,	TIFF						
Metadata format	SMU	JRF E	RMS	5																				
Quantity	200	files																						
OAIS Relevance			Pre-l	ngest	t			Ing	est -	Stora	age						Stor	rage	– Ace	cess				
E-ARK Format			E	E-ARk	(SIP	Х			E	-ARK	AIP										E	-ARK	DIP	
specifications			S	SIARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geod	data	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
					Х																			

Scenario 4	Provide access to records from governmental institution through RESTful services
Description	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 standardized linking/access
	methods that are implemented in the agencies' digital infrastructure / web site.
OIAS relevance	Access
Use-case	Access single ERMS records via CMIS Browser
	(To be consolidated with a CMIS interface access solution)
E-ARK specifications	SMURF
E-ARK Tools	CMIS Browser
Data	Records and metadata exported from native ERMS (via DELTA) at Ministry of Justice of Estonia
Description	Data set consists of different documents of Ministry of Justice from different series.
Data type	DDOC (a file format holding Estonian digital signature information), DOCX, PDF, TIFF
Metadata format	SMURF ERMS

Quantity	200	files																						
OAIS Relevance		l	Pre-l	ngest	t			Ing	est -	Stora	age						Stor	rage	– Acc	cess				
E-ARK Format			E	-ARk	(SIP				E	-ARK	AIP										E	-ARK	DIP	Х
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	SFSB					Geod	lata	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	× CMIS portal/viewer

Please note that you can find more details with screenshots on scenario execution in the previous deliverable <u>D2.4</u> <u>Pilot Documentation</u>.

Additional scenarios

Additional scenario	Extract records with EF	RMS	Export Module and ing	est ir	nto Preservica (Joint scenar	io with NAE)	
Description	The National Archives of	of Est	tonia was supposed to ι	ise th	e ERMS Export Module to	export records from ERM	IS
	but because of the late	dep	loyment of the tool NAE	had	to use a local export tool to	o complete the full-scale	
	pilot. To test the ERMS	Ехро	ort Module a joint addit	ional	scenario has been execute	d. DNA exported the reco	ords
	from Alfresco ERMS wi	th th	e newly deployed ERMS	Exp	ort Module and sent the SM	/URF ERMS file to NAE	
	where a SIP was create	d, ar	nd ingested to Preservic	a. Wi	th this additional scenario e	every step that was origin	nally
	planned to be tested in	Pilo	t 3 has been successfull	y tes	ted.		
OIAS relevance	Pre-Ingest, Ingest						
Use-case	Extract and Ingest ERM	S red	cords based on MoReq2	010			
E-ARK specifications	SMURF ERMS						
E-ARK Tools	ERMS Export Module						
Data	ERMS system of The Da	nish	School of Media and Jo	urna	lism (Danmarks Medie- og	Journalisthøjskole) (DMJ)	X)
Description	Different kinds of letter	rs an	d documents				
Data type	Records from Alfresco	ERM	S				
Metadata format	EAD						
Quantity	121 files, 17 MB						
OAIS Relevance	Pre-Ingest		Ingest - Storage		Storag	e - Access	
E-ARK Format	E-ARK SIP	Х	E-ARK AIP			E-ARK DIP	Х
specifications	SIARD 2.0		SMURF ERMS	х	SMURF SFSB	Geodata	

|--|

Additional scenario	ERM	1S Ex	port	Mod	ule s	cena	rio w	ith lo	ocal E	RMS	syst	em D	DELTA	4										
Description	This	addi	ition	al pilo	ot co	mbin	ies se	evera	l too	ls an	d tes	ts th	e E-A	RK w	orkfl	ow ii	n full	fron	n the	begi	nnin	g to t	he er	nd.
	Reco	ords	from	the	local	DELT	FA sy	stem	wer	e exp	orte	d wit	h ER	MS E	xpor	t Mo	dule	then	a SII	P was	s crea	ted	and	
	inge	sted	into	Pres	ervic	a. Fir	hally	the a	icces	s was	s pro	video	d by (CMIS	Port	al Vie	ewer							
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd Ir	ngest	ERM	lS red	cords	base	ed or	n Mol	Req2	010												
E-ARK specifications	SMU	JRF E	RMS	5																				
E-ARK Tools	ERN	1S Ex	port	Mod	ule																			
Data	Sele	cted	reco	ords f	rom	DELT	A ER	MS s	yster	n fro	m pa	rtne	r con	npan	y Wis	serca	t							
Description	Diffe	erent	t kind	ds of	docu	men	ts																	
Data type	Reco	ords	from	DEL	ta ef	RMS																		
Metadata format	Not	relev	vant																					
Quantity	A sn	nall a	mou	int of	reco	ords																		
OAIS Relevance		l	Pre-l	ngest	:			Ing	est -	Stora	age						Sto	rage	– Ace	cess				
E-ARK Format			E	E-ARk	SIP	Х			E	-ARK	AIP										E	-ARK	DIP	Х
specifications			S	SIARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geod	data	
E-ARK Tools	kit			ETP)			TA)			Ē										kit				
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	_	x			x		_									-								x

Execution report

The focus of Pilot 3 was the export of electronic records and their metadata from EDRM systems of Estonian public sector institutions, transfer and ingest to the NAE digital repository. 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.

Data has been 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.

NAE was supposed to use the ERMS export module to select and export records from the ERMS but the version compatible with the local DELTA system could not be launched before November 2016. The half year execution period of the full-scale pilots ended in October so NAE has decided to use the native export functionality of DELTA ERMS to create the E-ARK SMURF input for the SIP and perform an additional scenario with ERMS Export Module later. At the end two complete additional scenarios have been run, one in cooperation with the Danish National Archives.

Scenario	Started	Completed	Summary
1. Extract records from EDRM, create SIP and	May	November	After the very long preparation and local development
ingest to Preservica (Data set 1)	2016	2016	period the scenario has been successfully executed.
2. Provide access to records through RESTful	September	November	Access scenarios could start only after the ingest
services (Data set 1)	2016	2016	scenarios have been concluded. The scenario successfully
			completed. The SMURF file content is accessible through
			CMIS Portal Browser linked from producers corresponding
			web page.
3. Extract records from EDRM, create SIP and	May	December	After the very long preparation and local development
ingest to Preservica (Data set 2)	2016	2016	period the scenario has been successfully executed.
4. Provide access to records through RESTful	September	December	Access scenarios could start only after the ingest
services (Data set 2)	2016	2016	scenarios have been concluded. The scenario successfully
			completed. The SMURF file content is accessible through
			CMIS Portal Browser linked from producers corresponding
			web page.

Experience with piloted tools and specifications within the Pilot 3 was positive, they are compatible and widely usable.

Additional scenarios

Scenario	Started	Completed	Summary
Extract records with ERMS Export Module and ingest into Preservica (Joint scenario with NAE)	November 2016	December 2016	The joint scenario was a real success story. The preparations at both sites resulted in a smooth cooperation in order to export the selected records at DNA and create the ingest and provide access to data at NAE.
ERMS Export Module scenario with local ERMS system DELTA	November 2016	December 2016	This pilot was actually more than an additional scenario. The complete full-scale scenario that NAE planned to execute within the full-scale pilot has been performed. It's a wall-to-wall scenario from pre-ingest to access.

Changes to the original plans

NAE was supposed to use the ERMS export module to select and export records from the ERMS but the version compatible with the local DELTA system could not be launched before November 2016. The half year execution period of the full-scale pilots ended in October so NAE decided to use the native export functionality of DELTA ERMS to create the E-ARK SMURF input for the SIP and perform an additional scenario with ERMS Export Module later. At the end two complete additional scenarios have been run, one in cooperation with the Danish National Archives.

Feedback report

The following table summarizes the feedback communication between the pilot staff and tool developers or format specification providers.

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
ERMS Export Module	For the complete issue history, please refer to the GitHub page:
	https://github.com/magenta-aps/erms-export-ui-module
Used in additional scenario	Exporting ERMS Records
Data (input / output)	Tested with real-
Performance	Good
Issues	No issues left at scenario completion
Wishes	
Comments	
Experiences and recommended	
practices	

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
Universal Archiving Module (UAM)	
Used in tasks	SIP creation
Data (input / output)	Tested with two data sets of DELTA ERMS records
Performance	Good
Issues	No issues left at scenario completion
Wishes	None
Comments	None
Experiences and recommended	None
practices	
E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
CMIC Dentel Deserves	
CMIS Portal Browser	
Used in tasks	Access
	Access Tested with two data sets of DELTA ERMS records
Used in tasks	
Used in tasks Data (input / output)	Tested with two data sets of DELTA ERMS records
Used in tasks Data (input / output) Performance	Tested with two data sets of DELTA ERMS records Good
Used in tasks Data (input / output) Performance Issues	Tested with two data sets of DELTA ERMS records Good No issues left at scenario completion
Used in tasks Data (input / output) Performance Issues Wishes	Tested with two data sets of DELTA ERMS records Good No issues left at scenario completion None

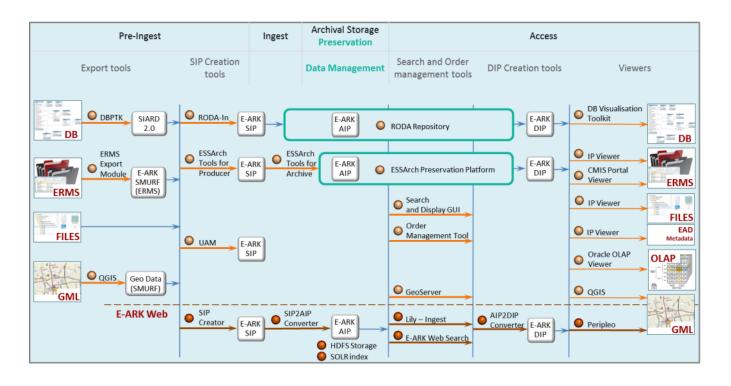
Although the tools and specifications proved to be usable, we are still planning to look for more possibilities to reduce the human factor and automate the workflow in the steps where it is possible in order to make the process even more scalable in the future.

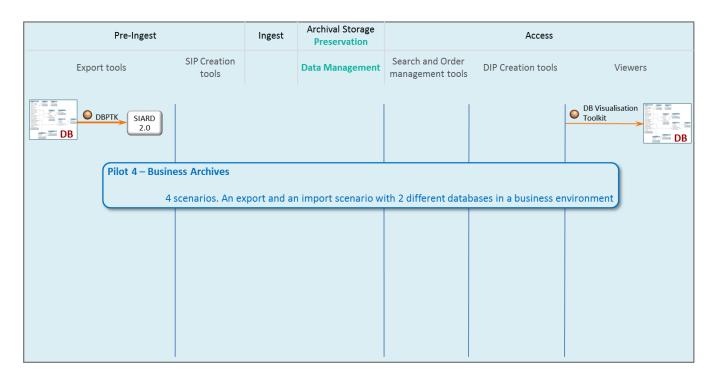
Recommended practices and further recommendations

The following table contains the recommended practices and further development suggestions collected during pilot execution and evaluation.

Category	Relates to	Recommended practices / Further developments
Recommended	UAM	Recommendations to data providers/agencies:
practices		 Allocate enough time for the first attempt of the transfer as there are plenty of useful functionalities in UAM which need time to get acquainted with;
		 The quality of ERMS exported data and metadata may not be sufficient for long time preservation and therefore it is necessary to consider whether the data may need to be rearranged and enriched with additional descriptive metadata before; Subsequent archival transfers will require less time.
		Recommendations to archives: - Continue UAM training in agencies; - Look for possibilities to enhance the user-friendliness and intuitive usage of UAM.
Recommended practices	CMIS Portal Browser	 Very useful and necessary tool which provides access to transferred data directly to digital archive. It allows users to see the document in the latest archival format; The tool is easy to configure. Link of the external interface of the digital archive will be given to
		the agency to configure the tool;
		- Easy to administer users. One administrator role will be given to the agency who can manage all others.
		- It is crucial to have a search feature but as far as this is not available there is need to explain data providers/agencies differences in EDHS and archival classification.
		- Security issues need to be solved for real production implementation (public network, first login)

Pilots 4 - Business Archives





Pilot 4	Bus	iness	Arc	hives	;																			
Task leader	Nati	ional	Arch	nives	of Es	toni	a																	
Supported by	Esto	nian	Busi	ness	Arch	ives																		
Scope	Pre-	inge	st pr	epara	ation	and	trans	sfer o	of bu	sines	s rec	ords	to a	digit	al ar	chive	e solu	ition	in a	busir	ness a	archi	ve	
Object	bes	ooke	busi	ness	syste	em tł	nat co	ontai	ns da	ataba	ise re	ecord	ls											
Short description	com reco	ipany ords.	/ is c The	ompi busir	rised ness a	of pi archi	ivate ves p	e bus ilot i	iness n the	ses ir e E-A	i Esto RK pi	onia f rojec	for an t will	rchiv I focu	ing a	nd p tran	reser sfer	vatic of da	on of		рар	er ar	of the Id dig m a	
Contacts	Nan	ne (T	itle)							E-m	ail								Sky	be				
Contact Person	Raiv	vo Ri	uusa	lepp)					raiv	0@6	eba.e	e						raiv	oruu	ı			
Pilot staff member	Ats	Ran	d							ats.	ranc	@el	ba.e	<u>e</u>					atsr	and				
OAIS Relevance		I	Pre-l	nges	t			Ing	est -	Stor	age						Sto	rage	– Ac	cess				
E-ARK Formats			E	-AR	(SIP				E	-ARK	AIP										E	-ARK	DIP	
			S	IARD	2.0	Х			SMU	RF EF	RMS				SMU	JRF S	SFSB					Geo	data	
E-ARK Tools	× Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	× Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
Scenario 1	~	ratio	n an	l d Ing	est o	f bus	ines	s rec	ords	from	besi	ooke	busi	ness	syste	em (I	Data	set 1)	^	I	I		
	_				fron										-,				7					
Scenario 2			-					-		-	hoc	aaka	huci	nocc	syste		Jata	cot 2	<u>\</u>					
Scenario 3	_			-							nes	JUKE	busi	11622	syste	:::: (L	Jaid	set Z	1					
Scenario 4	Extr	actin	ig red	cords	fron	n dat	abas	e (Da	ata se	et 2)														

Scenario 1	Mig	ratio	n an	d Ing	est o	of bu	sines	s rec	ords	from	n bes	poke	e bus	iness	syst	em								
Description	Expo	ort b	usine	ess re	cord	s fro	m be	spok	e bu	sines	s sys	tem.	Inge	st to	local	arch	ival s	syste	m of	EBA.				
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd Ir	ngest	relat	iona	l dat	abas	e bas	ed oi	n SIA	RD 2	.0											
E-ARK specifications	E-AF	rk si	P, SIA	ARD 2	.0																			
E-ARK Tools	Data	abase	e Pre	serva	tion	Tool	kit																	
Data	Rec	ords	from	besp	oke	busiı	ness	syste	em															
Description	Busi	iness	syst	em w	ith 1	4 tak	oles.	The o	datak	ase o	conta	ains a	appro	oxima	tely	12 00	00 re	cord	s.					
Data type	MS-	SQL	as m	df																				
Metadata format	non	e																						
Quantity	mor	e tha	an 12	000	rows	5																		
OAIS Relevance			Pre-l	ngest	:			Ing	est -	Stora	age						Stor	rage	– Ace	cess				
E-ARK Format			E	-ARK	SIP				E	-ARK	AIP										E	-ARK	DIP	
specifications			S	IARD	2.0	х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	lata	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	Х																							

Extra	actin	ig re	cords	fror	n da	tabas	se																
Extra	actin	g rec	cords	from	n dat	abas	e cor	ntaini	ing n	o doo	cume	nts.											
Acce	ess (r	not D	IPs ir	volv	ed o	nly re	estor	ing d	ata f	rom	SIAR) pao	kage	s)									
Acce	ess d	ataba	ases v	via D	BVTł	(SQ	L)																
SIAR	D 2.	0																					
Data	abase	e Pre	serva	tion	Tool	kit																	
Reco	ords	from	besp	oke	busi	ness	syste	em															
Busi	ness	syst	em w	ith 1	4 tal	oles.	The o	datak	base	conta	ins a	ppro	oxima	tely	12 00)0 re	cord	s.					
MS-S	SQL	as m	df																				
none	е																						
mor	e tha	an 12	000	rows	5																		
		Pre-l	ngest				Ing	gest -	Stora	age						Sto	rage	- Acc	cess				
		E	-ARK	SIP				E	-ARK	AIP										E	-ARK	DIP	
		S	IARD	2.0	х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	data	
Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	× Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	Extra Acce SIAR Data Recc Busi MS- none mor	Extractin Access (r Access d SIARD 2.0 Database Records Business MS-SQL a none more tha	Extracting rec Access (not D Access datab SIARD 2.0 Database Pre Records from Business syst MS-SQL as m none more than 12 Pre-I E S	Extracting records Access (not DIPs in Access databases SIARD 2.0 Database Preserva Records from besp Business system w MS-SQL as mdf none more than 12 000 Pre-Ingest E-ARK SIARD	Extracting records from Access (not DIPs involv Access databases via D SIARD 2.0 Database Preservation Records from bespoke Business system with 1 MS-SQL as mdf none more than 12 000 rows Pre-Ingest E-ARK SIP SIARD 2.0	Extracting records from dat Access (not DIPs involved of Access databases via DBVTH SIARD 2.0 Database Preservation Tool Records from bespoke busin Business system with 14 tab MS-SQL as mdf none more than 12 000 rows Pre-Ingest E-ARK SIP SIARD 2.0 X type	Extracting records from databas Access (not DIPs involved only re Access databases via DBVTK (SQ SIARD 2.0 Database Preservation Toolkit Records from bespoke business Business system with 14 tables. MS-SQL as mdf none more than 12 000 rows Pre-Ingest E-ARK SIP SIARD 2.0 X involution (Signal Content of Co	Access (not DIPs involved only restor Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business syste Business system with 14 tables. The of MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ing SIARD 2.0 Ver-Ingest Ing SIARD 2.0 Ver-Ingest Ing (qam kuching Ver-Ingest Ing Ver-Ingest Ing SIARD 2.0 Ver-Ingest Ing (qam kuching Ver-Ingest Ing Ver-Ingest Inge Ver-Ingest Inge Ver-Ingest Inge Ver-Ingest Inge Ver-Ingest Inge Ver-Ingest Inge Ver-Ingest Ingest Inge Ver-Ingest Inge	Extracting records from database contain Access (not DIPs involved only restoring d Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - E-ARK SIP SIARD 2.0 X SMU tig ool ungest - E-ARK SIP Lagent - SIARD 2.0 Access database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database MS-SQL as mdf none More than 12 000 rows SIARD 2.0 X SMU tig ool ungest - E-ARK SIP Access database SIARD 2.0 X SMU Access database Records from bespoke business system Business system with 14 tables. The database NS-SQL as mdf none More than 12 000 rows SIARD 2.0 X SMU Access database SIARD 2.0 SIARD	Extracting records from database containing in Access (not DIPs involved only restoring data fr Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database of MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Stora Pre-Ingest SIARD 2.0 X SMURF EF the database of the table of the table of tabl	Extracting records from database containing no doc Access (not DIPs involved only restoring data from S Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database containing MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP SIARD 2.0 X SMURF ERMS in the storage of the storage	Extracting records from database containing no docume Access (not DIPs involved only restoring data from SIARI Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains a MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Pre-Ingest SIARD 2.0 SIARD 2.0 Multiple and the second of the s	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD pac Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains appro MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Ingest - Sto	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD package Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approxima MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP SIARD 2.0 Vigon data from Database Contains approxima MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Ingest - Storage Ingest - Storage Ingest - Storage Ingest - Storage Ingest - Storage Ingest Ingest Ingest - Storage Ingest Ingest - Storage Ingest	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP SIARD 2.0 MS SIARD 2.0 MS SIARD 2.0 MS SQL as mdf none more than 12 000 rows SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 MS SQL as mdf none More than 12 000 rows SIARD 2.0 SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 SIARD 2.0 Y SIARD 2.0 SIARD 2.0 SIARD 2.0 Y SIARD 2.0 SIARD 2.0 SIARD 2.0 SIARD 2.0 MO SIARD 2.0 SIARD	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately 12 00 MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Pre-Ingest SIARD 2.0 X SMURF ERMS SMURF S SMURF S till a difficult of the second of th	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately 12 000 re MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Stor Pre-Ingest SIARD 2.0 X SMURF ERMS SMURF SFSB tild Gibard Contains approximately 12 000 restoration SIARD 2.0 MS-SQL as mdf none more than 12 000 rows SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 Access database contains approximately 12 000 restoration Tools for Archive (ETA) None Not tool for the storage Stor SIARD 2.0 N SMURF SFSB SMURF SFSB Tools for the storage Stor SMURF SFSB SMURF SF	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately 12 000 records MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Storage E-ARK SIP I E-ARK AIP SIARD 2.0 SIARD 2.0 MOT UNIT I I I I I I I I I I I I I I I I I I	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately 12 000 records. MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Storage - Acc SIARD 2.0 V (q) vivin big of vivi	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately 12 000 records. MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Storage Access SIARD 2.0 Pre-Ingest Ingest - Storage Storage - Access VI (QI) Interest Ingest - Storage Storage - Access VI (QI) Interest Ingest - Storage Storage - Access VI (QI) Interest Ingest - Storage Storage - Access Records from Jacua Interest Ingest - Storage Storage - Access VI (QI) Interest Ingest - Storage Ingest Ingest - Storage Ingest Ingest - Storage Ingest	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately 12 000 records. MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage Storage - Access E-ARK SIP I Ingest - Storage Storage - Access SIARD 2.0 Pre-Ingest Ingest - Storage Storage - Access E-ARK SIP I Ingest - Storage Storage - Access SIARD 2.0 SIARD 2.0 Business system with 14 tables. The database contains approximately 12 000 records. MS-SQL as mdf none more than 12 000 rows E-ARK SIP I Ingest - Storage Storage - Access SIARD 2.0 SIARD	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately 12 000 records. MS-SQL as mdf none more than 12 000 rows Pre-Ingest E-ARK SIP SIARD 2.0 X SIARD 2.0 Pre-Ingest E-ARK SIP V SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 Access from business system Business system with 14 tables. The database contains approximately 12 000 records. MS-SQL as mdf none Pre-Ingest SIARD 2.0 Y SIARD 2.0 SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 Y SIARD 2.0 SIARD 2.0 SIARD 2.0 Y SIARD 2.0 SIARD 2.0 SIA	Extracting records from database containing no documents. Access (not DIPs involved only restoring data from SIARD packages) Access databases via DBVTK (SQL) SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database contains approximately 12 000 records. MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ingest - Storage E-ARK SIP SIARD 2.0 X SMURF ERMS SMURF ERMS SMURF ERMS SMURF ERMS SMURF SFSB Access and an

Scenario 3	Mig	ratio	n an	d Ing	est o	of bus	sines	s rec	ords	fron	ı bes	poke	e bus	iness	syst	em								
Description	Expo	ort b	usine	ess re	cord	s froi	m be	spok	e bu	sines	s syst	tem.	Inge	st to	local	arch	ival s	syste	m of	EBA.				
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd Ir	ngest	relat	iona	l data	abas	e bas	ed o	n SIA	RD 2	.0											
E-ARK specifications	E-AF	rk si	P, SIA	ARD 2	2.0																			
E-ARK Tools	Data	abase	e Pre	serva	ation	Tool	kit																	
Data	Reco	ords	from	l besp	ooke	busir	ness	syste	em															
Description				em w he w									•								com	olete		
Data type	MS-	SQL	as m	df																				
Metadata format	non	e																						
Quantity	mor	e tha	an 20	00 00	0 rov	٧S																		
OAIS Relevance			Pre-l	ngest	:			Ing	sest -	Stora	age						Stor	rage ·	– Aco	cess				
E-ARK Format			E	E-ARK	SIP				E	-ARK	AIP										E	-ARK	DIP	
specifications			S	SIARD	2.0	Х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	lata	
E-ARK Tools	× Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	X																							L

Scenario 4	Extracting records from dat	abase					
Description	Extracting records from dat	abase containing no doo	cume	nts.			
OIAS relevance	Access (not DIPs involved or	nly restoring data from S	SIAR	D packages)			
Use-case	Access databases via DBVTk	((SQL)					
E-ARK specifications	SIARD 2.0						
E-ARK Tools	Database Preservation Tool	kit					
Data	Records from bespoke busin	ness system					
Description	Business system with 63 tak	oles (+several history an	d sup	port tables that are not	t nee	ded for a complete	
	structure of the working dat	tabase). The database c	ontai	ns approximately 200 0	00 re	ecords.	
Data type	MS-SQL as mdf						
Metadata format	none						
Quantity	more than 200 000 rows						
OAIS Relevance	Pre-Ingest	Ingest - Storage		Stor	rage	– Access	
E-ARK Format	E-ARK SIP	E-ARK AIP				E-ARK DIP	Х
specifications	SIARD 2.0	SMURF ERMS	Х	SMURF SFSB		Geodata	

|--|

Please note that more details with screenshots on scenario execution are provided in the deliverable <u>D2.4 Pilot</u> <u>Documentation</u>.

Execution report

The Estonian Business Archives (EBA) wanted to perform only one pre-ingest scenario in a test environment according to plans in D2.3 Detailed Pilot Requirements but as they worked with the tool, wished to substantially extend their work. EBA had good experience with the Database Preservation Toolkit SIARD 2.0 and also wanted to try the Database Visualization Toolkit. Finally EBA have performed 4 scenarios in "business-as-usual" manner, ingesting the SIARD files into their local preservation repository and accessing them through DBVTK.

Scenario	Started	Completed	Summary
1. Migration and Ingest of business records	April	September	Scenario performed successfully. Tools worked as
from bespoke business system (Data set 1)	2016	2016	required.
2. Extracting records from database	August	September	Scenario performed successfully. Tools worked as
(Data set 1)	2016	2016	required.
3. Migration and Ingest of business records	September	October	Scenario performed successfully. Tools worked as
from bespoke business system (Data set 2)	2016	2016	required.
4. Extracting records from database	September	October	Scenario performed successfully. Tools worked as
(Data set 2)	2016	2016	required.

Changes to the original plans

There were no changes. The scenarios have been performed according to plans in DoW and D2.3 Detailed Pilot Requirements.

Feedback report

The following table summarizes the feedback communication between the pilot staff and tool developers or format specification providers.

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
Database Preservation Toolkit	For the complete issue history, please refer to the GitHub page:
(version2.0.0-beta4.2)	https://github.com/keeps/db-preservation-toolkit
Used in tasks	Data extraction – in scenario 1 and 3
Data (input / output)	Input: Business system with 14 tables. The database contains approximately 12 000
	records + Business system with 63 tables with approximately 200 000 records
	Output: SIARD2.0 packages.
Performance	Very good
Issues	There have been several issues with DBPTK related SIARD 2.0 output. KEEP Systems has
	corrected all the bugs and the response time was excellent. After the completion of the
	scenarios no known issues remained.
Wishes	None
Comments	None
Experiences and recommended	After correcting the early bugs the tool functioned properly.
practices	

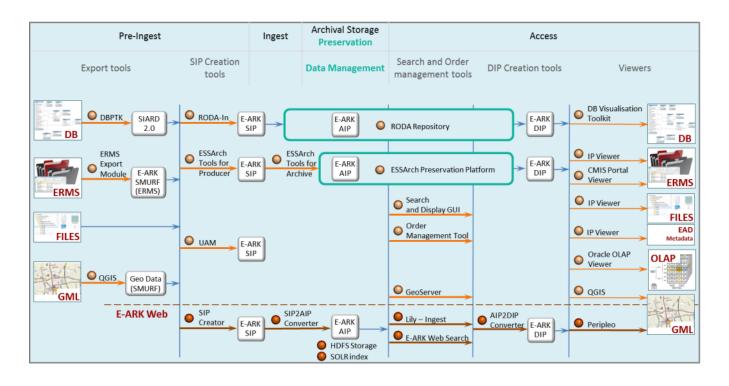
E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
Database Visualization Toolkit	
Used in task	Access – in scenario 2 and 4
Data (input / output)	Input: SIARD 2.0 packages
	Output: Restored DB tables
Performance	Good
Issues	No issues found
Wishes	None
Comments	None
Experiences and recommended	None
practices	

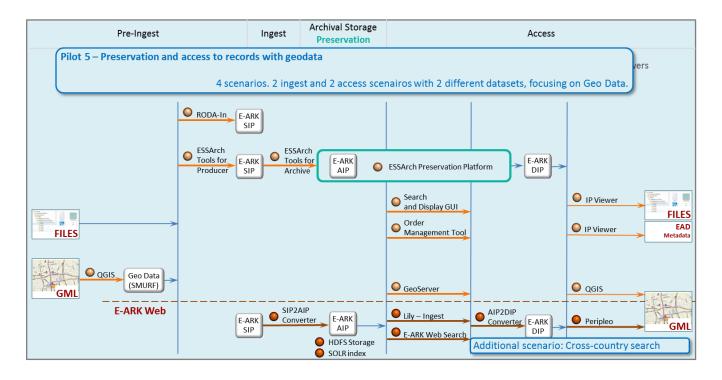
Recommended practices and further recommendations

The following table contains the recommended practices and further development suggestions collected during pilot execution and evaluation.

Category	Relates to	Recommended practices / Further developments
Recommended practices	SIARD 2.0	Manual validation requires a lot of time without SIARD 2.0 validation tools.

Pilots 5 - Preservation and access to records with geodata





Pilot 5	Pres	serva	tion	and	acce	ss to	reco	ords v	with	geod	lata													
Task leader	Nati	onal	Arch	nives	of Sl	oven	ia																	
Supported by	Dan	ish N	atior	nal A	rchiv	es																		
Scope	Pilot	t will	prov	/e th	at the	e SIP	and	DIP i	mple	emen	tatio	ns fu	lfill s	peci	fic re	quire	emer	nts fo	or the	e reco	ords	cont	aining	3
																-	-				-		prove	
											a ope	en da	ta m	ethc	od, di	rect	acces	ss in	the a	rchiv	es al	nd u	se GIS	5
	_	as s																						
Object										-													ippor	
		proje							noni	nent	01 56	ecte	u L-	ANK	arcin	vart	0015	prov	luen		e exe	amp		J v v
Short description									dize	d me	thod	for i	nges	ting	geo (data	will b	e de	velo	ped.	This	will a	allow	the
		-											-	-	-									
	of th	ne art	t too	ıls.																				
Contacts	Nam	ne (Ti	itle)							E-m	ail								Skyp	be				
Contact Person	Gre	gor Z	Zavr	šnik	()					gre	gor.z	avrs	nik@	gov	/.si				gre	gor.z	avrs	nik		
Pilot staff member	Aler	nka S	Starr	man	()					aler	nka.s	tarn	nan(@go	v.si									
Pilot staff member	Anja	a Pau	ulič (()						Anja	a.Pa	ulic@)go	/.si										
Pilot staff member	Joze	Anja Paulič () Anja.Paulic@gov.si Joze Skofljanec () joze.skofljanec@gov.si Pre-Ingest Ingest - Storage Storage - Access																						
OAIS Relevance		Pre-Ingest Ingest - Storage Storage – Access																						
E-ARK Formats		Pre-Ingest Ingest - Storage Storage - Access E-ARK SIP X E-ARK AIP X E-ARK DI														(DIP	Х							
																data	Х							
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	AITZUIP (E-AKK Web) Database Visualization Toolkit IP Viewer Peripleo Oracle (OLAP Viewer) Control (victor)				
			Х	Х			Х	Х				Х	Х	х	Х	Х	Х	Х	Х		Х	Х		
Scenario 1	SIP (Creat	ion a	and I	ngest	t of r	ecor	ds wi	ith G	eoda	ta (D	ata s	et 1-	2)										
Scenario 2		rch ai	nd A	ccess	s info	rmat	ion ι	using	Geo	data	(Dat	a set	1-2)											
	Sear							-			-													
Scenario 2	Sear SIP (rch ai	ion a	and I	ngest	t of r	ecor	ds wi	ith Ge	eoda	ta (D	ata s	et 3)											

Republi SIP creations than 10 geodata EARK SI ARK gui package Pre-Ing Other (E-ARK SI RODA-I Two set 1.) Rec	c of Sl tion a 00 red a SIP c P spee deline e was est, In SIP Cro IP, E-/ n, ESS is from	oveni ind in cords. reatic cificat es for perfo gest eatior ARK A	ia). Igest . Arc on. P tions geoc	of at hivis rodu for g data	t leas t crea icer c geoda SIP c	at one ates a create ata. <i>A</i> reation	e sma a Sub es a S Archi	all ve miss SIP cc vist t	ctor (ion a ontair	geoda greei ning g	ata so ment geoda	et wi for S ata, a	th les SIP cr accor	ss tha reation rding	an 10 on, ao to Su	00 ree	cords ding t	s and to E-A n agre	l one ARK g eeme	with guide	mor	for
SIP creat than 10 geodata EARK SI ARK gui packaga Pre-Ing Other (E-ARK S RODA-I Two sei 1.) Rec	tion a 00 rec a SIP c P spec deline e was est, In 6IP Cro IP, E-/ n, ESS is from	ind in cords. reation cificat es for perfo gest eation ARK A	igest . Arc on. P tions geoc orme	hivis rodu for រូ data	t crea icer c geoda SIP c	ates a create ata. A reatie	a Sub es a S Archi ⁱ	omiss SIP co vist t	ion a ontaiı	greei ning (ment geod	t for S ata, a	SIP cr accor	reatio rding	on, a to Su	ccord	ding t	to E-A n agre	ARK g eeme	guide	lines	for
than 10 geodata EARK SI ARK gui packago Pre-Ing Other (E-ARK S RODA-I Two sei 1.) Rec	00 red a SIP c P spee deline e was est, In 6IP Cro IP, E-/ n, ESS s fron	cords. reatic cificat es for perfo gest eatior	. Arc on. P tions geoc orme	hivis rodu for រូ data	t crea icer c geoda SIP c	ates a create ata. A reatie	a Sub es a S Archi ⁱ	omiss SIP co vist t	ion a ontaiı	greei ning (ment geod	t for S ata, a	SIP cr accor	reatio rding	on, a to Su	ccord	ding t	to E-A n agre	ARK g eeme	guide	lines	for
geodata EARK SI ARK gui package Pre-Ing Other (E-ARK S RODA-I Two sei 1.) Rec	a SIP c P spee deline e was est, In SIP Cro IP, E-/ n, ESS rs fron	reation cificat es for perfo gest eatior ARK A	on. P tions geoc orme	rodu for g data	icer c geodi SIP c	reato ata. <i>A</i> reatio	es a S Archi	SIP co vist t	ontaiı	- ning Į	geod	ata, a	accor	rding	to Su		-	n agre	eeme	-		
EARK SI ARK gui package Pre-Ing Other (E-ARK S RODA-I Two see 1.) Rec	P spee deline e was est, In 6IP Cro IP, E-/ n, ESS	cificat es for perfo gest eatior ARK A	tions geoo orme	for g data	geoda SIP c	ata. A reati	Archi	vist t			-			-		ubmi	ssior	-		ent, b	ased	on
ARK gui packagu Pre-Ing Other (E-ARK S RODA-I Two sei 1.) Rec	deline e was est, In SIP Cro IP, E-7 n, ESS es fron	es for perfo gest eatior ARK A	geoo orme	data	SIP c	reati			echn	icallv	valio											511
packag Pre-Ing Other (E-ARK S RODA-I Two se 1.) Rec	e was est, In SIP Cro IP, E-/ n, ESS	perfo gest eatior ARK A	orme				on. A	rchi		,	vant	lates	the	subm	hitteo	d SIP	pack	age,	ассо	rding	; to E	-
Pre-Ing Other (E-ARK S RODA-I Two se 1.) Rec	est, In SIP Cro IP, E-/ n, ESS s fron	gest eatior ARK A		d. Ar	n AIP				ist co	onfirr	ns, tl	nat c	onter	nt va	lidati	ion o	f the	subr	nitte	d SIP		
Other (E-ARK S RODA-I Two se 1.) Rec	SIP Cro IP, E-/ n, ESS	eatior ARK A	n and			is ge	nerat	ted fi	rom t	he Sl	P an	d get	s ing	ester	d into	o the	arch	ival r	epos	itory		
E-ARK S RODA-I Two set 1.) Rec	IP, E-/ n, ESS s fron	ARK A	n and																			
RODA-I Two se 1.) Rec	n, ESS s fron			d Ing	est o	f reco	ords	with	Geod	lata)												
Two set 1.) Rec	s fron	. بامد	AIP (v	vith (GeoD	ata)																
1.) Rec		Arch	Tool	s Arc	hive	(ETA), SIP	2AIP	(E-A	RK V	Veb),	ESS/	Arch	Prese	ervat	ion P	latfc	orm, E	EAD E	Edito	r, QG	ilS
		n the	Surv	veyin	g and	d Maj	oping	g Aut	horit	y of t	he R	epub	lic of	f Slov	/enia	:						
	ords a	nd me	etad	ata c	of mu	nicip	alitie	s as	valid	until	1994	4, exp	porte	d frc	om G	URS,	data	base	:			
2.) Rec	ords a	nd me	etad	ata c	of adr	ninis	trativ	/e un	nits ui	ntil 1	994,	ехро	rted	from	۱ GUF	۲S						
Record																						
GML do	document with metadata in XML format, ESRI Shapefile, csv																					
ISO 191	15 (IN	ISPIRE	E)																			
62 reco	rds (c	ca. 3N	ИB) +	+ 120)4 reo	cords	(cca	. 12,4	4 MB)												
	Pre-l	ngest				Ing	est -	Stora	age						Stor	rage	– Aco	cess				
	E	-ARK	SIP	Х			E	-ARK	AIP	Х									E	-ARK	DIP	Х
	S	IARD	2.0				SMU	RF EF	RMS				SML	JRF S	FSB	Х				Geoc	lata	х
			-		3				_						-							
olkit		(ЕТ	e		(ETA			orm										olkit				
10 L		Icer	Inpo	6	ive			latf			_	ō						Tot				
tion le		npo,	ğMc	Weł	Arch	(qa		on F			GU	it To					(qa	tion			er)	L
Iodu	D 19115 (INSPIRE) records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage E-ARK SIP X E-ARK AIP SIARD 2.0 SMURF ERMS SMURF SFSB X														iev	ewe						
rese rt N		olfc	vrchi	(E-7	ols 1	-AR	osito	eser	ge		Dis	agei				Sea	ARF	'isua		P <	CMIS portal/viewer	
se P Xpo	ata SIP creation. Producer creates a SIP containing geodata, according to Submission agreement, based SIP specifications for geodata. Archivist technically validates the submitted SIP package, according to E guidelines for geodata SIP creation. Archivist confirms, that content validation of the submitted SIP age was performed. An AIP is generated from the SIP and gets ingested into the archival repository. Ingest, Ingest r (SIP Creation and Ingest of records with Geodata) K SIP, E-ARK AIP (with GeoData) A-In, ESSArch Tools Archive (ETA), SIP2AIP (E-ARK Web), ESSArch Preservation Platform, EAD Editor, QC sets from the Surveying and Mapping Authority of the Republic of Slovenia: ecords and metadata of administrative units until 1994, exported from GURS, database ecords and metadata of maps with Geodata document with metadata in XML format, ESRI Shapefile, csv 9115 (INSPIRE) cords (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest E-ARK SIP SIARD 2.0 SMURF ERMS SIARD 2.0 SMURF ERMS SIARD 2.0 SIARD 2.0 SMURF ERMS SIARD 2.0 SMURF ERMS SIARD 2.0 SIARD 2.0 SMURF ERMS SIARD 2.0 SMURF ERMS SMURF SFSB X Geodata (1) (1) (1) (1) (1) (1) (1) (1)														orta							
aba VIS E	DA-I	Arcl	vers	cre	Arcl	ZAIF	DAF	Arcl	FS-S	.R Ir	rch	ler P	<u>ب</u>	ser	S	RK	2DII	aba	/iew	iple	<u>cle</u>	IS p
	ō	ESS	Uni	SIP	ESS	SIP	RO	ESS	P	ō	ê	2	<u> </u>	e e	U,	4	E E	Dat	4	Per	Ľ.	5
Dai					Х	Х		x	I	Ň	S	0		U	o X	Ш	4			_	0	5
	GML do ISO 191 62 reco	GML docume ISO 19115 (IN 62 records (co Pre-li E	GML document with ISO 19115 (INSPIR 62 records (cca. 31 Pre-Ingest E-ARK SIARD VILLE SIARD	GML document with m ISO 19115 (INSPIRE) 62 records (cca. 3MB) - Pre-Ingest E-ARK SIP SIARD 2.0 SIARD 2.0 Hong net E-Systech 100 for Producer (ETP) Nool for Producer (ETP) SIARD 2.0 Content (ETP) Nool for Producer (ETP) SIARD 2.0 SIARD 2.0	GML document with metad ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 120 Pre-Ingest E-ARK SIP X SIARD 2.0 (1 Yuk web) title and the second sec	GML document with metadata GAL document with metadata ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 12U records Pre-Ingest X SIARD 2.0 Image: Colspan="2">Vertical Acchining Module Image: Colspan="2">Image: Colspan="2">Image: Colspan="2">Image: Colspan="2">Image: Colspan="2">Image: Colspan="2">Image: Colspan="2" Telescolspan="2" Telescolspan="2	GML document with metadata in XM ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records Pre-Ingest Ing E-ARK SIP X SIARD 2.0	GML document with metadata in XML for ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca Pre-Ingest Ingest - E-ARK SIP X E SIARD 2.0 SMU	GML document with metadata in XML format, ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12, Pre-Ingest Ingest - Stor E-ARK SIP X E-ARK SIARD 2.0 SMURF E	GML document with metadata in XML format, ESRI ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB Pre-Ingest Ingest - Storage E-ARK SIP X E-ARK AIP SIARD 2.0 SMURF ERMS	GML document with metadata in XML format, ESRI Shap ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage E-ARK SIP X SIARD 2.0 SMURF ERMS	GML document with metadata in XML format, ESRI Shapefile, ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage E-ARK SIP X SIARD 2.0 SMURF ERMS	GML document with metadata in XML format, ESRI Shapefile, csv ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage E-ARK SIP X SIARD 2.0 SMURF ERMS	GML document with metadata in XML format, ESRI Shapefile, csv ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage E-ARK SIP X SIARD 2.0 SMURF ERMS	GML document with metadata in XML format, ESRI Shapefile, csv ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage E-ARK SIP X SIARD 2.0 SMURF ERMS	GML document with metadata in XML format, ESRI Shapefile, csv Ingest - Storage Stor 19115 (INSPIRE) Pre-Ingest Ingest - Storage Storat Module Pre-Ingest Storage SMURF ERMS SMURF ERMS SMURF ERMS SMURF SFSB Ingest - Storage SMURF SMURF ERMS SMURF SMURF ERMS Ingest - Storage SMURF SFSB Ingest - Grave and tool for hour (E-ARK MeP) N Ingest of the second of the	GML document with metadata in XML format, ESRI Shapefile, csv ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage SMURF SIP X SMURF ERMS IP X SMURF ERMS X SMURF ERMS X Ingest - Storage SMURF SFSB X SMURF Ingest (E-ARK MP) X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X	GML document with metadata in XML format, ESRI Shapefile, csv ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage Storage - Acceleration V Pre-Ingest Ingest - Storage Storage Storage SIARD 2.0 SMURF ERMS SMURF SFSB X vigest Ingest - Storage SMURF SIARD 2.0 SMURF second colspan="6">SMURF SFSB X V Management Lool Ingest - Grade Ingest - Grade SMURF SIARD 2.0 SMURF second colspan="6">SMURF SIARD 2.0 V Management Lool Ingest - Grade Ingest -	GML document with metadata in XML format, ESRI Shapefile, csv ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage E-ARK SIP X SIARD 2.0 SMURF ERMS SMURF SFSB X	GML document with metadata in XML format, ESRI shapefile, csv Ingest - Storage Storage - Access Storage - Access Pre-Ingest Ingest - Storage Storage - Access Storage - Storage - Access Storage - Access SMURF SFSB X SMURF ERMS X SMURF SFSB X Access SMURF Second Ingest - Storage SMURF Second A A A Ingest - Storage Storage A A A A A Ingest - Mangolipha G Ingest - Storage <td>GML document with metadata in XML format, ESRI Shapefile, csv ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage Storage - Access Y Ingest - Storage Storage - Access SIARD 2.0 X SMURF ERMS X Ingest - Storage SMURF SIAR (E-ARK (PR) X Geoco Vigest - Jack (ELA) X SMURF SFSB X Geoco Numagement Lool Numagement Lool N ageocion None X SMURF SFSB X Geocion None N Augeoritation Lool Augeoritation Lool Augeoritation Lool None Augeoritation Lool Augeoritation Augeoritation </td>	GML document with metadata in XML format, ESRI Shapefile, csv ISO 19115 (INSPIRE) 62 records (cca. 3MB) + 1204 records (cca. 12,4 MB) Pre-Ingest Ingest - Storage Storage - Access Y Ingest - Storage Storage - Access SIARD 2.0 X SMURF ERMS X Ingest - Storage SMURF SIAR (E-ARK (PR) X Geoco Vigest - Jack (ELA) X SMURF SFSB X Geoco Numagement Lool Numagement Lool N ageocion None X SMURF SFSB X Geocion None N Augeoritation Lool Augeoritation Lool Augeoritation Lool None Augeoritation Lool Augeoritation Augeoritation	

Scenario 2	Search and Access information using Geodata
Description	Create DIP from AIP containing record with Geodata. Present Geodata information with QGIS along with
	content and metadata from DIP.
	A data object containing geodata can be identified by using search criteria as specified by E-ARK Tool
	requirement specification after search index was updated from an AIP. Selected data objects are selected and
	order is issued. DIP is prepared according to order specification and end user credentials. DIP file structure with
	file descriptions (mime type, short description) is presented to the end user. Geodata from the order can be
	accessed in the designated viewer (QGIS). The user checks authenticity of the DIP by accessing PREMIS
	documentation. Access to DIP is documented and captured metadata can be exported.
OIAS relevance	Access
Use-case	Other (Access of records with Geodata)
E-ARK specifications	E-ARK AIP, E-ARK DIP (with GeoData)
E-ARK Tools	Search and Display GUI, Order Management Tool, Lily – Ingest, ESSArch Preservation Platform, E-ARK Web
	(Search), AIP2DIP (E-ARK Web), IP Viewer, QGIS, Geoserver, Peripleo
Data	Two sets from the Surveying and Mapping Authority of the Republic of Slovenia:

Description Data type Metadata format	2.) F Reco GMI ISO	Records Drds L doc 1911	rds a and i cume .5 (IN	and n nd m meta nt wi ISPIR	etad data th m E)	ata o of m etad	of adr iaps v ata ii	minis with n XM	Geod Geod	ve un lata mat,	its u ESRI	ntil 1 Shap	994,	expo					5, dat	abas	e			
Quantity OAIS Relevance	62 r		<u> </u>	ca. 31		+ 120)4 reo		•)					Sta		A	0.00				
E-ARK Format				ngest -ARK				Ing	est - F	-ARK	-	х					Sto	rage	- Acc	ess	F	-ARK	DIP	x
specifications				SIARD	-				SMU						SMU	JRF S	FSB	х				Geod		X
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	c Search and Display GUI	Conter Management Tool	k Lily - Ingest	c Geoserver	¢ QGIS	E-ARK Web Search	k AIP2DIP (E-ARK Web)	Database Visualization Toolkit	k IP Viewer	c Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
												Х	Х	х	Х	Х	х	Х	х		Х	Х		

Scenario 3	SIP	Crea	tion	and l	nges	t of r	ecor	ds w	ith G	eoda	ita																				
Description	Crea	ate S	IP fro	om re	cord	s and	d met	adat	a exp	oorte	d fro	m AF	rso (Envi	ronm	ienta	l Age	ency o	of Re	publi	ic of :	Slove	nia).								
	SIP o	creat	ion a	nd in	igest	of at	t leas	t one	e vec	tor g	eoda	ta wi	ith at	leas	t 250) reco	ords.	Data	is ex	kport	ed di	irectl	y fro	m							
	thei	r ow	n sys	tem i	nto (GML	form	at. A	nd tl	neir s	yster	n als	o exp	oorts	INSF	PIRE r	neta	data													
	Arch	nivist	crea	tes a	Sub	missi	on a	greei	ment	for S	SIP cr	eatio	on, ac	cord	ling t	o E-A	RK g	uide	lines	for g	eoda	ita Sl	Ρ								
	crea	tion	. Pro	ducer	crea	ates a	a SIP	cont	ainin	g geo	odata	, асс	ordi	ng to	Subr	nissi	on ag	greer	nent	, base	ed or	ו EAR	K SIP)							
	spec	cifica	tions	forg	geod	ata. A	Archi	vist t	echn	ically	/ valio	dates	s the	subn	nitte	d SIP	pack	age,	ассс	ording	g to E	-ARk									
	-			-																		o pac	kage								
	was	perf	orme	ed. Ar	n AIP	is ge	enera	ted	from	the S	SIP ar	nd ge	ts in	geste	ed int	o the	e arcl	nival	repo	sitor	у.										
OIAS relevance	Pre-	Inge	st, In	gest																											
Use-case	Oth	er (S	IP Cre	eatio	n and	d Ing	est o	f rec	ords	with	Geod	lata)																			
E-ARK specifications	E-AF	rk Si	P, E-/	ARK A	AIP (v	vith (GeoD	ata)																							
E-ARK Tools	ESSA	Arch	Tool	s Pro	duce	r (ET	P), E	SSArc	ch To	ols A	rchiv	e (ET	ΓA), Ε	SSAr	ch Pr	reser	vatio	n Pla	tfor	n, EA	ND Ed	litor,	QGIS	;							
Data	Reco	ords	and ı	meta	data	of N	atura	200	0 are	eas cr	eate	d in 2	2004	, exp	ortec	d fror	n AR	SO d	ataba	ase											
Description	Reco	ords	and ı	meta	data	of m	aps v	with	Geod	lata																					
Data type	GM	L doc	ume	nt wi	th m	etad	ata ii	n XM	L for	mat,	ESRI	Shap	pefile																		
Metadata format	INSF	PIRE																			ess E-ARK D										
Quantity	286	reco	rds (cca. S	9,6 N	1B)													age – Access E-ARK DI Geodat												
OAIS Relevance			Pre-l	ngest				Ing	est -	Stora	age						Sto	rage	– Ac	cess											
E-ARK Format			E	-ARK	SIP				E	-ARK	AIP										E	-ARK	DIP								
specifications			S	IARD	2.0	Х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	data								
E-ARK Tools				P)			7			_																					
	olkit			ET (ET	e		(ET/			forn										olkił				1							
	D To			ncer	npo	(q	ive			Plat			=	lo						٦To											
	atior	ule		rodi	ВM	We	Arch	(də)		ion			v GU	nt To				_	eb)	ation			/er)	L.							
	erva	Mod		or P	ivin	ARK	for	(E-ARK Web)	٥Ŋ	rvat			play	imei				arch	××	aliz			viev	iewe							
	Pres	ort		ool f	Arch	r (E-	sloc	-AR	osit	rese	age	×	d Dis	Jage	بد			b Se	-AR	Visu			AP	al/s							
	ase	Expo	Ļ	hΤα	sal /	atoi	μĭ		Rep	ιΗ	stor	nde	anc	Mar	ges	rver		Wel	P (E	ase 1	ver	o,	[0	ortio							
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	atab	Viev	eriple	acle	CMIS portal/viewei							
	Da	ER	RC	X X	Ŀ	SII	X X	X X	RC	E X	H	X X	Se	ō	Ci)	Ö	ð	ш	A	Da	đ	Pe	ō	5							
				^			~	^		^		^											<u> </u>	i							

Scenario 4	Sea	rch a	nd A	ccess	s info	rmat	tion	using	g Gea	dota														
Description	Crea	ate D	IP fro	om A	IP co	ntair	ning r	ecor	d wit	h Ge	odat	a. Pre	esent	t Geo	odata	info	rmat	ion w	vith (QGIS	alon	g wit	h	
	con	tent	and r	neta	data	from	DIP.																	
	A da	ata o	bject	cont	ainin	ig ge	odata	a can	be i	denti	fied	oy us	ing s	earc	h crit	eria	as sp	ecifie	ed by	/ E-A	RK To	ol		
	requ	uiren	nent	speci	ficat	ion a	fter s	searc	h inc	lex w	as up	odate	ed fro	om a	n AIP	. Sele	ected	data	a obj	ects	are se	elect	ed an	d
	orde	er is i	ssue	d. DII	P is p	repa	red a	iccor	ding	to or	der s	peci	ficati	on a	nd er	nd us	er cre	eden	tials.	DIP	file s	truct	ure w	/ith
			•	-		•••					•												in be	
					-													-		essin	g PRE	MIS		
	doc	umei	ntatio	on. A	ccess	s to D	OIP is	docu	ımer	ted a	and c	aptu	red n	neta	data	can k	e ex	porte	ed.					
OIAS relevance	Acc																							
Use-case				s of re																				
E-ARK specifications	E-Al	rk ai	P, E-/	ARK [) DIP	with	GeoD	Data)																
E-ARK Tools	Sear	rch a	nd Di	isplay	/ GUI	, Orc	der N	lanag	geme	ent To	ool, L	ily –	Inges	st, ES	SArc	h Pre	serv	ation	l Plat	form	n, E-A	ARK V	Veb	
	(Sea	rch),	AIP2	2DIP	(E-AF	RK W	eb), l	P Vie	ewer	, QGI	S, Ge	oser	ver, F	Perip	leo									
Data	Rec	ords	and r	meta	data	of N	atura	200	0 are	eas cr	eate	d in 2	2004,	, exp	ortec	d fror	n AR	SO d	atab	ase				
Description	Rec	ords	and r	meta	data	of m	aps v	with	Geoc	lata														
Data type	GM	L document with metadata in XML format, ESRI Shapefile PIRE																						
Metadata format	INS	PIRE	RE																					
Quantity	286	reco	rds (cca. 9	9,6 N	1B)																		
OAIS Relevance			Pre-li	ngest	:			Ing	est -	Stora	age						Stor	rage	– Ac	cess				
E-ARK Format			E	-ARK	SIP				E	-ARK	AIP	Х									E	-ARK	DIP	Х
specifications			S	IARD	2.0			:	SMU	RF EF	RMS				SMU	JRF S	SFSB	х				Geo	data	х
E-ARK Tools	it			TP)			(A)			E										it				
	N N			er (E	ule		e (El			tfor										oolk				
	n To			luce	lodi	eb)	hive	_		ı Pla			n	loo					_	n To				
	'atio	dule		Proc	√ Bu	Ň	. Arc	(E-ARK Web)		tion			ay G	ent 1				ч	Veb)	atic			wer	/er
	serv	Mo		for	hivil	-ARI	s for	RK V	tory	erva			ispla	eme				earc	N XS	ualiz			Vie	view
	Pre	ort		00	Arc	or (E	loo	E-AI	posi	rese	rage	Xe	id Di	nag	st	-		s de	E-AF	Visı			LAP	tal/
	ase	Exp	-l	ch T	rsal	eatc	ch T		Re	сhР	Sto	Inde	h an	Μa	nge	erve		Ŵ) dio	ase	wer	eo	e 0	port
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	× Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
		Ш	R	Ш		S	ш	S	8	ш	T	x	x	X	X	X	X	X	≺ X		⊔ X	_ X	0	U
			l												L					1	1		<u>لـــــــــ</u>	

Additional scenario	Cross-country search with E	E-ARK Web (joint scenario v	vith NAH)										
Description	The SOLR index and E-ARK V	Neb infrastructure theoretic	cally makes it possible to perform a federated search										
	over more than one archive	. When the SOLR index of th	ne other archival institution can be "seen" by the search										
	engine (e.g. one institution	has access rights to the othe	ers SOLR) then it can make a common list of the result.										
	The National Archives of Slo	ovenia and the National Arch	nives of Hungary both have an E-ARK implementation at										
	their pilot sites. This scenari	io is a simple feasibility stud	y of cross-country search.										
OIAS relevance	Access												
Use-case	Search and Display												
E-ARK specifications													
E-ARK Tools	E-ARK Web												
Data	Test data in the SOLR index												
Description	The SOLR index of the two a	archives will be theoretically	connected in this sceanrio										
Data type	Not relevant												
Metadata format	Not relevant												
Quantity	small												
OAIS Relevance	Pre-Ingest	Ingest - Storage	Storage - Access										
E-ARK Format	E-ARK SIP	E-ARK AIP	E-ARK DIP										

specifications			S	IARD	2.0				SMU	RF EF	RMS				SMU	JRF S	SFSB					Geod	lata	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
												Х						Х						

Please note that more details with screenshots on scenario execution are provided in the deliverable <u>D2.4 Pilot</u> <u>Documentation</u>.

Execution report

Two pilots (5, 7) decided to use many tools also testing their compatibility beside their core functionality. The pilot of the Slovenian National Archives (NAS) was focusing on Geodata. NAS has tested the ESSArch tools and E-ARK Web tools with SMURF Geodata specification checking their compatibility with the E-ARK Geodata standard and with each other from SIP creation to accessing graphical Geodata information. E-ARK Web has two deployment options: full deployment and virtual environment. The virtual environment is a compact solution for electronic archiving therefore could be very useful for smaller archives. NAS used the virtual E-ARK Web deployment solution.

Scenario	Started	Completed	Summary
1. Migration and Ingest of business records	April	September	After a longer the incompatibility errors were corrected
from bespoke business system (Data set 1)	2016	2016	the scenario performed successfully. Tools basically worked as required.
2. Extracting records from database	July	October	Scenario could not be completed before the Search tool
(Data set 1)	2016	2016	was ready but after completion the scenario performed
			successfully. Tools worked as required.
3. Migration and Ingest of business records	April	October	After a longer the incompatibility errors were corrected
from bespoke business system (Data set 2)	2016	2016	the scenario performed successfully. Tools basically
			worked as required.
4. Extracting records from database	July	October	Scenario could not be completed before the Search tool
(Data set 2)	2016	2016	was ready but after completion the scenario performed
			successfully. Tools worked as required.

Additional scenarios	Started	Completed	Summary
Cross-country search with E-ARK Web	December	January	The scenario execution was stopped because of security
(joint scenario with NAH)	2016	2017	considerations by the archives. The cross-country search is technically feasible but from security point of view it is risky. In the future if the archives build the infrastructure
			to implement a publicly accessible E-ARK Web solution

	outside their firewall then it can be reached from the
	search engine of another archive with E-ARK Web.

Changes to the original plans

There were no major changes. The scenarios have been performed according to plans in DoW and D2.3 Detailed Pilot Requirements.

Feedback report

The following table summarizes the feedback communication between the pilot staff and tool developers or format specification providers.

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
ESS Arch tools	For the complete issue history, please refer to the GitHub page:
	https://github.com/ESSolutions/ESSArch_Tools_Producer
	https://github.com/ESSolutions/ESSArch_Tools_Archive
	https://github.com/ESSolutions/ESSArch EPP
Used in tasks	In all scenario
Data (input / output)	SIP creation and ingest with 3 different datasets
Performance	Good
Issues	There have been several issues at the beginning, mostly incompatibility problems
	between tools and between tools and the SIP specification. After the completion of the
	scenarios no known issues remained.
Wishes	None
Comments	None
Experiences and recommended	After correcting the early bugs the tool functioned properly.
practices	

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
RODA-In	For the complete issue history, please refer to the GitHub page:
(2.0.0 Alpha 7.4)	https://github.com/keeps/roda-in
Used in tasks	Create SIP - Create an E-ARK Sip Package
Data (input / output)	Input: Unstructured data
	Output: EARK SIP in a *.zip file
Performance	ОК
Issues	No issues left at the end of the pilot
Wishes	None
Comments	The tool is being translated to Slovenian language.
Experiences and best practices	None

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
E-ARK Web	For the complete issue history, please refer to the GitHub page:
(Virtual deployment)	https://github.com/eark-project/earkweb
Used in tasks	SIP to AIP conversion, Lilly ingest, SOLR search, AIP to DIP conversion
Data (input / output)	Input: 3 different data set

	Output: depending on component
Performance	ОК
Issues	No issues left at the end of the pilot
Wishes	None
Comments	None
Experiences and best practices	None

E-ARK Tool – Version	Issues (bugs, wishes, comments) Experiences / Recommended practices
Search & Display GUI	
Order Management Tool	
Used in tasks	Access
Data (input / output)	Input: E-ARK AIP
	Output: order
Performance	ОК
lssues	No issues left at the end of the pilot
Wishes	None
Comments	None
Experiences and best practices	None

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
IP Viewer	
Used in tasks	View DIP
Data (input / output)	Input: DIP
Performance	Good
Issues	None
Wishes	None
Comments	None
Experiences and best practices	None

Recommended practices and further recommendations

Lessons learned

We addressed a real need with our users.

When we started talking to our producers, who were cooperating as pilot sites, they welcomed our propositions. There is a real need for them to know how to archive all the spatial data, that has been accumulating for some years. The guidelines from this project gave them a way to finally structure geodata in a way it is suitable for the archives, as well as an input on how to adjust their current and future systems in order to automate this process.

Bridging the gap of limited network accesses

Since we used two different tools for packaging data it was shown, that a stand-alone tool, like Roda-In is more usable than a web based one (ESS ETP). We are working with different organisations with different types of network security policies, that often disable us from accessing the web based tool from within organisations network. It is also more practical to physically move large quantities of data on a portable disk drive as oposed to streaming it via network.

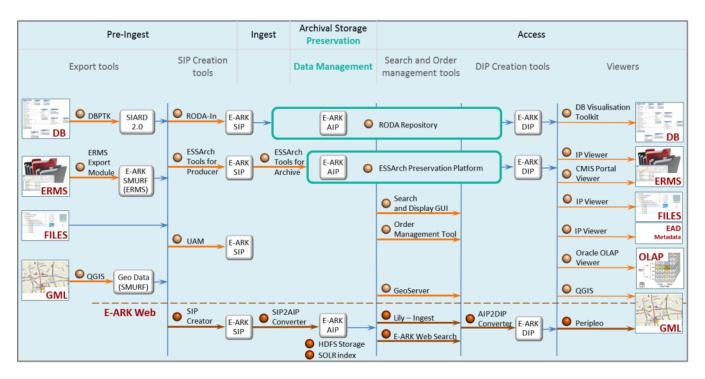
Full text search brings the archival experience closer to our users

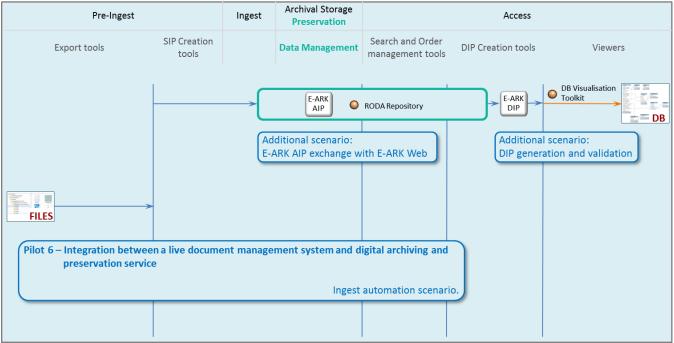
E-ARK Web based SOLR index with the Magenta Search interface brought us a new experience - full text search. Previously the only search option was using the catalogue. This brings our users an experience similar to the way of searching that they are used to already using (Google, Bing...). This provides better search results and less work for our archivists, but only if the data is well described. Therefore we need to assure, that we have good metadata descriptions.

Interoperability between systems – better communication between archives

Our experience using the general E-ARK IP structure through different applications has proven that using a common standard is a good way to ensure interoperability between different archives. This is important when using records that are the same across different archives within a country or even between countries across Europe (like the Natura 2000 record).

Pilots 6 - Integration between a live document management system and digital archiving and preservation service





Pilot 6	Integration between a live document management system and digital archiving and preservation service												
Task leader	KEEP SOLUTIONS (KEEP	PS)											
Supported by	Instituto Superior Técn	ico (IST)										
Scope	The goal of this pilot is two-fold. On one hand, KEEP SOLUTIONS will demonstrate that the pan-European SI												
	structure designed in th	structure designed in the WP3 is adequate to support the media types found in today's Elec											
	Management Systems	other hand, that the most											
	adequate and scalable	process to the preservation											
	service.												
Object	In order to achieve the goals of this pilot we will tap into two live Electronic Records Management Systems												
	(ERMS) and, based on the appraisal and selection strategies installed, extract, transform, aggregate and create Submission Information Packages (SIP) that conform to the A1:R21-European SIP format defined in WP3. The												
						vation services that follow							
			Information Packages from										
			_			served by the preservation							
Short description						. .	cations which defines how						
				order to mo	ove	e records between the thre	e stages of records keeping						
	- active, semi-active and inactive.												
	On a typical setting, a record that needs to be archived usually falls into one these three "ages": - Active - when the metadata and data are "live" being used and modified regularly.												
				-		d for a short period – say u							
						long-term repository for p							
						f information between the							
						ta is lost in the process. To							
							tions and orchestrates the						
	entire transfer process.			•									
	The pilot worked with o	data	from a public in	nstitution wh	ho	ose "active" records have b	een initially produced and						
	managed in an electror	nic re	ecords manager	ment system	n a	and then transferred to the	archival service of that						
	same institution for ter	npoi	rary conservatio	on - semi-act	tiv	ve stage.							
	The archival service is,	how	ever, not prepa	ared to face t	th	e challenges of long-term	digital preservation, so the						
	records that have been	sele	ected for perma	nent conser	rva	ation need to be transferre	d to a long-term digital						
	repository (the third "a	ge")	. This is where t	this pilot cor	me	es in.							
	The whole goal of the p	oilot	is to ensure tha	at the inform	na	tion package specifications	developed in E-ARK and						
	the integration procedu	ures	developed are	appropriate	e to	o support the transference	of records between a active						
	or semi-active archival	syste		-	vat	tion repository.							
Contacts	Name (Title)		E	E-mail			Skype						
Contact Person	Miguel Ferreira		r	mferreira@k	ke	ep.pt	jmaferreira						
Pilot staff member	Luís Faria		I	faria@keep.	.pt	<u>t</u>	luis100						
Pilot staff member	Hélder Silva hsilva@keep.pt hsilva_keep												
Pilot staff member	Sebastien Leroux sleroux@keep.pt slerouxatkeep												
Pilot staff member	Rui Rodrigues		<u>r</u>	eep.pt	rui.tiago.mr								
Pilot staff member	Ricardo Vieira		r	rjcv@tecnico	ricardojoao.vieira								
Pilot staff member	João Cardoso		j	oao.m.f.carc	do	oso@tecnico.ulisboa.pt	joao.m.f.cardoso						
OAIS Relevance	Pre-Ingest		Ingest - S	Storage		Storage	– Access						
E-ARK Formats	E-ARK SIP	Х	E-/	ARK AIP X			E-ARK DIP X						
	SIARD 2.0		SMUR	FERMS		SMURF SFSB X	Geodata						

E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	× RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
									X			-												
Scenario 1	Aut	oma	tic iı	nges	t of I	reco	rds f	rom	a se	emi-a	activ	e ar	chiva	al ma	anag	eme	ent s	yste	m					
Additional scenario	Inte	ntegration with OMT via E-ARK DIP																						
Additional scenario	Rep	osito	ry su	icces	sion	via E	-ARK	AIP (E-AR	RK AI	P exc	hang	ge ex	perir	nent	s)								

Scenario 1	Automatic ingest of records from a semi-active archival management system												
Description	This scenario aims to d	This scenario aims to demonstrate the ability to seamlessly transfer data from a semi-active records											
	management system to a long-term preservation repository with little or no human intervention.												
	The scenario is based on real-world operations already in place at a public organization since mid-2015.												
	scenario enhances the	scenario enhances the established practice by adding an additional component to its architecture that will b											
	responsible for the long	g-ter	m preservation of histor	rical ı	records once they reach t	thei	r inactive age. The long-	-					
	term preservation repo	sitor	ry runs as a back-end se	rvice	of the Archival Managen	nent	t System and aims to						
	support its data curation	support its data curation activities.											
OIAS relevance	Ingest												
Use-case	Other (Ingest of Archival Management Records using the SMURF profile.)												
E-ARK specifications	E-ARK SIP, E-ARK AIP												
E-ARK Tools	Repository Integration	Repository Integration Pipeline (RIP), RODA Repository											
Data	Historical records												
Description	Data used in this pilot s	scena	ario was comprised of a	colle	ction of digitised books r	elat	ed to the Peninsular Wa	ar					
	dating from 1778 to 18	34. T	he collection is compos	ed of	964 records stored in a	rela	tional database followir	ng					
	the semantic elements	of E	AD. The dataset also cor	ntains	s a total of 34.600 pages	of d	ocumentation in						
	uncompressed TIFF file	s at 3	300 dpi. The total amou	nt of	data is around 1.2 TB. Th	nis c	ollection can be inspect	ted					
	at its original location a	at htt	p://arquivo.cm-mafra.p	t/det	tails?id=173037.								
Data type	300 dpi uncompressed	TIFF	files										
Metadata format	EAD												
Quantity	964 records described	in EA	D containing a total of 3	84.60	0 pages of 300 dpi uncor	npre	essed TIFF files. The tota	al					
	amount of data is arou	nd 1.	.19 TB.										
OAIS Relevance	Pre-Ingest		Ingest - Storage		Stora	age ·	– Access						
E-ARK Format	E-ARK SIP	Х	E-ARK AIP	X E-ARK DI									
specifications	SIARD 2.0		SMURF ERMS		SMURF SFSB	Х	Geodata	х					

Additional scenario	Inte	grati	ion w	ith C	MT	via E	-ARK	DIP															Integration with OMT via E-ARK DIP An Archive uses a combination of the Order Management Tool (OMT) and E-ARK IP Viewer to provide access to													
Description	An A	Archi	ve us	es a	com	oinat	ion o	of the	e Ord	er M	anag	emer	nt To	ol (O	MT)	and	E-AR	K IP \	/iewe	er to	prov	ide a	ccess	s to												
	existing digital objects to its users. In order to articulate the RODA repository system with these tools, a new process has been developed for RODA that enables an archivist to create E-ARK compliant DIPs. These files can																																			
	proc	cess l	nas b	een o	devel	lope	d for	ROD	A tha	at ena	ables	an a	rchiv	vist to	o crea	ate E	-ARK	com	plian	t DIP	s. Th	ese f	iles c	:an												
	then be downloaded and added to the OMT workflows in order to be served to the end-user.																																			
	The workflow works by selecting an AIP and running a process that generates an E-ARK DIP. The resulting DIP																																			
	can be downloaded on the RODA user interface and then uploaded to the OMT to be delivered to the end-user. The DIP can also be consulted using the RODA's REST API, for example, to support a more advanced systems																																			
	integration approach.																																			
OIAS relevance	Access																																			
Use-case																																				
E-ARK specifications	E-ARK DIP																																			
E-ARK Tools	RODA Repository, Order Management Tool																																			
Data	Test data																																			
Description	Different kinds of letters and documents																																			
Data type	Not relevant																																			
Metadata format	Not relevant																																			
Quantity	small																																			
OAIS Relevance	Pre-Ingest Ingest - Storage Storage - Access															1																				
E-ARK Format			E	-ARK	SIP				E	-ARK	AIP										E	-ARK	DIP	Х												
specifications			S	IARD	2.0				SMU	RF EF	RMS				SMU	JRF S	SFSB					Geo	Geodata													
E-ARK Tools	it			(J			(A)			۶										Ŀ.																
	olki			r (E1	əlr		е (ЕТ			tfori										olki																
	n Tc			luce	lodı	eb)	hive	_		Pla			5	00						n To			_													
	atio	dule		Prod	Jβ	Ň	Arc	Veb)		tion			ν GI	int T				ء	/eb)	atio			ver)	er												
	serv	Moc		for I	hivir	-ARI	for	N XX	tory	ervai			spla	eme				earc	× ×	aliz			Viev	view												
	Pre	ort		00.	Arc	or (E	00	(E-ARK Web)	isoc	rese	rage	Xe	Ρ	nag	st	-		sb Se	E-AF	Visı			LAP	tal∕												
	ase	Exp	-In	ch T	rsal	eatc	chT		Rep	ch P	Sto	Inde	h an	Ra	nge	erve		Ň) elo	ase	wer	eo	e (O	port												
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewei												
		Ξ	æ	ш	D	SI	ш	SI	ž	ш	I	S	Š	o X		G	ď	ш	A		8	ď	0	Ū												
									^					~																						

Additional scenario	Repository succession via E-ARK AIP (E-ARK AIP exchange experiments)
Description	A repository system has reached the end of its expected lifetime. The head of the Archive has decided to move
	to a next-generation long-term digital repository system. This will unavoidably imply the migration of metadata
	records, millions of files, and terabytes of data from the legacy repository system to the newly adopted one.
	Because of the large scale of this operation, this procedure should entail careful planning, validation and
	support. However, to simplify the migration of data between the two systems, the head of the Archive opted
	for a repository system that is compliant with the E-ARK AIP specification. By doing so, the migration of data

v	-	,		was greatly simplified. Data and metadata does not need to be transformed, restructured or reshaped in any																			
	way. AIPs just need to be copied to the storage area of the new repository (or linked to) and the new repository																						
n	, needs t	-							Ū					•								•	
h	n orde	r to in	nplem	nent	the s	cena	rio, a	a sele	ctior	n of A	IPs v	vill be	e trai	nsfer	red f	rom	the R	ODA	repo	ositor	ry sys	stem	to
ť	he E-A	RK W	eb ref	feren	ice in	npler	nent	atior	. Pre	vious	s to t	he tr	ansfe	erenc	e, a j	proce	ess n	eeds	to be	e run	over	r the	
s	electe	d AIPs	that	will g	genei	rate a	a ma	nifes	t file	in th	e roc	ot of	the A	IP fo	lder	(met	s.xm	I). Af	ter re	eceiv	ing tl	ne Al	Ps,
E	E-ARK Web will re-index them thus merging them with the rest of its managed data.																						
OIAS relevance A	Archival Storage																						
Use-case																							
E-ARK specifications E	E-ARK AIP																						
E-ARK Tools R	RODA Repository, E-ARK Web																						
Data T	Test data																						
Description D	Different kinds of letters and documents																						
Data type N	Not relevant																						
Metadata format	Not relevant																						
Quantity s	small																						
OAIS Relevance	Pre-Ingest Ingest - Storage Storage - Access																						
E-ARK Format		I	E-ARK	(SIP				E	-ARK	AIP	Х									E	-ARK	DIP	
specifications	SIARD 2.0						SMURF ERMS						SMURF SFSB Geodata										
E-ARK Tools	4		(а.			A)			۶										t				
	Database Preservation Toolkit ERMS Export Module		ESSArch Tool for Producer (ETP)	ule		ESSArch Tools for Archive (ETA)			ESSArch Preservation Platform				_						Database Visualization Toolkit				
	й И		duce	Nod	(eb)	chive	-		ו Pla			IJ	Tool					(on T			(
	vatio dule		Pro	I gui	κw	r Ar	(E-ARK Web)	>	atior			ay G	ent '				ch	Neb	zatio			wer	ver
	E Mo		l for	chiv	E-AR	ls fo	RK \	itor	erva	e		lgsi	gem				ear	RK \	inali			v Vie	viev
	e Pro		Too	l Ar	tor (I	Too	(E-A	soda	Pres	orag	lex	nd D	ana	est	er		/eb S	(E-A	e Vis	5		OLAF	rtal/
	Database Preservatio ERMS Export Module	RODA-In	\rch	ersa	creat	Vrch	AIP	A Re	vrch	S-St	S Inc	ch a	R	Ing	serv	(0	ΝX	DIP	basi	IP Viewer	oleo	ile (C	od S
	Data	ROD	ESSA	Universal Archiving Module	SIP creator (E-ARK Web)	ESSA	SIP2AIP	RODA Repository	ESSA	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Data	IP Vi	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
			_	_	••	_	•••	x	_	x	X		-		-		_		_		_		

Please note that more details with screenshots on scenario execution are provided in the deliverable <u>D2.4 Pilot</u> <u>Documentation</u>.

Execution report

The aim of pilot 6 was to assess the efficacy of the E-ARK Information Package Specifications which defines how metadata and data should be packaged in order to move records between the three stages of records keeping - active, semi-active and inactive.

On a typical setting, a record that needs to be archived usually falls into one these three "ages":

- 1. Active when the metadata and data are "live" being used and modified regularly.
- 2. Semi-active when the metadata and data are archived for a short period say up to 5 years.
- 3. Inactive when the metadata and data are moved to a long-term repository for permanent conservation.

The pilot aims to do ensure the seamless transfer of information between the semi-active and the inactive stages in a way that ensures that no relevant data or metadata is lost in the process. To accomplish this goal, a special

integration tool was developed that implemented the package specifications and orchestrated the entire transfer process.

The pilot worked with data from a public institution whose "active" records have been initially produced and managed in an electronic records management system and then transferred to the archival service of that same institution for temporary conservation - semi-active stage. The archival service is, however, not prepared to face the challenges of long-term digital preservation, so the records that have been selected for permanent conservation need to be transferred to a long-term digital repository (the third "age"). This is where this pilot comes in.

The whole goal of the pilot was to ensure that the information package specifications developed in E-ARK and the integration procedures developed are appropriate to support the transference of records between an active or semiactive archival system and a long-term preservation repository.

Scenario	Started	Completed	Summary
1. Migration and Ingest of business records	May	July	Our initial claim was that a systems integration approach
from bespoke business system (Data set 1)	2016	2016	was one of the most effective ways to support demanding archival workflows. In our view, this claim has largely been proven. In a short amount of time, an automatic routine has been developed and implemented that is capable of moving millions of digital objects between the semi-active and inactive stages of an archival workflow with little or no human intervention.

Additional scenarios	Started	Completed	Summary
Integration with OMT via E-ARK DIP	December	January	Until the very end of the project we didn't know whether
	2016	2017	we would have time and resources to run these scenarios.
Repository succession via E-ARK AIP (E-ARK			The E-ARK DIP has been generated and the E-ARK AIP
			exported but the evaluation of the integration could not
AIP exchange experiments)			be finished. We are planning to finish the scenarios in the
			next couple of weeks.

Changes to the original plans

At the pilot planning phase the Porto Municipality also showed great interest in participating in an automatic ingest scenario. So a second, additional, scenario was planned with the same E-ARK component and infrastructure. Later they had some resource planning problems with their local developer who was needed to implement the producer-side infrastructure. The discussions and preparations continued until August 2016, when the Porto Municipality finally decided to delay the project. It is still possible that in the near future this additional scenario can be executed, but definitely not within the time frame of the current project.

Feedback report

The following table summarizes the feedback communication between the pilot staff and tool developers or format specification providers.

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
RODA Repository	For the complete issue history, please refer to the GitHub page:
	https://github.com/keeps/roda
Used in tasks	Ingest of records
Data (input / output)	Historical records, 300 dpi uncompressed TIFF files, 1,2 TB
Performance	Good
Issues	None
Wishes	None
Comments	None
Experiences and recommended	Real world usage brought new requirements to the ingest process of the repository but
practices	these have been solved by the RODA development team.

Recommended practices and further recommendations

This pilot allowed us to learn a few lessons. These are summarised next:

Requirements emerged from the real-world

Working with a real-world data and workflows enabled us to understand that additional requirements had to be accommodated by the repository system. For example, the ingest workflow had to be revised to support the capability of updating existing AIPs with information included in SIPs (called Update SIPs). Also, the full support for Update SIPs had to be added to the specification and software libraries. Moreover, in an unattended systems integration, resilience is an important characteristic. Retry mechanisms had been added to the RIP application to cope with network failures and temporary service unavailability.

Well-established patterns proved to be a successful formula

The RIP application follows a well-established software design pattern called "Pipes and Filters". This pattern makes use of a sequence of tasks (called "filters") that handle part of the entire processing workflow. Each filter is programmed to be simple and stateless. Streaming of data is used whenever possible, enabling the following filters to start processing data even before the entire set of data is completely processed by the previous filter. The most interest aspect of this pattern is the fact that it is possible to change filters in the chain of processing without breaking the processing workflow. This means that the same workflow can be used to process data from different data sources, thus enabling the reuse of the application in many different scenarios. For example, other scenarios have been experimented hat take as input a well-structured folder system and by merely changing the data source filter we were able to ingest data with very little effort.

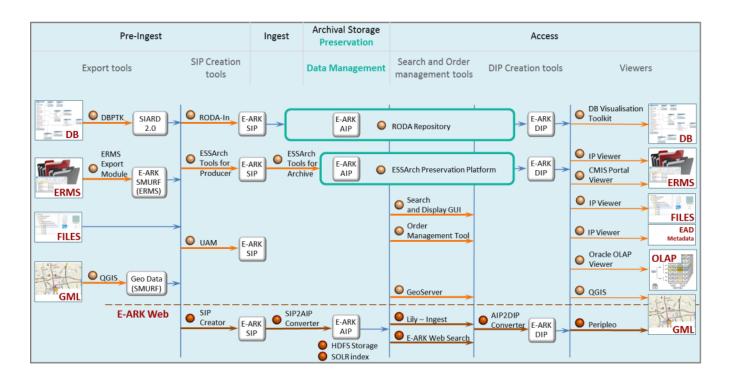
Systems integration is the way forward

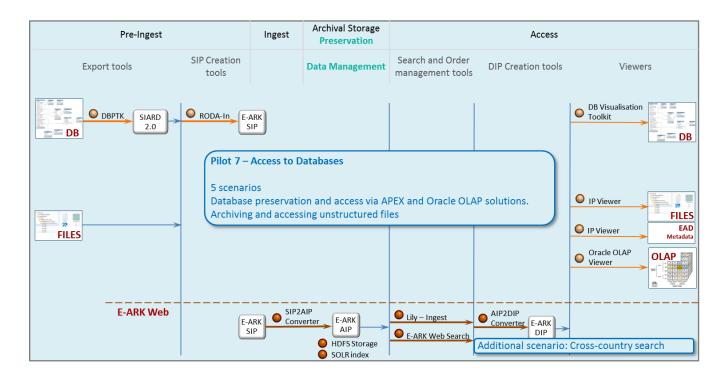
Our initial claim was that a systems integration approach was one of the most effective ways to support demanding archival workflows. In our view, this claim has largely been proven. In a short amount of time, an automatic routine has been developed and implemented that is capable of moving millions of digital objects between the semi-active

Page 70 of 100

and inactive stages of an archival workflow with little or no human intervention. There are always questions of accountability and quality assurance of the entire process, however, the repository side already supports a human validation step at the end of its ingest workflow. This helps to mitigate the previously outlined issues as in the end there is a human expert that attests the quality of the entire process.

Pilots 7 – Access to Databases





Pilot 7	Acce	ess to	o Dat	tabas	ses																			
Task leader	Nati	onal	Arch	nives	of H	unga	ry																	
Supported by	Dan	ish N	latio	nal A	rchiv	es																		
Scope		reser tent.	ntatio	on of	not	less 1	than	2 da	tabas	ses o	f diff	eren	t size	es and	d cor	nple	kities	with	n rest	ricte	d an	d op	en	
Object																							using valua	
Short description	exar the	mine origiı	the a nal st	appli truct	cabil ure a	ity of Ind ir	f data ntelle	a-wa ectua	reho l inte	use c erpre	conce tabili	epts i ity of	n an	arch ested	ival e data	envir a. The	onm e wo	ent ir rking	n ord	er to	mai	ntair	ilot w 1 botł ess w	า
Contacts	Nan	ne (T	itle)							E-m	ail								Skyp)e				
Contact Person	Zolt																							
Pilot staff member	Józs	József Mezei mezei.jozsef@mnl.gov.hu jmezei_92 Pre-Ingest Ingest - Storage Storage - Access																						
OAIS Relevance	Pre-Ingest Ingest - Storage Storage - Access																							
E-ARK Formats	E-ARK SIP X E-ARK AIP X E-ARK DI														DIP	Х								
																data	Х							
E-ARK Tools	c Database Preservation Toolkit	ERMS Export Module	k RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	k SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	sipzaip (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	sour Index	Search and Display GUI	Order Management Tool	 Lily - Ingest 	Geoserver	QGIS	E-ARK Web Search	k AIP2DIP (E-ARK Web)	Database Visualization Toolkit	k IP Viewer	Peripleo	c Oracle (OLAP Viewer)	CMIS portal/viewer
	х		Х			Х		х			Х	Х			Х			Х	х		Х		Х	
Scenario 1	SIP (Creat	ion a	and I	nges	t of c	old (n	ot no	orma	lized) dat	abas	e in S	SIARE	2.0	forn	nat							
Scenario 2	SIP (Creat	ion a	and I	nges	t of u	Instru	uctur	ed fi	les														
Scenario 3	"Ext	ract	SIAR	D Pa	ckage	e froi	n Pre	eserv	vica/E	E-ARI	K AIP													
Scenario 4	(APE	EX/O	racle	Bla	ccess	5)"																		
	•	-																						
Scenario 5	"Sea	arch a	and p	orese	ent SI	ARD	base	d inf	orma	ation	with	E-A	RK ac	cess	tool	S								

Scenarios

Scenario 1	SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format
Description	Create SIP from old (not normalized) database B25. The data is in CSV exports of DBASE files. Create both E-ARK
	and local SIPs and ingest them into E-ARK Web HDFS storage and Preservica archival repository. Both E-ARK
	and local AIPs are generated during the ingest.
OIAS relevance	Pre-Ingest, Ingest
Use-case	Relational database based on SIARD 2.0
E-ARK specifications	E-ARK SIP, E-ARK AIP

E-ARK Tools	DBP	TK, F	RODA	-In, S	SIP2A	IP (E	-ARK	Wel	b), HI	DFS-S	Stora	ge												
Data	Hun	garia	n Pr	osecu	ution	Offi	ce da	itaba	se															
Description	Old	(not	norn	nalize	ed) da	ataba	ase ir	۱CSV	expo	orts d	of DB	ASE 1	files.											
Data type	CSV	files																						
Metadata format	non	e																						
Quantity	mor	e the	en 30	0.00	0 cas	es ar	nd 50	00.00	0 nar	ne. (1,6 G	B)												
OAIS Relevance		l	Pre-l	ngest	:			Ing	est -	Stora	age						Stor	rage	– Acc	cess				
E-ARK Format			E	-ARK																				
specifications			S	IARD	RD 2.0 X SMURF ERMS SMURF SFSB Geodata																			
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	Х	Store then 300.000 cases and 500.000 name. (1/6 GB) FLAKK Export Module Storage Storage Storage Construction of for Producer (ETP) X Storage Construction of for Archiving Module Construction of for Archiving Module Construction of for Archiving Module SiP Creator (E-ARK Web) SIPZAIP (E-ARK Web) SiPZArch Tools for Archiving Module Construction Platform NODA Repository SMURF ESSB Construction Platform Construction Platform																						

Scenario 2	SIP	Crea	tion	and I	nges	t of ι	unstr	uctu	red f	iles														
Description	Crea	ate S	IP fro	om sc	anne	ed do	cum	ents	of th	e Me	eting	g min	utes	of th	ne Ce	ntral	Coin	nmet	tee o	of the	e Hur	igaria	n	
	Soci	alist	Party	y. The	e ima	ige fi	les a	re in	PDF	form	at wi	th EA	\D m	etad	ata. C	Creat	e bot	th E-/	ARK a	and l	ocal S	SIPs a	ind	
	inge	est th	iem i	nto B	27ar	nd Pr	eserv	vica a	rchiv	/al re	posit	ory.	Both	E-AF	RK an	d loc	al Al	Ps ar	e ger	nerat	ed du	uring	the	
	inge	est.																						
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Oth	er (E	xtrac	t and	l Inge	est co	ompu	iter f	iles f	rom	simpl	e file	e-syst	tem)										
E-ARK specifications	E-AF	rk si	P, E-/	ARK A	AIP																			
E-ARK Tools	ROD)A-In	, SIP2	2AIP	(E-AF	RK W	eb), l	HDFS	-Stor	rage														
Data	Scar	nned	mee	ting	minu	tes c	of the	Cen	tral (Comr	nitte	e of t	he H	unga	rian	Socia	list F	arty						
Description	Scar	nned	docu	umen	ts in	file s	syste	ms ir	n PDF	file	and c	orre	spon	ding	meta	idata	(EAD	D)						
Data type	PDF	/JPG	files	(rep	reser	ntatio	ons)																	
Metadata format	EAD																							
Quantity	123.225 files. (101 GB)																							
OAIS Relevance			Pre-l	ngest	t			Ing	est -	Stora	age						Stor	rage	– Ace	cess				
E-ARK Format			E	-ARK	SIP	Х			E	-ARK	AIP	Х									E	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EF	RMS				SMU	JRF S	FSB	Х				Geod	lata	
E-ARK Tools				6			2			_														
	Toolkit			(ЕТІ	e		ET/			orm										lkit				
				cer	Inpo	<u> </u>	ive			latf			_	0						To				
	tion	e		npo	ž	Wet	rch	(q		on P			GU	t To					(q	tion			er)	~
	rvat	npo		r Pr	ving	RK	orA	Ň	≥	/atio			olay	ueu				rch	We	lizat			ewe	wei
	ese.	Σ		ol fo	rchi	(E-A	ols f	(E-ARK Web)	sito	ser	e B		Disp	agen				Sea	ARK	isua			ΡŇ	/vie
	ie Pr	por	_	Toc	al Ai	tor	Toc		epo	Pre	orag	dex	pue	lana	gest	'er		/eb	(E-/	e Vi	Ŀ	•	OLA	rtal
	abas	IS E)	A-Ir	Arch	/ers	creator (E-ARK Web)	Arch	AIP	AR	Arch	S-St	R In	ch a	er ≤	- Ing	serv	s	K V	DIP	abas	iewe	plec	cle (S po
	Database Preservation	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP o	ESSArch Tools for Archive (ETA)	SIP2AIP	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
			X			X		X			X					-	-						-	

Scenario 3	Extract SIARD Package from Preservica/E-ARK AIP
Description	Access database information of the Hungarian Prosecution Office in SIARD format using APEX and OWB access.
	Both E-ARK and local DIPs are generated during access.

	Database Preservation T	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Pla	× HDFS-Storage	× SOLR Index	Search and Display GUI	Order Management Tool	× Lily - Ingest	Geoserver	QGIS	★ E-ARK Web Search	× AIP2DIP (E-ARK Web)	Database Visualization T	× IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
E-ARK Tools	Toolkit			er (ETP)	lule		e (ETA)			Platform				_						Toolkit				
specifications		E-ARK SIP E-ARK AIP X E-ARK DIP SIARD 2.0 X SMURF ERMS SMURF SFSB Geodata													X									
OAIS Relevance E-ARK Format		Pre-Ingest Ingest - Storage Storage - Access E-ARK SIP E-ARK AIP X E-ARK DIP X														v								
Quantity	mor	e the	en 30	0.00	0 cas	es ar	nd 50	0.00	0 nai	me. (1,6 G	B)												
Metadata format	non	е																						
Data type	CSV	files																						
Description	Old	not	norm	nalize	ed) da	ataba	ase ir	ו CSV	exp	orts o	of DB	ASE 1	files.											
Data							ce da			Jeare	,,, , .		(2	/		/	,	001						
E-ARK specifications E-ARK Tools						gest	F-AI	RK W	/eh (9	Searc	h)Δ	P2D	IP (F-	ARK	Weh)		DBV	тк					
Use-case		•				e via	ΑΡΕΧ	(and	Ora	cle Bl)													
OIAS relevance	Acce																							

Scenario 4	Sea	rch a	nd p	reser	nt SIA	RD	based	d info	orma	tion	with	E-AF	RK ac	cess	tools	;								
Description	Acc	ess d	ataba	ase ir	nform	natio	n of	the H	lunga	arian	Pros	ecuti	ion C	ffice	in SI	ARD	form	at us	ing H	HADC)OP k	based	l sea	rch
	and	acce	ess w	ith H	IVE Li	ily Pr	esen	tatio	n in l	ocal	envir	onm	ent.											
OIAS relevance	Acc	ess																						
Use-case	Acc	ess d	ata v	vith C	DLAP	via c	oracle	5																
E-ARK specifications	E-Al	rk ai	P, E-	ARK I	DIP																			
E-ARK Tools	HDF	S-Sto	orage	e , Lily	/ – In	gest,	, E-Af	RK W	/eb (S	Searc	h), A	P2D	IP (E·	ARK	Web)	,	DBV	ТΚ					
Data	Hun	garia	an Pr	oseci	ution	Offi	ce da	taba	se															
Description	Old	(not	norn	nalize	ed) da	ataba	ase ir	n CSV	/ exp	orts o	of DB	ASE	files.											
Data type	CSV	files																						
Metadata format	non	e																						
Quantity	mor	e the	en 30	0.00	0 cas	es ar	nd 50	00.00	0 nai	me. (1,6 G	B)												
OAIS Relevance		more then 300.000 cases and 500.000 name. (1,6 GB) Pre-Ingest Ingest - Storage Storage - Access																						
E-ARK Format			E	-ARk	(SIP				E	-ARK	AIP	Х									E	-ARK	DIP	Х
specifications			S	IARD	2.0	Х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geo	data	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
											х	Х			х			х	X				Х	

Scenario 5	Access information from unstructured files
Description	Create DIP from scanned documents of the Meeting minutes of the Central Coimmettee of the Hungarian

	Soci	alist	Party	y. The	e ima	ige fi	les a	re in	PDF	form	at wi	th EA	۹D m	etad	ata ir	ו E-A	RK W	/eb ⊦	IDFS	stora	ige ai	nd		
	Pres	ervio	ca. Cr	reate	botł	ו E-A	RK ar	nd lo	cal D	IPs.														
OIAS relevance	Acce	ess																						
Use-case	Acce	ess d	ataba	ases	via S	OLR ((no-s	ql)																
	Acce	ess d	ata fi	rom l	E-AR	K we	b/H	DFS s	stora	ge ar	nd fro	om lo	cals	syste	em. S	OLR i	s use	d fo	r sea	rch tł	าe fu	ll tex	t ind	ex
	gene	erate	ed of	the c	locui	nent	s.																	
E-ARK specifications	E-AF	rk ai	P, E-	ARK I	DIP																			
E-ARK Tools	HDF	S-Sto	orage	e, AIP	2DIP	(E-A	RK V	Veb),	, Lily	/ – Inį	gest,	E-AR	K W	eb (S	earch	ו) <i>,</i> Siı	ngle f	file V	iewr					
Data	Scar	nned	mee	ting	minu	tes c	of the	e Cen	tral (Comn	nitte	e of t	he H	lunga	rian	Socia	alist F	Party						
Description	Scar	nned	docu	umen	its in	file s	syste	ms ir	ו PDF	file a	and c	orre	spon	ding	meta	data	(EAI	D)						
Data type	PDF	/JPG	files	(rep	reser	ntatio	ons)																	
Metadata format	EAD																							
Quantity	123	.225	files.	(101	GB)																			
OAIS Relevance			Pre-l	ngest	:			Ing	sest -	Stora	age						Stor	rage	– Ace	cess				
E-ARK Format			E	-ARk	SIP				E	-ARK	AIP	Х									E	-ARK	DIP	X
specifications			S	IARD	2.0				SMU	RF EF	RMS				SMU	JRF S	SFSB	Х				Geod	data	
E-ARK Tools	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display GUI	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	P Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	Õ	Ē	~	ш	<u> </u>							S	S											

Additional scenario	Cross-country search with E-ARK Web (joint scenario with NAS)					
Description	The SOLR index and E-ARK Web infrastructure theoretically makes it possible to perform a federated search					
	over more than one archive	e. When the SOLR index of	the other archival institution	can be "seen" by the search		
	engine (e.g. one institution	has access rights to the ot	hers SOLR) then it can make a	common list of the result.		
	The National Archives of Slo	ovenia and the National A	rchives of Hungary both have a	an E-ARK implementation at		
	their pilot sites. This scenar	io is a simple feasibility stu	udy of cross-country search.			
OIAS relevance	Access	Access				
Use-case	Search and Display					
E-ARK specifications						
E-ARK Tools	E-ARK Web					
Data	Test data in the SOLR index					
Description	The SOLR index of the two	archives will be theoretica	theoretically connected in this sceanrio			
Data type	Not relevant					
Metadata format	Not relevant					
Quantity	small					
OAIS Relevance	Pre-Ingest	Ingest - Storage	Storage - Access			
E-ARK Format	E-ARK SIP	E-ARK AIP	E-ARK DIP SMURF SFSB Geodata			
specifications	SIARD 2.0	SMURF ERMS				

	E-A
	NRK Tools
	Database Preservation Toolkit
	ERMS Export Module
	RODA-In
	ESSArch Tool for Producer (ETP)
	Universal Archiving Module
	SIP creator (E-ARK Web)
	ESSArch Tools for Archive (ETA)
	SIP2AIP (E-ARK Web)
	RODA Repository
	ESSArch Preservation Platform
	HDFS-Storage
х	SOLR Index
	Search and Display GUI
	Order Management Tool
	Lily - Ingest
	Geoserver
	QGIS
х	E-ARK Web Search
	AIP2DIP (E-ARK Web)
	Database Visualization Toolkit
	IP Viewer
	Peripleo
	Oracle (OLAP Viewer)
	CMIS portal/viewer

Execution report

Two pilots (5, 7) decided to test tools' compatibility beyond their core functionality. The core of the Hungarian pilot infrastructure was the E-ARK Web. E-ARK Web has two deployment options, Hungary used the full deployment. In the beginning it was necessary to create a common understanding between AIT (as developer) and NAH (as user) of a very complex system. It was necessary to ensure that everyone understood how it works, and what the idea behind some of the features is. The AIT developers were eager to create a very usable set of components and helped in every way. At the end we think that E-ARK Web is very useful solution and it can be well combined with other E-ARK tools.

Scenario	Started	Completed	Summary
1. SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format	April 2016	September 2016	283 SIARD 2.0 packages have been created and ingested to Preservica.
2. SIP Creation and Ingest of unstructured files	May 2016	October 2016	3703 SIPs have been created and ingested to Preservica.
3. "Extract SIARD Package from Preservica/E- ARK AIP	June 2016	October 2016	Data Explorer (Oracle APEX) was used in this scenario for accessing the databases archived in SIARD 2.0 packages. Scenario has been successfully performed.
4. (APEX/Oracle BI access)"	October 2016	November 2016	Access to database information archived in SIARD 2.0 format was provided using HADOOP based search and access with Lily Presentation in local environment. By OWB the original model can be converted into a Data Warehouse model.
5. "Search and present SIARD based information with E-ARK access tools	September 2016	October 2016	DIP was successfully created for the archived scanned documents.

Additional scenarios	Started	Completed	Summary
Cross-country search with E-ARK Web (joint scenario with NAS)	December 2016	January 2017	The scenario execution was suspended because of security considerations by the archives. The cross-country search is technically feasible but from security point of view it is risky. In the future if the archives build the infrastructure to implement a publicly accessible E-ARK Web solution outside their firewall then it can be reached from the search engine of another archive with E-ARK Web.

Changes to the original plans

There were no major changes. The scenarios have been performed according to plans in DoW and D2.3 Detailed Pilot Requirements.

Feedback report

E-ARK Tool – Version	Issues (bugs, wishes, comments)			
	Experiences / Recommended practices			
E-ARK Web	For the complete issue history, please refer to the GitHub page:			
(Virtual deployment)	https://github.com/eark-project/earkweb			
Used in tasks	SIP to AIP conversion, Lilly ingest, SOLR search, AIP to DIP conversion			
Data (input / output)	Input: 2 different data set			
	Output: depending on component			
Performance	ОК			
lssues	At the beginning there were some issues, mostly with compatibility.			
	No issues left at the end of the pilot			
Wishes	None			
Comments	None			
Experiences and best practices	None			

E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
Database Preservation Toolkit	For the complete issue history, please refer to the GitHub page:
(version2.0.0-beta4.2)	https://github.com/keeps/db-preservation-toolkit
Used in tasks	Data extraction – scenario 1
Data (input / output)	Input: Hungarian prosecution office data
	Output: SIARD2.0 package
Performance	Excellent
Issues	There have been several issues with DBPTK related SIARD 2.0 output. KEEP Systems has
	corrected all the bugs and the response time was excellent. After the completion of the
	scenarios no known issues remained.
Wishes	A tool or function for automatic validation of SIARD 2.0 would be nice to have.
Comments	None
Experiences and recommended	None
practices	

E-ARK Tool – Version	Issues (bugs, wishes, comments)		
	Experiences / Recommended practices		
RODA-In	For the complete issue history, please refer to the GitHub page:		
(2.0.0 Alpha 7.4)	https://github.com/keeps/roda-in		
Used in tasks	Create SIP - Create an E-ARK SIP Package		
Data (input / output)	Input: Unstructured data		
	Output: EARK SIP in a *.zip file		
Performance	ОК		
lssues	No issues left at the end of the pilot		
Wishes	None		
Comments	None		
Experiences and best practices	None		

E-ARK Tool – Version	Issues (bugs, wishes, comments) Experiences / Recommended practices
IP Viewer	
Used in tasks	View DIP

D2.5 Recommended Practices and Final Public Report on Pilots

Data (input / output)	Input: DIP
Performance	Good
Issues	None
Wishes	None
Comments	None
Experiences and best practices	None

Recommended practices and further recommendations

AIT – E-ARK WEB

EARK WEB's SIP creator is a very simple application for real-life scenarios. We have therefore been using the more complex RODA-In instead.

Even if only ingesting one SIP we recommend to use the Batch SIP ingest, because it goes through almost every ingest task automatically, so you don't have to click and run every tasks manually! But in order to understand the workflow one should use it manually once or twice.

Please note that using Batch SIP Ingest AIPs won't get uploaded into Lily automatically. In a later step one can load the AIPs into Lily.

RODA-In

RODA-in offers a lot of features that makes SIP creation very easy and fast. Take your time and examine all the possibilities.

If you select a folder tree and drop it in the centre, and want to fill out the metadata cells with similar data: you can just hold CTRL and select every SIP in the centre field, and fill out the metadata cells on the right, and hit OK. Now you have the similar metadata for the selected SIPs. Some metadata cells cannot be the same.

We had many folders in a root folder, and every single folder had two subfolders. We had dropped them into the centre field and used the second option, that means every single folder will be an SIP. On the right side we created a second representation and we separated those two folders into rep1 and rep2. The type of the files were jpg in the first and pdf/a in the second folder.

DBPTK/DBVTK

If you would like to use DBVTK and DBPTK, make sure the version of DBPTK is compatible with DBVTK version that you would like to use or later you might have to recreate every single SIARD file.

When you make an export from an Oracle DB with DBPTK, and you want to import it into your own database: you might have to recreate the same environment to import the SIARD into, because there could be a problem with the tablespace names.

Oracle Warehouse Builder and OLAP Viewer

This is a very nice and informative way of presenting data. It should be noted, however, that the whole procedure of creating this result requires a lot of effort. This not an automatic procedure of DIP creation.

External evaluations

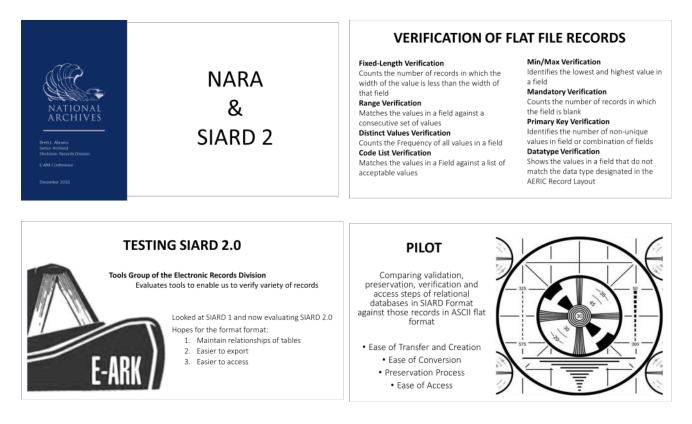
We have been encountering a growing interest about the E-ARK project and its results in the archival community. At DLM Forum meetings and at the E-ARK Final Conference we have talked to people who have not only showed general interest about E-ARK tools and format specifications but have plans to try them in the near future and asked for support in specific problems.

Promoting and supporting external evaluation of our products has been primary task at WP2. An external evaluation or validation, according to the Description of Work, is an evaluation or implementation of E-ARK products by members of DLM Forum and DPC or third parties outside the project with limited involvement from consortium members.

Organization	Title	Scenario Description	Data set
National Archives and	Testing SIARD 2.0	NARA has performed 1 pre-ingest, 1 pre-	Status: Completed
Records Administration		ingest/ingest and 1 access scenarios archiving 2	
(NARA, USA)		different databases as SIARD 2.0 files with Database	
		Preservation Toolkit. NARA has generated SIARD 2.0	
		files from databases, created SIPs in local format and	
		ingested them to their local preservation system.	
Ministerio de Hacienda y	Archiving complete	MinHAP plans to test DBPTK for archiving databases.	Status: In progress
Función Pública (MinHAP)	databases	They are generating SIARD 2.0 files from MySQL and	
		later from Oracle databases. Also testing E-ARK SIP	
		creation tools for creating E-ARK SIP format	
		information packages in the future but today	
		MinHAP uses the Spanish SIP standard.	
Swiss Federal Archive	SIARD 2.0 validation	Testing DBPTK and validate DBTK's SIARD 2.0 output.	Status: Under
(SFA)		The new version of SIARD has been developed in	preparation
		cooperation by the E-ARK project and the Swiss	
		Federal Archive.	
		SFA plans to test DBTK and validate the created	
		SIARD 2.0 files.	
Agenda Open Systems	Testing the possible use	Agenda Open Systems is an Alfresco service provider	Status: Under
	of ERMS Export Module	in Slovenia. They are interested in the product. The	preparation
		latest version with source code has been sent to AOS	
		lately.	
National Archives of Chile	Piloting E-ARK toolset	The NACh has no electronic archival solution so far.	Status: Preliminary
(NACh)	for electronic archiving	They had been planning to launch one when they	arrangements are in
		heard about the E-ARK project. We've been having	progress at the archive
		several conversations over the possibilities of trying a	in order to test and
		subset of E-ARK tool portfolio with their consultant	launch their first
		Daniel Cáceres in the subject. They are really	electronic archival
		interested but organizational and IT arrangements go	solution.
		very slowly. At the time of this report there is no	
		official decision about the project.	

The following organisations have performed (or performing) external evaluation activities during the project:

The following slides are from the presentation by Brett Abrams of NARA at the E-ARK Final Conference, at Budapest.



Please note that at moment of finishing this document some of the above external evaluation scenarios are still in progress. Since they are outside of the project E-ARK had no influence on resource planning or scheduling these activities.

We have found it very encouraging that major external organisations are already starting to work with our project tools in preparation to deploy them operationally.

E-ARK project members are committed to promote and support above and later external evaluations after the official ending of the project.

Pilot evaluation

This chapter provides an evaluation of the pilots against their goal given as detailed success criteria by the document D2.3 Detailed Pilot Requirements.

Work Package 2 Objectives (according to the Description of Work):

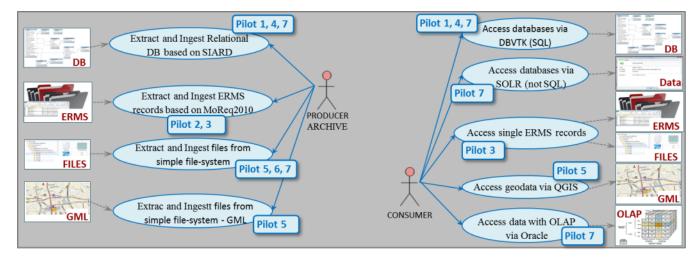
The overall objective of this work package is to ensure that the scenarios implemented at 7 identified pilot sites are both realistic and relevant. That is, that they bring together a meaningful subset at each site of the use cases that define establish a general model of the E-ARK service.

Project level pilot success evaluation

Pilot level success criteria as defined in D2.3 Detailed Pilot Requirements

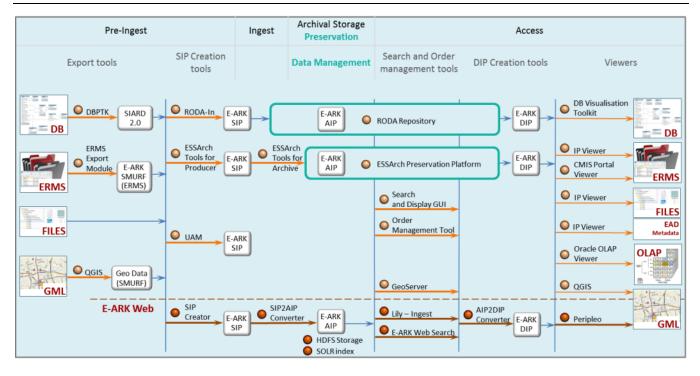
No #	Requirement	MoSCoW	Comment
7.2	The whole E-ARK full-scale pilot is successful if all the high-level E-ARK use cases are piloted in at least one of the pilots	М	
7.3	The whole E-ARK full-scale pilot is successful if all of the core E-ARK tools are piloted in at least one of the pilots	М	
	The whole E-ARK full-scale pilot is successful if most of the E-ARK web (Integrated Prototype) tools are piloted in at least one of the pilots	М	

E-ARK uses-cases



	Use Case	Pilot	Scenario	Succesfull?
Pre-Ingest	Extract and Ingest relational database based on SIARD 2.0	Pilot 1	Scenario 1-4	
		Pilot 4	Scenario 1-4	V
		Pilot 7	Scenario 1	
		External	NARA,	
		evaluation	MinHAP, SFA	
	Extract and Ingest ERMS records based on MoReq2010	Pilot 2	Scenario 1-3	1
		Pilot 3	Scenario 1,3	V
		Pilot 1,3	Additional sc.	
	Extract and Ingest computer files from simple file-system	Pilot 5	Scenario 1,3	
	– GML			
	Extract and Ingest computer files from simple file-system	Pilot 5	Scenario 1,3	1
	- Other (please specify)	Pilot 6	Scenario 1	
		Pilot 7	Sceanrio 2	
Ingest	Ingest E-ARK SIP (Generate E-ARK AIP)	Pilot 2	Scenario 1-3	
Access		Pilot 5	Scenario 1,3	
		Pilot 6	Scenario 1	
		Pilot 7	Scenario 1-2	
	Access databases via DBVTK (sql)	Pilot 4	Scenario 1-4	1
		Pilot 1	Additional sc.	V
	Access databases via SOLR (no-sql)	Pilot 5	Scenario 3	
		Pilot 7	Scenario 3-5	v
	Access single ERMS records	Pilot 3	Scenario 2,4	
		Pilot 2	Additional sc.	v
	Access geodata via qgis	Pilot 5	Scenario 2,4	\checkmark
	Access data with OLAP via oracle	Pilot 7	Sceanrio 4	

E-ARK tools and format specifications



	Tools	Pilot	Scenario	Succesfull?
Pre-Ingest	Database Preservation Toolkit	Pilot 1	Scenario 1-4	
		Pilot 4	Scenario 1,2	v
		Pilot 7	Scenario 1	
		External	NARA,	
		evaluation	MinHAP, SFA	
	ERMS Export Module	Pilot 1	Additional sc.	
		Pilot 3	Additional sc.	v
	RODA-In	Pilot 5	Scenario 1	
		Pilot 7	Scenario 1,2	v
	ESSArch Tool Producer (ETP)	Pilot 2	Scenario 1-3	
	- Redesigned UI, E-ARK compatible version	Pilot 2	Additional sc.	v
		Pilot 5	Scenario 3	
	Universal Archiving Module	Pilot 3	Scenario 1,3	\checkmark
	SIP creator (E-ARK Web)	Pilot 7	Scenario 2	\checkmark
Ingest	ESSArch Tools Archive (ETA)	Pilot 3	Scenario 1,3	
		Pilot 5	Scenario 2	v
	SIP2AIP (E-ARK Web)	Pilot 5	Scenario 1,2	
		Pilot 7	Scenario 1,2	v
	RODA Repository	Pilot 6	Scenario 1	\checkmark
	ESSArch Preservation Platform	Pilot 3	Scenario 1,3	\checkmark
	HDFS-Storage	Pilot 7	Scenario 1-5	\

	Tools	Pilot	Scenario	Succesfull?
Access	SOLR Index	Pilot 5	Scenario 1-4	1
		Pilot 7	Scenario 1-5	V
	Search and Display GUI	Pilot 5	Scenario 2,4	\checkmark
	Order Management Tool	Pilot 5	Scenario 2,4	\checkmark
	Lily – Ingest	Pilot 5	Scenario 2,4	1
		Pilot 7	Scenario 3-5	V
	Geoserver	Pilot 5	Scenario 2,4	\checkmark
	QGIS	Pilot 5	Scenario 1-4	\checkmark
	E-ARK Web Search	Pilot 7	Scenario 3-5	\checkmark
	AIP2DIP (E-ARK Web)	Pilot 5	Scenario 2,4	1
		Pilot 7	Scenario 3-5	v
	Database Visualization Toolkit	Pilot 4	Scenario 2,4	1
		Pilot 1	Additional sc.	V
	IP Viewer	Pilot 5	Scenario 2,4	1
		Pilot 7	Scenario 5	V
	Peripleo	Pilot 5	Scenario 2,4	\checkmark
	Oracle (OLAP Viewer)	Pilot 7	Scenario 4	\checkmark
	CMIS portal/viewer	Pilot 3	Scenario 2,4	_

	Use Case	Pilot	Scenario	Successful?
Information		Pilot 2	Scenario 1-3	
Package format		Pilot 3	Scenario 1,2	~
specification	E-ARK SIP	Pilot 5	Scenario 1,2	
	(Supplier Information Package)	Pilot 6	Scenario 1	
		Pilot 7	Scenario 1,2	
		Pilot 2	Scenario 1-3	1
	E-ARK AIP	Pilot 5	Scenario 1,2	~
	(Archival Information Package)	Pilot 6	Scenario 1	
		Pilot 7	Scenario 1,2	
	E-ARK DIP	Pilot 3	Scenario 2,4	1
		Pilot 5	Scenario 2,4	~
	(Dissemination Information Package)	Pilot 7	Scenario 3-5	\ \ \ \
Content type		Pilot 1	Scenario 1-4	1
specification		Pilot 4	Scenario 1-4	~
	SIARD 2.0	Pilot 7	Scenario 1	
		External	NARA,	
		evaluation	MinHAP, SFA	

E-ARK SMURF ERMS	Pilot 2 Pilot 3 Pilot 1,3	Scenario 1-3 Scenario 1-4 Additional sc.	\checkmark
E-ARK SMURF SFSB	Pilot 5 Pilot 6 Pilot 7	Scenario 1-4 Scenario 1 Scenario 2,5+D14	\checkmark
E-ARK SMURF Geodata	Pilot 5	Scenario 1-4	\checkmark

Pilot and scenario level success evaluation

The full-scale pilots have pilot level and scenario level success criteria defined in D2.3 Detailed Pilot Requirements. The following table provides the evaluation details at both levels.

Pilot / Scenario	Success criteria	Successful?
Pilot 1	The following E-ARK tools will be tested in a pilot environment: Database Preservation Toolkit	\checkmark
Scenario 1	Extract records from MS SQL Server database containing 50-60 tables and about 90.000 records. (95% success rate)	\checkmark
Scenario 2	Extract records from MySQL database about 5 million records.(95% success rate)	\checkmark
Scenario 3	Extract records from MS SQL Server database containing documents. (95% success rate)	\checkmark
Scenario 4	Extract records from MS SQL Server database containing documents. (95% success rate)	\checkmark
Pilot 2	The following E-ARK tools will be tested in a pilot environment: ESSArch Tools Producer (ETP), ESSArch Tools Archive (ETA), ESSArch Preservation Platform (EPP).	\checkmark
	This pilot will be considered a success if we are able to use and evaluate these tools in all three scenarios, producing an output that can be stored in depot. The National Archives of Norway have been using an earlier version of EPP in production for a couple of years, the ETP and ETA are newly developed software from which user experience will be gathered and disseminated during piloting.	
	The new version of ETP was tested in an additional scenario because of the incompatibilities at the producer IT infrastructure. The ETP tool has also been tested in Pilot 5.	\checkmark
Scenario 1	Ingest around 20 GBs of EDRMS data from public producer converted into Noark 4 output	\checkmark
Scenario 2	Ingest around 5 GBs of EDRMS data from public producer converted into Noark 4 output	\checkmark
Scenario 3	Ingest around 335.000 registered persons (105 MB) from the national registry of licenced hunters.	\checkmark
Pilot 3	The following E-ARK tools will be tested in a pilot environment: ERMS Export Module (see Aditional Scenario), UAM (Universal Archival Module), E- ARK CMIS Browser (Yes/No)	\checkmark
	The ERMS Export Module was tested in 2 additional scenarios because of the late deployment of the appropriate version corresponding to local producer's requirements.	\checkmark

Scenario 1	Extract records from EDRM, create and ingest SIP of different documents of Ministry of Justice with different retention period (95% success rate)	\checkmark
Scenario 2	Provide access to archived records of Ministry of Justice (95% success rate)	\checkmark
Scenario 3	Extract records from EDRM, create and ingest SIP of different documents of Ministry of Justice with different retention period (95% success rate)	\checkmark
Scenario 4	Provide access to archived records of Ministry of Justice (95% success rate)	\checkmark
Pilot 4	The following E-ARK tools were tested in a pilot environment: Database Preservation Toolkit (Done), RODA-In (see note below) RODA-In wasn't used in this pilot because the native SIP creation tool was required to ingest into the preservation system of the Business Archives. RODA-In, on the other hand, was tested in Pilot 5 and 7.	\checkmark
Scenario 1	Exporting records from database for more than 12 000 business records from bespoke business system	\checkmark
Scenario 2	Importing records to database for more than 12 000 business records from bespoke business system	\checkmark
Scenario 3	Exporting records from database with files for more than 200 000 business records from bespoke business system (success rate 85% due complicated database architecture)	\checkmark
Scenario 4	Importing records to database with files for more than 200 000 business records from bespoke business system (success rate 85% due complicated database architecture)	\checkmark
Pilot 5	The following E-ARK tools will be tested in a pilot environment: ESSArch Tools Producer (ETP), ESSArch Tools Archive (ETA), ESSArch Preservation Platform (EPP), Search and Display GUI, Order Management Tool, IP Viewer, along with components of the Integrated Prototype (E-ARK Web): Order Submission Service(see note below), Lily-Ingest, Geoserver, Peripleo, with the integration of QGIS (Yes/No) In the final order management solution of WP5 Order Submission Service is not a separate software component any more. The planned functionality has been implemented in the Order Management Tool.	\checkmark
Scenario 1	SIP creation, verification and ingest of more than 1000 records with a vector geodata layer. (90% success rate)	\checkmark
Scenario 2	Finding, accessing, modifying and exporting a DIP containing a vector geodata layer of more than 1000 records. (90% success rate)	\checkmark
Scenario 3	SIP creation, verification and ingest of more than 200 records with a vector geodata layer. (90% success rate)	\checkmark
Scenario 4	Finding, accessing, modifying and exporting a DIP containing a vector geodata layer of more than 200 records. (90% success rate)	\checkmark
Pilot 6	Test the E-ARK compatible RODA Repository in a pilot environment. (Yes/No)	\checkmark
Scenario 1	Ingest of no less that 900 historical records in E-ARK SIP format automatically generated by a specially developed integration tool (90% success rate)	\checkmark
Scenario 2	At the pilot planning phase the Porto Municipality also showed great interest in participating in an automatic ingest scenario. So a second scenario was planned with the same E-ARK component and infrastructure. Later they had some resource	Postponed (Outside scope of

	planning problems with their local developer who was needed to implement the producer-side infrastructure. The discussions and preparations continued until August 2016, when the Porto Municipality finally decided to delay the project. It is still possible that in the near future this scenario can be executed, but definitely not within the time frame of the current project, so we had to cancel this scenario and at that time it was too late to start another.	DoW)
Pilot 7	The following E-ARK tools will be tested in a pilot environment: DBPTK, RODA-in and DB viewer (Sofia) using Oracle OLAP Viewer, along with components of the Integrated Prototype (E-ARK Web): SIP2AIP, HDFS-Storage, Lily-Igest, Search, AIP2DIP (Yes/No)	\checkmark
Scenario 1	Create SIP and Ingest more than 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)	\checkmark
Scenario 2	Create SIP and Ingest more than 30.000 pages of scanned pdf images of meeting minutes of the former Hungarian Socialist Party. (95% success rate)	\checkmark
Scenario 3	Provide access for more than 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)	\checkmark
Scenario 4	Provide access for more than 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)	\checkmark
Scenario 5	Provide access for more than 30.000 pages of scanned pdf images of meeting minutes of the former Hungarian Socialist Party. (95% success rate)	\checkmark

Referenced Documents

In this document the following external document references have been used:

D2.1 General Model 1.0

http://eark-project.com/resources/project-deliverables/5-d21-e-ark-general-pilot-model-and-use-case-definition

D2.3 Detailed Pilot Requirements

http://eark-project.com/resources/project-deliverables/60-23pilotsspec

D2.4 Pilot Documentation

Part 1: http://eark-project.com/resources/project-deliverables/87-d24docs-p1-1

Part 2: http://eark-project.com/resources/project-deliverables/88-d24docs-p2-1

The latest version of the General Model can be found in the E-ARK Knowledge Base and also accessible from the E-ARK project web site: <u>http://eark-project.com/resources/general-model</u>

Appendix 1 – Extract from E-ARK DoW

E-ARK will pilot an end-to-end OAIS-compliant e-archival service covering ingest and reuse of structured and unstructured data addressing the needs of data subjects, data owners and data users. It will integrate tools currently in use in partner organisations, and provide a framework for providers of these, and similar tools, to ensure compatibility and interoperability. The project has three phases resulting in a set of tool instantiations, a validated pilot platform and a set of recommended practices based on evaluation of the pilot. This approach supports the planned three-tier piloting strategy (full-scale pilot, shorter 'stretch' pilots and external validation).

The work has been organised into six work packages, as shown in the diagram below. Specialist skills are associated with each WP and this grouping of activities also reduces inter-dependences between work packages and localises risk. The detailed definition of the work required in each work package includes a diagrammatic 'product flow' diagram. These express the flows and dependences within and between work packages.

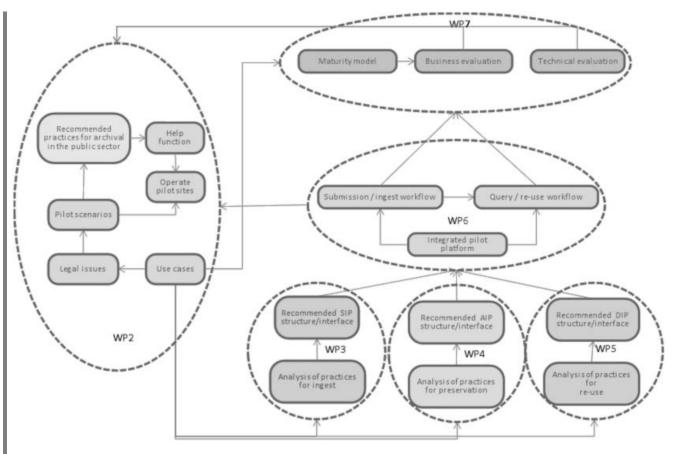


Figure 1: E-ARK – Overall Approach

WP2 is concerned with ensuring that the needs of each pilot site are addressed in the work packages that actually deploy the tools, and that the pilot scenarios are achievable and reflect any legal and logistical constraints. It also supervises the acquisition of appropriate data from the data-owners working with each pilot site and, finally, documents the knowledge gained from the pilot in the form of recommended practices.

WP3, WP4 and WP5 are responsible for the information packages that encapsulate the content and related metadata that is being archived, respectively during the workflows for **submission** (SIP - the data structures used by the data owner to enable ingestion of the content), **archival** (AIP - the data structures used by the repository operator to enable preservation functions) and **dissemination** (DIP – the data structures used for extraction and reuse of content). The mapping of SIP to AIP and AIP to DIP provide the mechanism for integration of tools/services in the pilot and compliance with these three data-structures provides the mechanism for interoperability between tools/services.

WP6 provides access to ingest and re-use tools/services to be deployed in the pilot, based on the implementation of a repository supporting the open source AIP schema from WP4. Pilot sites can either use this open-source solution or work with their platform-providers to implement SIP/AIP and AIP/DIP mappings of their own, supported through their community of interest within the project.

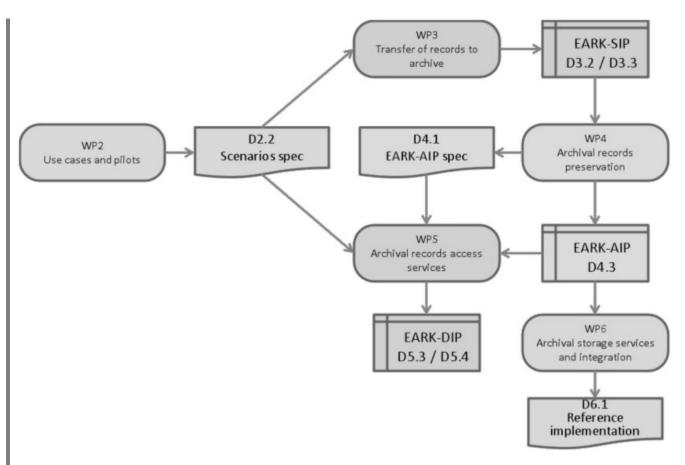
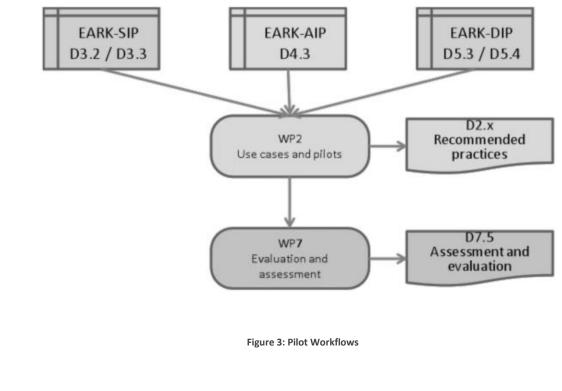


Figure 2: E-ARK Technical Integration

WP7 is responsible for evaluating the pilot service from technical and commercial perspectives based on criteria established for each scenario by WP2 and will utilise a maturity model developed in the TIMBUS project. Following the pilot deployments, both technical and business evaluations will be carried out and stored in a knowledge base, based on the indicators created for each pilot component. For example, a formal specification of the pilot ingest workflow will include information about how it has been developed and tested.



More specifically, there are two distinct work-streams orchestrating the work required to integrate the pilot service and the work required to deploy, support and evaluate the pilot. This is summarised above, one leading to the WP6 deliverable for an "Integrated Platform Reference Implementation" (M24) and the other leading to the WP7 deliverable "Pilots Assessment – Final" (M36).

Piloting, which is the responsibility of WP2, consists of seven instances of parts of the E-ARK service.

The full scale pilots planned in the E-ARK Description of Work (DoW)

T2.5.1 Full scale pilot no. 1. – SIP creation of relational databases

Task leader: Danish National Archives.

Supported by: Magenta

Scope: Not less than 4 databases of different sizes and complexities (one contains several million records)

Object: Creating SIPs for relational databases using the tool created in WP3, T3.3: SIP Creation Tools, for further evaluation.

Participants: Danish National Archives (digital archive), Magenta, the data provider institution creating the archival records.

Resource plan: 8 person months for setting up the pilot (assisting the archivists and data provider in preparing the transfer), carrying out the pilot (transfer, quality checking, metadata amendments), testing the results and reporting.

Timeframe: M28-M33

Preconditions: M03.3 and M03.4

Position in the project: DNA will pilot SIP creation and ingest specified by WP3

Contribution to the project outcome: the pilot demonstrates the applicability of the project outcomes in creating SIPs from relational databases

T2.5.2 Full scale pilot no. 2. – SIP creation and ingest of records

Task leader: National Archives of Norway

The main part of the pilot includes the export of electronic records and their metadata from EDRM systems and databases of Norwegian public sector institutions, transfer and ingest them to the NAN digital repository.

Scope: Not less than 2 transfers of unstructured records with mixed restricted and unrestricted material, and not less than 1 transfer of structured records.

Object: Extract data from EDRMS and databases, create SIPs for structured and unstructured records using ESSArch Tools, ingest the SIPs to the repository using ESSArch Preservation Platform, for further evaluation.

Participants: National Archives of Norway (digital archive), data provider

Resource plan: 6 person months for setting up the pilot (assisting the archivists and data provider in preparing the transfer), carrying out the pilot (transfer, quality checking, metadata amendments), testing the results and reporting

Position in the project: NAN will pilot SIP creation and ingest specified by WP3

Timeframe: M28-M33

Preconditions: M03.3 and M03.4

Contribution to the project outcome: the pilot demonstrates the applicability of ESSArch Tools and the ingest functions of ESSArch Preservation Platform.

Data owners: to be defined at the time of the pilot.

Platform: ESSArch Tools will be used to create the SIPs, and ESSArch Preservation Platform will be used to create and manage the AIPs, both delivered by ES Solutions. NAN IT-department is responsible for the systems operation.

T2.5.3 Full scale pilot no. 3. – Ingest from government agencies

Task leader: National Archives of Estonia

The main part of the proposed pilot 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.

In addition 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.

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); The whole set will include about 5000 records (but depends on the exact agency of course).

Objects: EDRMS at a governmental agency (Alfresco), records preparation tool (UAM), digital preservation and access systems (SDB, AIS);

Participants: National Archives of Estonia (digital archive), one governmental agency (data provider), general public (access to records);

Number of users: Archivists at NAE (dealing with the ingest and preservation, about 3 persons); archivists at the agency (about 2-3 persons preparing the export/transfer and providing means for continuous in-house usage), general public - we have around 1000 daily users at the archives virtual reading room / AIS but obviously we are not able to predict how many of these will actually access and use the information ingested through the pilot;

Resource plan: about 4 person months (includes updates to the EDRMS installation at the agency, to UAM and SDB/AIS, setting up and running the pilot).

Position in the project: NAE will implement and pilot the records export requirements, SIP format and transferingest workflow specified by WP3 and the access services specified by WP5;

Timeframe: setting up pilot sites through M25 – M27, running the pilot for six months through M28 – M33, which means that the records are available for the general public for at least three months;

Preconditions: M03.3, M03.4, M04.2, M05.4, M05.6.Records are available at the agency in digital form and enriched with metadata; it is possible to export the records; records export, preparation, transfer, ingest and access functionalities have been updated according to project deliverables in Alfresco, UAM, SDB and AIS;

Contribution to the project outcome: the pilot demonstrates the applicability of the project outcomes inside the framework of Estonian public sector legislation and the tools applied at NAE.

Platform and data owners: a specific data provider has not been selected for NAE, NAE notified the Ministry of Economics and Communication (in charge for co-ordinating e-Gov and electronic records management in Estonia) and they have promised their full support when it comes to actually selecting the specific agency. We are aiming to use Alfresco as the commercial system which we ingest data FROM (there are about 10-20 agencies in Estonia who use it – so quite a few possibilities). SDB is the preservation platform which we employ to ingest data.

T2.5.4 Full scale pilot no. 4. – Business archives

Task leader: National Archives of Estonia

Supported by: Estonian Business Archives

Estonian Business Archives, Llc. is a privately owned archiving services provider. The main client base of the company is comprised of private businesses in Estonia for archiving and preservation of both paper and digital

records. The business archives pilot in the E-ARK project will focus on transfer of electronic records from private companies to the digital archive solution of the Estonian Business Archives and their subsequent description required for archiving and preservation.

Scope: Transfer of business records to a digital archive solution in a business archive, quality control, enhancement of description and AIP creation.

Object: bespoke business system that contains records (pilot will test an annual batch of ca 4,500 records); financial and CRM systems that contain records (pilot will test an annual batch of ca 15,000 records).

Participants: Estonian Business Archives, Llc (digital archive), two private companies (data providers).

Number of users: The archived business records are for the sole use of their owner-company only.

Resource plan: 4 person months for setting up the pilot (assisting the companies' archivists in preparing the transfer; setting up and configuring the IT infrastructure at EBA), carrying out the pilot (transfer, quality checking, metadata amendments, AIP creation), testing the results and reporting.

Position in the project: The pilot will report on the suitability of the ES Tools and ES Preservation Platform for processing electronic records from business systems.

Timeframe: M25-M27: setting up the pilot sites; M28-M31: running the pilots; M32-M33: testing and reporting.

Preconditions: M03.3, M03.4, M04.2, M05.4, M05.6.

Contribution to the project outcome: The business archives pilot will provide a view how the tools developed by the project can be implemented in the private sector setting. The pilot will assess to what extent these tools add value to the existing archiving services and workflows established in the corporate sector. The nature of objects used in the pilot – business information systems that contain or manage records – is slightly different from the public sector use cases that mostly rely on EDRM systems or databases of records.

Platform and data owners: The systems that records will be transferred from and the current digital archive solution at the EBA are all bespoke solutions.

T2.5.5 Full scale pilot no. 5. – Preservation and access to records with geodata

Task leader: National Archives of Slovenia.

Supported by: Danish National Archives

During the e-ARK project the standardised 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.

Scope: Pilot will prove that the SIP and DIP implementations fulfil 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 (via open data method, direct access in the archives and use GIS data as search criteria in the DIP contents).

Object: pilot report with recommendations about urgent improvements and possible future improvements support for WP6 & WP7 setting up the work environment of selected E-ARK archival tools provide real life examples how the project deliverables can be used

Position in the project: Pilot will prove usability of specification and tools for supporting ingest (WP3 D03.3) and access (WP5 D5.3, D5.4) of archival records with specific data. Uses specifications and tools for supporting ingest (WP3 D03.2, D03.2) and access (WP5 D5.2, D5.3, D5.4)

Participants: National Archives of Slovenia (digital archives), Danish National Archives (best practice exchange)

Resource plan: 7 person months (6 pm for National Archives of Slovenia 1 pm for DNA)

Preconditions: M03.3, M03.4, M04.2, M05.4, M05.6.

Timeframe: M25-M27: setting up the pilot sites; M28-M31: running the pilots; M32-M33: testing and reporting.

Platform: DBExport Tool

T2.5.6 Full scale pilot no. 6. – Seamless integration between a live document management system and a long-term digital archiving and preservation service

Task leader: KEEP SOLUTIONS

RODA (Repository of Authentic Digital Records) is a long-term digital repository system that implements an ingest workflow that not only validates SIPs, but also checks its contents for virus, does format identification, extracts technical metadata, and migrates file formats to more "preservable" surrogates. RODA also provides access to digital information in several forms such as search/navigate over available metadata as well as online visualisation and download of originals, preservation formats and dissemination derivatives. Administration interfaces allow back-office users to manage fonds/collections and define rules for preservation actions. All interactions between users (human and machines) and the repository are logged for security and accountability reasons. RODA ensures that ingested data is authentic by recording PREMIS metadata on all actions performed by the repository, records provenance in archival metadata standards such as ISAD(g), and ensured integrity and availability by frequently monitoring data and making sure that it has not been tampered with. More recently, RODA has been enhanced to support preservation plans developed in Plato, thus proving a full-cycle preservation environment for digital objects ensuring usability and readability of ingested data.

RODA currently supports the Digital Archiving and Preservation Service at the Portuguese National Archives. This service allows public bodies to submit digital content to the archiving service for long-term preservation. The Digital Archiving and Preservation Service takes care of the necessary procedures to keep data accessible for long periods of time (in the scale of decades). Producers have special privileges in the system, allowing them to manage their data and change the structure of their fonds/collections. Data is submitted via SIP files that need to be manually prepared by producers using an offline tool called RODA-in.

Scope and objectives: The goal of this pilot is two-fold. On one hand, Keep Solutions demonstrates that the pan-European SIP structure designed in the WP3 is adequate to support the media types currently supported by RODA (i.e. relational databases, text documents, video, audio and images) and, on the other hand, that the most adequate and scalable form of ingest is to automate the SIP creation process. In order to achieve this, we will tap into a running Document Management System and, based on appraisal and selection strategy installed, we will extract, transform, aggregate and create Submission Information Packages that conform to the pan-European SIP format defined in WP3 that are ready to be ingested in RODA.

Participants: In this pilot we will make use of data produced by several bodies of the Portuguese public administration. One already confirmed is a project partner, the IST. The IST is a Portuguese public university that delivers top quality higher education and engages in research, development and innovation activities. In its activities, several forms of content with high administrative, legal, financial and informational value are produced every day. During the project lifetime the IST will engage in a parallel project to re-engineer a large part of the technology that supports its administrative services, which will include the acquisition and deployment of an integrated archival system. This makes this pilot an excellent example as information assets to be ingested from the actual production systems are expected to be highly unstructured and in desperate need of preservation. Besides the IST, the consortium will also take advantage of the role that AMA plays in the structure of the Portuguese Public Administration to complement this case with more data providers.

Resource plan: 7 person months. 6 PM for KEEPS for development, testing and integration and 1 PM for IST for consulting and liaison with the departments that will provide data to the pilot.

Position in the project: RODA already supports preservation actions and dissemination interfaces for 5 media types. This pilot will focus on enhancing the ingest process by connecting the long-term repository to the Document Management Systems active at the data producer's location this way demonstrating SIP suitability for packaging various content types and scalability by providing a seamless ingest process that requires little or no human intervention.

Timeframe: Between M25–M27 the pilot will be deployed. Between M28–M33 the ingest process will run in parallel with the SIP creation process.

Preconditions: pan-European SIP format defined (WP3). RODA must be enhanced to support the new SIP format (WP3). Automatic SIP creation tool/middleware must be developed to integrate the data provider DMS with the long-term repository.

Contribution to the project outcome: The pilot will demonstrate that the pan-European SIP structure designed in the WP3 is adequate to support the content types currently supported by RODA (i.e. relational databases, text documents, video, audio and images) and, on the other hand. The pilot will also demonstrate and provide a framework for automatic SIP creation and DMS-Repository interoperability showing the scalability of whole ingest process.

Platform and data owners: The owner of the data in this pilot will be the IST. Multiple systems are currently in place to support document management processes, e.g. an internally developed records management system called "DOT", a commercial workflow software called eDocLink, and an archival management system called ICA-Atom. In this pilot a prioritization of existing platforms will be made to choose the ones that will be included in the pilot.

T2.5.7 Full scale pilot no. 7. – Access to databases

Task leader: National Archives of Hungary.

Supported by: Danish National Archives

NAH will extract structured content from an Oracle database with the tools developed by WP3. The pilot will examine the applicability of data-warehouse concepts in an archival environment in order to maintain both the

original structure and intellectual interpretability of ingested data. The working prototype for access will be a userfriendly web-based application based on the DIP specification of WP5.

Scope: Representation of not less than 2 databases of different sizes and complexities with restricted and open content.

Objects: Extract data from the EDRMS and the databases, create SIPs for structured and unstructured records using the ESSArch Tools, ingest the SIPs to the repository using the ESSArch Preservation Platform, for further evaluation.

Participants: National Archives of Hungary (digital archives), data provider

Resource plan: 6 person months for setting up the pilot (assisting the archivists and the data provider in preparing the transfer; setting up and configuring the IT infrastructure at NAH), carrying out the pilot (transfer, quality checking, metadata amendments, AIP creation), testing the results and reporting.

Position in the project: NAH will primarily implement and pilot the applicability of specifications and tools related to access (WP5 D5.3, D5.4). The pilot will also prove usability of specifications and tools for supporting ingest (WP3 D03.3) of archival records.

Resource plan: 7 person months (6 pm for National Archives of Slovenia 1 pm for DNA)

Preconditions: M03.3, M03.4, M04.2, M05.4, M05.6.

Timeframe: M25-M27: setting up the pilot sites; M28-M31: running the pilot; M32-M33: testing and reporting.

Contribution to the project outcome

Data owner: Prosecution Service of Hungary

Platform: DBExport Tool, Oracle APEX, development in Java