



# **DELIVERABLE**

Project Acronym: E-ARK

**Grant Agreement Number:** 620998

Project Title: European Archival Records and Knowledge Preservation

## **DELIVERABLE DETAILS**

DELIVERABLE REFERENCE NO.	2.5
DELIVERABLE TITLE	Recommended Practices and Final Public Report on Pilots
REVISION	1.0

	AUTHOR(S)
Name(s)	Organisation(s)
István Alföldi	National Archives of Hungary (NAH)
István Réthy	National Archives of Hungary (NAH)
	REVIEWER(S)
Name(s)	Organisation(s)
Andrew Wilson	University of Brighton (UoB)
Clive Billenness	Silversity of Brighton (505)

F	Project co-funded by the European Commission within the ICT Policy Support Programme											
	Dissemination Level											
Р	Public	Х										
С	Confidential, only for members of the Consortium and the Commission Services											

# **REVISION HISTORY AND STATEMENT OF ORIGINALITY**

## **Submitted Revisions History**

Revision No.	Date	Authors(s)	Organisation	Description
0.1	05/01/17	István Alföldi	NAH	Draft document structure
0.2	12/01/17	István Rethy	NAH	Extended document structure
0.3	19/01/17	István Rethy	NAH	First Draft
0.4 – 0.7	23/01/17	István Alföldi	NAH	Draft
0.8	26/01/17	István Rethy	NAH	Internal Review
0.9	30/1/17	István Alföldi	NAH	Final Review
1.0	7/2/17	István Alföldi	NAH	Final version

## Contributions are also acknowledged from

Anders Bo Nielsen (DNA)

Phillip Mike Tømmerholt (DNA)

Alex Thirifays (DNA)

Hans Fredrik Berg (NAN)

Terje Pettersen-Dahl (NAN)

Arne-Kristian Groven (NAN)

Tarvo Kärberg (NAE)

Karin Oolu (NAE)

Raivo Ruusalepp (EBA)

Ats Rand (EBA)

Gregor Završnik (NAS)

Boris Domajnko (NAS)

Joze Skofljanec (NAS)

Miguel Ferreira (KEEPS)

Zoltán Lux (NAH)

Mezei József (NAH)

Clive Billenness (UoB)

## Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise.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

## **Table of Contents**

EXECUTIVE SUMMARY	
PLANNING AND EXECUTING THE E-ARK PILOTS	4
PILOT PLANNING IN THE DESCRIPTION OF WORK (DOW)	4
PILOT PLANNING DURING THE PROJECT	4
PILOT PREPARATION	5
PILOT EXECUTION	11
PILOT EVALUATION	18
OVERVIEW OF THE E-ARK PILOTS	19
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	23
PILOTS 1 - SIP CREATION ON RELATIONAL DATABASES	23
Scenarios	25
Execution report	28
Changes to the original plans	30
Feedback report	30
Recommended practices and further recommendations	31
PILOTS 2 - SIP CREATION AND INGEST OF RECORDS	32
Scenarios	34
Execution report	37
Changes to the original plans	37
Feedback report	37
Recommended practices and further recommendations	38
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 SERVICE	64
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	
Scenarios	
Execution report	
Changes to the original plans	
Feedback report	
Recommended practices and further recommendations	
EXTERNAL EVALUATIONS	
ILOT EVALUATION	
PROJECT LEVEL PILOT SUCCESS EVALUATION	
PILOT AND SCENARIO LEVEL SUCCESS EVALUATION	
FEFRENCED DOCUMENTS	01

D2.5 Recommended Practices and Final Public Re	port on Pilots
--	----------------

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

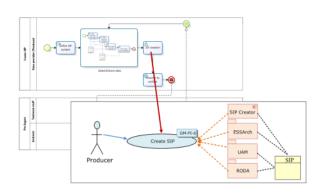
Pilots were planned to take place in the third year of the project when all tools and format specifications were ready to be tested, but pilot related activities started at very beginning and accompanied the tool development and format specification work throughout the project.

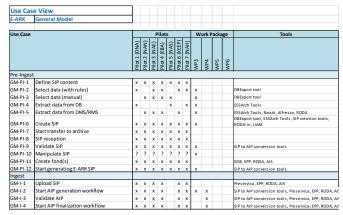
#### **General Model 1.0**

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.





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.

## **Pilot Card example**

Pilot #1	SI	P C	cre	ati	ior	10	n r	ela	tic	ona	al c	lat	ab	as	es														
Status	•	In In	stal stal	lati lati	ned on s on r cuti	tari	ly	ted		√ √ - -		OAI	S re	elev	vanc	e	Dra-Ingast	16-1118631	Ingest	F-ARK SIP		E-ARK AIP	Preservation		Data	Management	E-ARK DIP		Access
	•	Pi	lot (	exe	cuti	on (	com	plet	ed	-							Х												
Task leader	Da	nish	n Na	itioi	าal <i>เ</i>	Arcł	nive	S																					
Supported by	Ma	iger	nta																										
Contacts	Na	me	(Tit	le)											e-m	ail					Sk	уре							
Contact Person	Phi	illip	Mil	ke T	øm	mei	hol	t							pmt	t@s	a.dŀ	(			ph	illip	ton	nme	erho	lt_r	igsa	rkiv	et
Contact Person	An	der	s Bo	Ni	else	n									abn	@sa	a.dk	:											
Pilot staff members																													
Scope			•					sev							n to	ol w	ith i	not	less	thar	า 4 (	data	base	es o	f diff	ferer	nt siz	es a	ind
Object	fur	ther	eva	luat	ion																						Too		
Short description		_			•			reat SIP, i										lect	ed d	atak	oase	wit	h th	e Di	3ext	ract	tool	. Aft	er
Timeframe	•	8-M		urar	ice c	) I C	ucii.	JII , I	u ici	cube	JCK V	VIII	oc g	IVCI	110	7713													_
Preconditions	MO	3.3,	MO	3.4	(Do)	W)																							_
E-ARK tools					<u> </u>	Ĺ									I													ı	
	★ 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 Platform	HDFS-Storage	Catalogue (E-ARK web)	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)	<b>DBPTK</b>	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
Pilot Scenarios					•		•								•	•							•		•			•	
Scenario 1	Ext	rac	ting	rec	cord	ls fr	om	data	aba	se (	Data	a Se	t 1)	- d	atal	oase	wi	th n	o d	ocu	mei	nts							
Scenario 2	Extracting records from database (Data Set 1) - database with no documents  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								data														(lar	ge)						
Links		Process and use case information Pilot definition  Fest data specification 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:

Task leader Supported by Scope Object	Pilo dat arcl crit Pilo WP	ot wi ta, te thival teria ot re	h Na ill prosst the I use in the	ation ove ne in of (	that stru GIS c	l Are the ction	Chiv SIP ns (f is po	es and or th	DIP i	impl		ntat	ions	fulf	ill cr					. mbc	<b></b>			_				_					
Scope	Pilo dat arcl crit Pilo WP pro	ot wi ta, te thival teria ot re	ill pr est th I use in th port	ove ne in of C ne D	that istru GIS c IP cc	t the ictioi data	SIP ns (f is pc	and or thossib	ne pr	•		ntat	ions	fulf	ill cr					. m. th.c	<b>.</b>												
·	dat arcl crit Pilo WP pro	ta, te hival teria ot re	est th I use in th port	ne in of C ne D	istru GIS d IP co	ictioi lata	ns (f is po	or th	ne pr	•		ntat	ions	fulf	iller		٠.			. m.b.c	£	. 1	anish National Archives										
Object	WP pro	6 & '		with		ilot will prove that the SIP and DIP implementations fulfill specific requirements for the records containing GIS ata, test the instructions (for the producer and for the archive) regarding all phases of ingest, to prove that the rchival use of GIS data is possible (via open data method, direct access in the archives and use GIS data as search riteria in the DIP contents).																											
	Dur	Filot report with recommendations about urgent improvements and possible future improvements support for VP6 & 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	arcl	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	M2.	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	Nar	me (	Title	e)							E-m	ail									Skyp	ре											
Contact Person	Gre	egor	Zavr	šnik	: ()						gre	gor.z	avrs	nik(	ωgον	v.si					gregor.zavrsnik												
Pilot staff member	Ale	enka	Star	man	ı ()						aler	ıka.s	tarn	nan(	@go	v.si																	
Pilot staff member	Joz	e Sk	oflja	nec	()						joze	.skc	fljar	nec@	وgoر	/.si																	
OAIS Relevance			Pre	e-Ing	<b>gest</b> E-ARK	SIP	х	-	<b>est -</b> E-ARK	Stoi	age							St	orag	e - A	Acces	SS			E-	-ARK	DIP	х					
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	X ESSArch Tool Producer (ETP)	X ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	X ESSArch Preservation Platform	HDFS-Storage	X ICA-AtoM Catalogue	X OMT - Search and Dsiplay GUI	X Order Submission Service	X OMT - Order Management Tool	X Lily - Ingest	X ESSArch Preservation Platform	X E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	X IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	× QGIS	X Geoserver	× Peripleo	Orade (OLAP Viwer)					

Scenario 2	Sea	arch	and	l Ac	cess	inf	orm	atio	n u	sing	Gea	idot	a															
Description			ate DIP from AIP containing record with Geodata. Present Geodata information with QGIS along with cent and metadata from DIP.																									
	A d	lata	a object containing geodata can be identified by using search criteria as specified by E-ARK Tool																									
			rement specification. Selected data objects are selected and order is issued. DIP is prepared according																									
			der specification and end user credentials. DIP file structure with file descriptions (mime type, short																									
		•	cription) is presented to the enduser. Geodata from the order can be accessed in the designated viewer																									
			IS). The user checks authenticity of the DIP by accessing PREMIS documentation. Access to DIP is																									
	dod	cumented and captured metadata can be exported.																										
OAIS Relevance		Pre-Ingest Ingest - Storage Storage - Access																										
				Е	-ARK	SIP		Е	-ARK	AIP	Х		1						1						E	-ARK	DIP	Х
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 Platform	HDFS-Storage	X ICA-AtoM Catalogue	X OMT - Search and Dsiplay GUI	X Order Submission Service	X OMT - Order Management Tool	X Lily - Ingest	X ESSArch Preservation Platform	X E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	X IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	X QGIS	X Geoserver	X Peripleo	Oracle (OLAP Viwer)
Use-case	Acc	Access geodata via QGIS																										
Note	_		<u> </u>					ita a	nd p	ores	ent g	geo	data	wit	h Q	GIS												$\neg$

Pilot 5	Pilot Data										
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 administrative units until 1994 exported from GURS  The Surveying and Mapping Authority of the Republic of Slovenia)									
Description	Records and metadata of map										
Data type	GML document with metadata	in XML forr	nat, ESRI Shapefile, csv								
Metadata format	ISO 19115 (INSPIRE)										
less	62 records (cca. 3MB)										
Data Set 2	Records and metadata of Natu	ıra 2000 area	s, exported from ARSO								
Description	Records and metadata of map	s with Geod	ata								
Data type	GML document with metadata	in XML forr	nat								
Metadata format	ISO 19115 (or INSPIRE)										
less	1209 records (cca. 10 MB)										

OAIS Process					Pre-Ingest				
					Fre-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		Ī		QGIS			ESS Arch ETP		ESS Arch ETP
Used local tools			Existing archival system	Producers tools	Producer tools, open convesion tools	MS Excel, Inspire Metadata Creator			
Perfomer (actor)		Producer + Archivist + Technical Specialist	Producer	Producer	Producer	Producer	Producer		Producer
Prelemineries and Start condition	n	Official archival records 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, SIF
Output		Submission Agreement	Data selection list	Extracted data	Additional Data and documentation	Inspire.xml, MS excel w. metadata	E-ARK SIP		Submited SIP
Scenario 3	SIP Creation and Ingest of	of geodata in GML format							
Used E-ARK tool	,	ľ					ESS Arch ETP		ESS Arch ETP
Used local tools			Existing system	Producers tools	Producer tools	Producer GIS system, MS Excel			
Perfomer (actor)		Producer + Archivist + Technical Specialist	Producer	Producer	Producer	Producer	Producer		Producer
Prelemineries and Start condition	n	Official archival records 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, SIF
Output		Submission Agreement	Data selection list	Extracted data	Additional Data and documentation	Inspire.xml, MS excel w. metadata	E-ARK SIP		Submited SIP

Pilot 5	Pilot Preparation												
						Preparat	tion 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				
reparation tasks related to the software	from Software Component Matrix				Yes /		No needed / Configured /						
omponents	(for E-ARK tools)	from Scenarios sheet	from Processes sheets	Yes /No / (issue)	(planned date of availability)	Installed / (issues)	(issues)	Yes / (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				
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	r ???				
Component 8.	Search and display GUI	Scenario 2, 4	Access	No	No	Not installed	Need support form AIT	Further knowladge transfer	r 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	r Yes				
Component 10.	OMT	Scenario 2, 4	Access	No	No	Not installed	Need support form Magenta	Further knowladge transfer	r NO				
Component 11.	Archival Catalogue (EAD based)	Scenario 1, 2, 3, 4	Ingest, Access	No	No	Not installed	Need input	Further knowladge transfer	r NO				
Component 12.	Lilly	Scenario 1, 2, 3, 4	Ingest, Access	Yes	Yes / in 2/2 April	Not installed	Need support from AIT	Further knowladge transfer					
Component 13.	Geoserver	Scenario 2, 4	Access	Yes	Yes	Installed	None needed	Yes	Yes				
•													
				Prep	ration status	·							
Pilot dataset	Dataset #	Scenario	Data selected	Legal issues	Data available	Dataset ready for Pilot							
reparation tasks related to pilot data	from Pilot Data sheet	from Scenarios sheet	Yes / (issues)	None / (issue)	Yes / (planned date) / (issue)	Ready/(issue)							
lovenian Register of spatial units sele	ects Data set 1	1.	2 Yes	None	Yes	Yes							
latura 2000 dataset	Data set 2		4 Yes	None	Yes	Yes							
	_			Prepration statu	3								
nfrastructure	Scenario	Process	Element selected	Issues	Element ready for Pilot								
astractare	Scenario		Licinciit selecteu	sues	Licincine ready for Prior								
reparation tasks related to pilot													
nfrastructure	from Scenarios sheet	from Processes sheets	Yes / (issues)	None / (issue)	Ready / (issue)								
'irtual 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

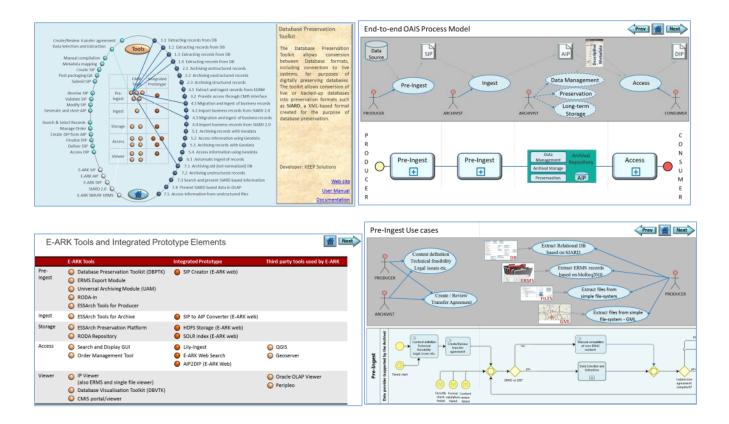
From the beginning of 2016 weekly progress meetings were held via a Webex teleconference service. The pilot representatives and staff members along with technical work package leads and some of the tool developers were regular members of these meetings.

## 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: <a href="http://eark-project.com/resources/general-model">http://eark-project.com/resources/general-model</a>

# 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	Status	Comment
	Completed		
Number and Title of pilot scenario	date	0 %	
	date	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	any info on tool execution that could be important to tool developers
practices	

## 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 at the end of the reporting period (Not started, Delayed, Started, Pending, Aborted, Completed)
duration of the process (only for Completed tasks)
reason for Pending / Aborted / Delayed status or any important comments at process step level
name of the task within the process step (each task must have a separate process step table, see sample on Pilot 7)
empty if detail fields are filled
or summary of tools if detail fields are empty (Manual, Local tool name)
tool name (Indicates if a tool is not developed by using E-ARK → "local")
(mandatory for E-ARK tools)
input summary
output summary
task actor (e.g. Archivist, IT specialist, Technical administrator, etc.)
any performance related info
all bugs, wishes, comments (that were entered in any of the lists provided by the tool developer)
any important info on tool execution
empty or "None" or "Not relevant"
empty if detail fields are filled
or summary of input data if detail fields are deleted
input data description
type of content
format of the metadata
volume of input data
further data manipulation activities (if any)
empty if detail fields are filled
or summary of input data if detail fields are deleted
output data description
type of content
format of the metadata
volume of output data
further data manipulation activities (if any)
further task-internal data manipulation (if any)
description of the data manipulation activities
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 → "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: <a href="http://eark-project.com/resources/project-deliverables/87-d24docs-p1-1">http://eark-project.com/resources/project-deliverables/87-d24docs-p1-1</a>

Part 2: <a href="http://eark-project.com/resources/project-deliverables/88-d24docs-p2-1">http://eark-project.com/resources/project-deliverables/88-d24docs-p2-1</a>

#### 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.

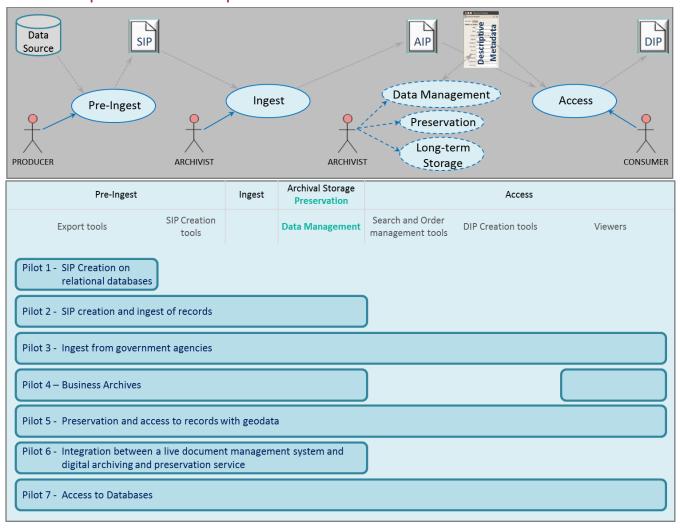
Some pilots were focusing on specific tools or processes of the OAIS models (1, 2, 4, 5, 6), others on archival and access of specific content types (4, 5, 7), one on automated ingest (6), and two pilots had scenarios to test the full spectrum of the OAIS processes along with the reference implementation: 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.

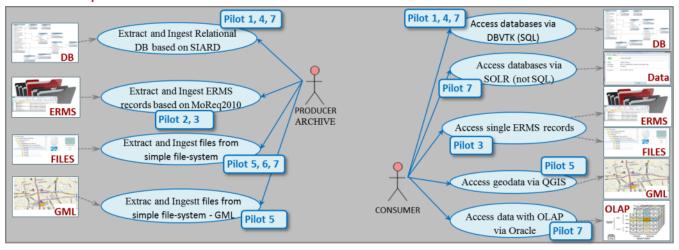
The following tables and graphs present the pilots and their relationships to other E-ARK elements. They help positioning the pilot scenarios on the OAIS map and among the various E-ARK tools and format specifications.

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

# Full-scale pilots and OAIS process

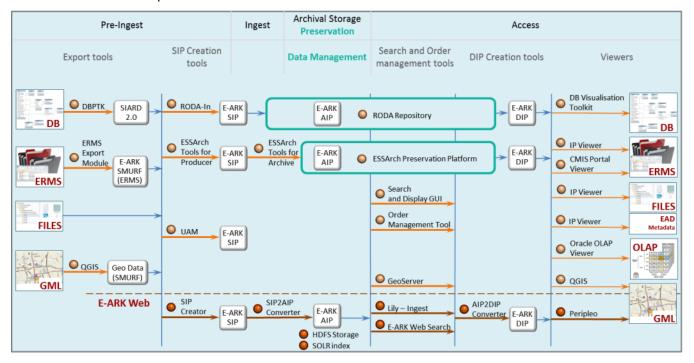


# Full-scale pilots and E-ARK uses-cases

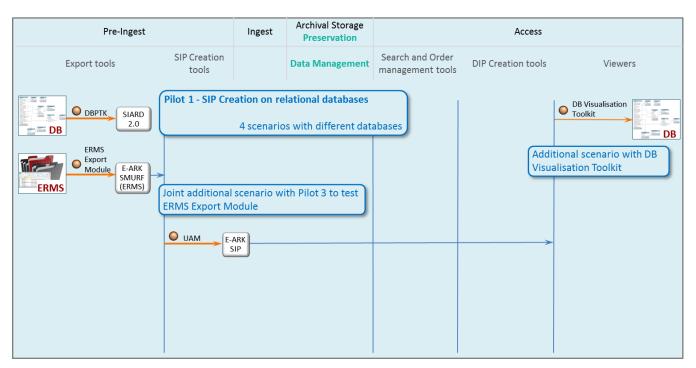


# Pilots using E-ARK tools and format specifications

### **E-ARK Tools and Format Specifications**



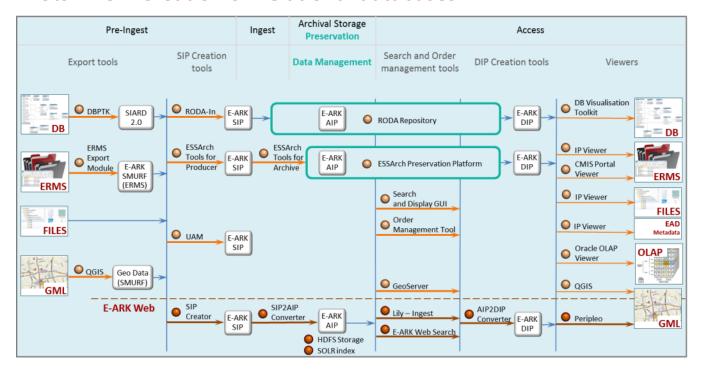
#### Pilot 1 - Danish National Archives

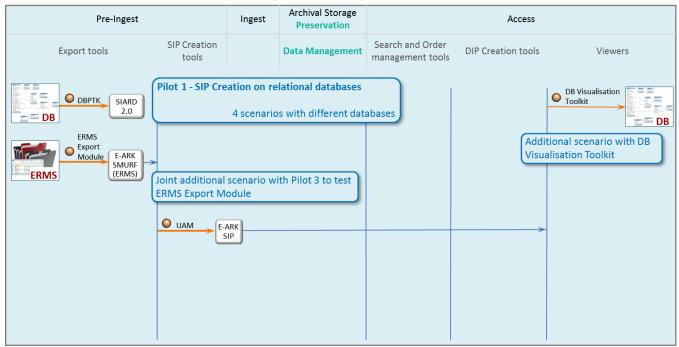


# Pilot report

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	Crea	tion	on re	latio	nal c	latak	ases	5															
Task leader	Dan	ish N	latio	nal A	rchiv	es																		
Supported by	Mag	genta	ı																					
Scope					Pilot (one								tool	with	not l	ess tl	nan 4	4 dat	abas	es of	diffe	erent	sizes	;
Object		ating uatio		for r	elatio	onal	datal	oases	s usir	ng th	e too	l cre	ated	in W	′P3, T	ГЗ.З:	SIP (	Creat	ion T	ools	for	furth	er	
Short description		The goal of the pilot is to make four successful data extractions from live authentic databases into the SIARD 2.0 format.																						
Contacts	Nan	ne (T	itle)							E-m	ail								Skyp	е				
Contact Person	And	ers B	o Ni	elser	1					<u>abn</u>	@sa.	<u>dk</u>												
Pilot staff member	Phillip Mike Tømmerholt									pmt	@sa.	dk							phill vet	iptoı	mme	rholt	_rigs	arki
OAIS Relevance			Pre-I	nges	t			Ing	est -	Stor	age						Sto	rage	– Ac	cess				
E-ARK Formats			E	-AR	SIP				Е	-ARK	AIP										Е	-ARK	DIP	
			S	IARD	2.0	Х			SMU	RF EI	RMS		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	geis	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
Companie 1	X			اديد	£	۰ - ام		- /5		-+ 4\	۰ ـ ام	- l		ul	ام									
Scenario 1					fron														`					
Scenario 2					fron			•									its (I	arge	)					
Scenario 3					fron												/1	-1						
Scenario 4					fron							apas	e wii	ın do	cum	ents	(ıarg	e)						
Additional scenario	<u> </u>				Data																			
Additional scenario	Extr	act r	ecor	ds wi	th EF	RMS	Expo	rt Mo	odule	e and	linge	st in	to Pr	eser	vica (	Joint	sce	nario	with	NAE	)			

# Scenarios

Extracting records from database (Data Set 1)																						
Extrac	ting	recor	ds fron	n dat	abase	e con	ntaini	ng no	o doc	ume	nts.											
Pre-In	gest	:																				
Extrac	t an	d Inge	st rela	tiona	l data	abase	e bas	ed o	n SIA	RD 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 about																						
90.000																						
Micro	Microsoft SQL Server 2008																					
Not relevant																						
small																						
	Pi	re-Inge	est			Ing	est -	Stora	age						Sto	rage	- Acc	ess				
		E-A	RK SIP				Е	-ARK	AIP		E-ARK				( DIP							
		SIAI	RD 2.0	Х		:	SMU	RF EF	RMS		SMURF SFSB					Geodata						
× Database Preservation Toolkit	ERMS Export Module	RODA-In	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
	Pre-In Extrac SIARD Datab Healt Datab 90.00 Micro Not re small	Extracting Pre-Ingest Extract an SIARD 2.0 Database Health sys Database 90.000 rec Microsoft Not releva small P	Extracting recor Pre-Ingest Extract and Inge SIARD 2.0 Database Preser Health system fi Database contai 90.000 records i Microsoft SQL S Not relevant small  Pre-Inge E-A  SIAI	Extracting records from Pre-Ingest Extract and Ingest related SIARD 2.0 Database Preservation Health system from The Database containing in 90.000 records in the result of the Pre-Ingest SIARD 2.0  Pre-Ingest Pre-Ingest E-ARK SIP SIARD 2.0  Universal Archiving Module Module Pre-Ingest	Extracting records from dat Pre-Ingest Extract and Ingest relationa SIARD 2.0  Database Preservation Tool Health system from The Da Database containing inform 90.000 records in the main Microsoft SQL Server 2008 Not relevant small  Pre-Ingest E-ARK SIP SIARD 2.0  X  SIARD 2.0  X  A (a)  SIARD 2.0  A (b)  SIARD 2.0  A (c)  SIARD 2.0  SIARD	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  (Label Module (ETA) SIARD 2.0  SIARD 2.0  SIARD 2.0  ESSArch Tools for Archiving Module (ETA)  SIARD 2.0  SIA	Extracting records from database correlations of the part of the p	Extracting records from database containing Pre-Ingest  Extract and Ingest relational database base SIARD 2.0  Database Preservation Toolkit  Health system from The Danish National States and Ingest relational database base Database containing information from responding informatio	Extracting records from database containing not pre-Ingest  Extract and Ingest relational database based of SIARD 2.0  Database Preservation Toolkit  Health system from The Danish National Serum Database containing information from reporte 90.000 records in the main table.  Microsoft SQL Server 2008  Not relevant small  Pre-Ingest  E-ARK SIP  SIARD 2.0  X  SMURF ER  SIARD 2.0  X  SMURF ER  SIARD 2.0  E-ARK Mep)  E-ARK Mep)  BY  SIARD 2.0  SIARD 2.0  SMURF ER  SMURF ER  SIARD 2.0  SMURF ER  SMURF ER  SIARD 2.0  SMURF ER  SMURF	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 info 90.000 records in the main table.  Microsoft SQL Server 2008  Not relevant  Small  Pre-Ingest  E-ARK SIP  SIARD 2.0  SIARD 2.0  SMURF ERMS  SIARD 2.0  SMURF ERMS  B-ARK AIP  SIARD 2.0  SMURF ERMS  B-ARK AIP  SIARD 2.0  SMURF ERMS  B-ARK AIP  SIARD 2.0  SMURF ERMS	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  SMURF ERMS  BIARD 2.0  SIARD 2.0  SMURF ERMS  BIARD 2.0  SIARD 2.0  SMURF ERMS  SIARD 2.0  SMURF ERMS  SMURF ERMS  SOLK Index  SOLK IND	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  SMURF ERMS  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  KODA Repository  F-ARK SIP  SIARD 2.0  SIA	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  Silard 2.0  Silard 2.0  Silard 2.0  Silard 2.0  Silard 3.0  Silard 2.0  Silard 2.0  Silard 3.0  Silard 3.0  Silard 4.0  Silard 4.0  Silard 5.0  Silard 6.0  Silard 6.	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 90.000 records in the main table.  Microsoft SQL Server 2008  Not relevant  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  SOUR Index  F-ARK AIP  SIARD 2.0  SOUR Index  SOUR Index  Sourch and Display Gul  Order Management Tool  Order Management Tool  Indest  Sourch Tool  Sourch Tool	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  SIARD 2.0  SIARD	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  E-ARK SIP  SIARD 2.0  X  SMURF ERMS  SMURF ERMS  SMURF SFSB  E-ARK AIP  SIARD 2.0  SIARD 2.0  SMURF ERMS  SMURF ERMS  SMURF SFSB  Geoseuver  Geoseuv	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 lever 90.000 records in the main table.  Microsoft SQL Server 2008  Not relevant small  Pre-Ingest  SIARD 2.0  SIARD 2.0	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  SIARD 2.0  E-ARK SIP  SIARD 2.0  E-ARK SIP  SIARD 2.0  E-ARK Mep  SIARD 2.0  E-ARK Mep  SIARD 2.0  E-ARK Mep  SIARD 2.0  E-ARK Mep  SIARD 2.0  BE-ARK Mep  SIARD 2.0  SMURF SFSB  SMURF SFSB  SMURF SFSB  Alabability - Indeest  Geoserver  GOGIS  E-ARK Mep  Alabability - Indeest  Geoserver  GOGIS  BE-ARK Mep  SIARD 2.0  Alabability - Indeest  GE-ARK Mep  SIARD 2.0  Alabability - Indeest  BE-ARK Mep  SMURF SFSB  SMURF SFSB  SMURF SFSB  Alabability - Indeest  Geoserver  GOGIS  BE-ARK Mep  SWURF SFSB  Alabability - Indeest  GOGIS  BE-ARK Mep  Alabability - Indeest  GOGIS  BE-ARK Mep  SWURF SFSB  SMURF SFSB  SMUR	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  E-ARK SIP  SIARD 2.0  E-ARK Mep  SIARD 2.0  SIARD 2.0	Extracting records from database containing no documents.  Pre-Ingest  Extract and Ingest relational database based on 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  SIARD 2.0  Table (E-ARK NIP)  SIARD 2.0  SIARD 2.0  Table (E-ARK NIP)  Table (	Extracting records from database containing no documents.  Pre-Ingest  Extract and Ingest relational database based on 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  E-ARK SIP  E-ARK SIP  E-ARK SIP  E-ARK SIARD 2.0  Tight - Ingest - Storage  E-ARK SIP  SIARD 2.0  Tight - Ingest - Storage  E-ARK SIP  E-ARK Mep  Database Access  E-ARK Mep  Database Access  E-ARK SIP  E-ARK Mep  Database Access  E-ARK SIP  E-ARK Mep  Database Access  E-ARK Mep  Databa	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  E-ARK SIP  SIARD 2.0  Datapase Access  E-ARK SIP  SIARD 2.0  SIARD 3.0  SIARD 3.0  SIARD 3.0  SIARD 4.0  SIARD 3.0  SIARD 4.0  SIA

Scenario 2	Extracting records from	ı dat	abase (Data Set 2)							
Description	Extracting records from	dat	abase containing no docu	ıme	nts.					
OIAS relevance	Pre-Ingest	e-Ingest								
Use-case	Extract and Ingest relati	ktract and Ingest relational database based on SIARD 2.0								
E-ARK specifications	SIARD 2.0	ARD 2.0								
E-ARK Tools	Database Preservation	atabase Preservation Toolkit								
Data	Registry of Cultural Events from Kultunaut Aps									
Description	Database from the commercial company Kultunat Aps, which holds information about cultural events at a									
	national level, from eve	nts a	arranged by local commu	ıniti	es to cultural events fro	m th	e Danish cultural			
	institutions. The databa	se c	ontains more than 5 milli	ion r	records.					
Data type	MySQL									
Metadata format	Not relevant									
Quantity	large									
OAIS Relevance	Pre-Ingest Ingest - Storage Storage - Access									
E-ARK Format	E-ARK SIP	RK SIP E-ARK AIP E-ARK				E-ARK DIP				
specifications	SIARD 2.0	SIARD 2.0 X SMURF ERMS SMURF SFSB				Geodata				

ERMS Export Module  ERMS Export Module  RODA-In  ESSArch Tool for Producer (E  Universal Archiving Module  SIP creator (E-ARK Web)  ESSArch Tools for Archive (ET  SIP2AIP (E-ARK Web)  RODA Repository  ESSArch Preservation Platfor  HDF5-Storage  SOLR Index  Search and Display GUI  Order Management Tool  Lily - Ingest  Geoserver  QGIS  E-ARK Web Search  AIP2DIP (E-ARK Web)  Database Visualization Toolk  IP Viewer  Peripleo  Oracle (OLAP Viewer)  CMIS portal/viewer		E-ARK Tools
Arch Tool for Producer Versal Archiving Modul Creator (E-ARK Web) Arch Tools for Archive Creator (E-ARK Web) Arch Tools for Archive CAIP (E-ARK Web) DA Repository Arch Preservation Platt FS-Storage IR Index Irch and Display GUI Ier Management Tool Ier Management Tool Ier Management Tool Is Server IS RK Web Search Cable (E-ARK Web) Cab	Х	atabase Preservation To
Arch Tool for Producer versal Archiving Modul creator (E-ARK Web) Arch Tools for Archive- ZAIP (E-ARK Web) DA Repository Arch Preservation Platt FS-Storage R Index Index Index Index RR Undex Server Ingest Server Is RK Web Search ZDIP (E-ARK Web) abase Visualization Tool ilewer ipleo Is portal/viewer		RMS Export Modul
versal Archiving Modul creator (E-ARK Web) Arch Tools for Archive- ZAIP (E-ARK Web) DA Repository Arch Preservation Platf FS-Storage -R Index rch and Display GUI ler Management Tool -Ingest DServer IS RK Web Search 2DIP (E-ARK Web) :abase Visualization Too riewer ipleo icle (OLAP Viewer) IS portal/viewer		
creator (E-ARK Web) Arch Tools for Archive Arch Tools for Archive 2AIP (E-ARK Web) DA Repository Arch Preservation Platt F5-Storage -R Index Irch and Display GUI ler Management Tool -Ingest Sserver IS RK Web Search 2DIP (E-ARK Web) abase Visualization Too ilewer ipleo Icle (OLAP Viewer) IS portal/viewer		SSArch Tool for Producer (
Arch Tools for Archive Arch Tools for Archive ZAIP (E-ARK Web) DA Repository Arch Preservation Platt F5-Storage -R Index Irch and Display GUI ler Management Tool -Ingest Sserver IS RK Web Search ZDIP (E-ARK Web) abase Visualization Tool iewer ipleo Icle (OLAP Viewer) IS portal/viewer		niversal Archiving Modu
Arch Tools for Archive  ZAIP (E-ARK Web)  DA Repository  Arch Preservation Platt  FS-Storage  R Index  rch and Display GUI  ler Management Tool  ler Management Tool  - Ingest  Sserver  IS  RK Web Search  2DIP (E-ARK Web)  abase Visualization Too  fiewer  ipleo  icle (OLAP Viewer)  IS portal/viewer		IP creator (E-ARK W
DA Repository Arch Preservation Plat Arch Preservation Plat FS-Storage IR Index Irch and Display GUI Ier Management Tool Ier Management Tool Server IS RK Web Search ZDIP (E-ARK Web) abase Visualization To iewer ipleo Icle (OLAP Viewer) IS portal/viewer		Tools for Archiv
DA Repository Arch Preservation Plat F5-Storage -R Index Irch and Display GUI ler Management Tool -Ingest Sserver IS RK Web Search SDIP (E-ARK Web) abase Visualization To iewer ipleo Icle (OLAP Viewer) IS portal/viewer		IP2AIP (E-ARK W
Arch Preservation Plat FS-Storage R Index rch and Display GUI ler Management Tool - Ingest DServer IS RK Web Search 2DIP (E-ARK Web) abase Visualization To fiewer ipleo Isle (OLAP Viewer) IS portal/viewer		ODA Reposit
FS-Storage  R Index Irch and Display GUI ler Management Tool - Ingest DServer IS RK Web Search 2DIP (E-ARK Web) abase Visualization To iewer ipleo Icle (OLAP Viewer) IS portal/viewer		SSArch Preservation Platf
Irch and Display GUI Ier Management Tool Ingest Sserver IS RK Web Search 2DIP (E-ARK Web) abase Visualization To fiewer ipleo Is portal/viewer		
ler Management Tool - Ingest - Ingest - Ingest - Ingest - Ingest - Is - Ingest - Is		OLR Ind
ler Management Tool - Ingest server IS RK Web Search 2DIP (E-ARK Web) abase Visualization To //iewer ipleo icle (OLAP Viewer) IS portal/viewer		earch and Display
- Ingest sserver IS RK Web Search 2DIP (E-ARK Web) :abase Visualization To flewer ipleo icle (OLAP Viewer) IS portal/viewer		rder Management Too
IS RK Web Search 2DIP (E-ARK Web) tabase Visualization To Viewer ipleo scle (OLAP Viewer)		
ARK Web Search IP2DIP (E-ARK Web) atabase Visualization To Viewer eripleo racle (OLAP Viewer) MIS portal/viewer		osen
-ARK Web Search IP2DIP (E-ARK Web) atabase Visualization To viewer viewer eripleo rracle (OLAP Viewer) MIS portal/viewer		QGIS
atabase Visualization To viewer  ripleo racle (OLAP Viewer)		-ARK Web Sea
atabase Visualization To Viewer eripleo racle (OLAP Viewer) MIS portal/viewer		IP2DIP (E-ARK
viewer eripleo racle (OLAP Viewe MIS portal/viewer		atabase Visualization
eripleo racle (OLAP Viewe MIS portal/viewer		Viewe
racle (OLAP Viewe		Peripleo
MIS portal/view		racle (OLAP Viewe
		MIS portal/view

Scenario 3	Extr	actin	ng red	cords	fron	n da	tabas	se (D	ata S	et 3)														
Description	Extr	actin	g rec	ords	fron	n dat	abas	e cor	ntaini	ng d	ocum	ents												
	The	DNA	will	go to	the	prod	lucer	's site	e wit	h the	tool	on a	USB	. The	DNA	will	toge	ther	with	the	prod	ucer	use t	he
	tool	and	mak	e exti	ractio	ons i	nto t	wo fo	orma	ts: SI	ARDI	OK ar	nd SI	ARD2	.0.									
OIAS relevance	Pre-	Inge	st																					
Use-case	Extr	act a	ct and Ingest relational database based on SIARD 2.0																					
E-ARK specifications	SIAF	RD 2.	0 2.0																					
E-ARK Tools	Data	abase	e Preservation Toolkit																					
Data	Adn	ninist	rativ	e sys	ystem from The Danish National Archives																			
Description	Data	abase	e con	tainii	ng information about all incoming scientific research data, and public deliveries of research								h											
	data	a. Dat	tabas	e coi	ntaining BLOBs/documents. Size 131 gigabyte.																			
Data type	Mic	osoft SQL Server 2008																						
Metadata format	Not	t relevant																						
Quantity	sma	II																						
OAIS Relevance		I	Pre-li	ngest	:			Ing	est -	Stora	age						Sto	rage	- Acc	ess				
E-ARK Format			Е	-ARK	SIP			E-ARK AIP					E-ARK DIP											
specifications			S	IARD	2.0	Х	SMURF ERMS				SMURF SFSB Geoda						data							
E-ARK Tools	kit			ЕТР)			:TA)			E										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
	Х						İ			İ	İ													i

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  Metadata format  Quantity  OAIS Relevance  E-ARK Format	-	relev	QL Server 2008  elevant  Pre-Ingest Ingest - Storage Storage - Access  E-ARK SIP E-ARK DIP																					
specifications		SIARD 2.0 X							SMU	RF EF	RMS		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

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				ds wi												-					-			
Description	NAE	was	supp	osec	toι	ıse tl	ne ER	MS I	Ехро	rt Mo	dule	to e	xpor	t reco	ords 1	from	ERM	1S bu	t bec	ause	of th	ne lat	e	
	dep	loym	ent c	of the	too	l nae	had	to u	se a l	ocal	expo	rt to	ol to	com	plete	the	full-s	cale	pilot	. To t	est t	he EF	RMS	
	Expo	ort N	1odul	le a jo	oint a	addit	ional	scer	nario	has l	oeen	exec	uted	l. DN	A exp	orte	d the	erec	ords 1	from	Alfre	sco	ERM:	3
	with	the	new	ly de <sub>l</sub>	ploye	ed ER	MS E	хроі	rt Mo	dule	and	sent	the S	SMU	RF EF	RMS 1	ile to	NAI	E whe	ere a	SIP v	vas c	reate	≗d,
	and	inge	sted	to Pr	eser	vica.	With	this	addi	tiona	l sce	nario	eve	ry ste	p th	at wa	as ori	igina	lly pla	anne	d to l	oe te	sted	in
	Pilot	t 3 ha	3 has been successfully tested.																					
OIAS relevance			ngest, Ingest																					
Use-case	Extr	act a	act and Ingest ERMS records based on MoReq2010																					
E-ARK specifications	SMU	JRF E	RF ERMS																					
E-ARK Tools	ERIV	1S Ex	Export Module																					
Data	ERIV	1S system of The Danish School of Media and Journalism (Danmarks Medie- og Journalisthøjskole) (DMJX)																						
Description	Diffe	fferent kinds of letters and documents																						
Data type	Reco	ecords from Alfresco ERMS																						
Metadata format	EAD	AD																						
Quantity	121	files	, 17 N	ИΒ																				
OAIS Relevance			Pre-li	ngest	:			Ing	est -	Stora	age						Sto	rage	- Acc	ess				
E-ARK Format			Е	-ARK	SIP				Е	-ARK	AIP										E-ARK DIP			
specifications			S	IARD	2.0				SMU	RF EI	RMS	Х			SMU	JRF S	FSB					Geo	data	
E-ARK Tools	_			(d.			8			_										4				
	olkit			. (ET	le		(ET/			forn										olki				
	υTo			ncer	npo	Q	ive			Plat			_	00						n T0				
	atio	ule		rod	МB	We	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 GGIS E-ARK Web Search AIP2DIP (E-ARK Web) Database Visualization Toolkit IP Viewer Peripleo					ver)	-											
	erve	Mod		or P	ivin	ARK	for	(E-ARK Web)	ory	rvat			pla	me				arch	≥  >	aliza			/iev	ewe
	Pres	ort l		oolf	Arch	, (E-	sloc	-AR	osit	ese	age	×	Σ	Jage	+			b Se	-AR	Visu			AP \	a ×
	Database Preservation Toolkit	ERMS Export Module	드	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	h T	Р (Е	RODA Repository	h Pı	HDFS-Storage	SOLR Index	anc	Mar	Lily - Ingest	rver		E-ARK Web Search	AIP2DIP (E-ARK Web)	ase 1	ver	90	Oracle (OLAP Viewer)	CMIS portal/viewer
	tab	MS	RODA-In	SArc	iver	cre	SArc	SIP2AIP	DA	SArc	FS-(	LR I	arch	der	y - Ir	Geoserver	QGIS	1 RK	P2DI	tab	IP Viewer	Peripleo	acle	AIS F
	Da	ER	RO	ES:	,	SIF	ES	SIF	RO	ES	모	SO	Se	ō	É	Ge	ğ	F-7	₹		В	Pe	ō	5
																				X				<u> </u>

# **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	SIARD2.0:
(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 "eCH-0165 SIARD Format Specification 2.0". There is no automatic tool for this yet.  SIARDDK:  100% extraction of all tables and their data. The DNA has validated against "Executive Order on Submission Information Packages" and found no errors in the product.

Extracting records from database (Data Set 2) - database with no documents (large)	June 2016	September 2016	SIARD2.0:  100% extraction of all tables and their data. The DNA has manually validated the SIARD-package up against the "eCH-0165 SIARD Format Specification 2.0". There is no automatic tool for this yet.  SQL Server:  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.  SIARDDK:
			100% extraction of all tables and their data. The DNA has validated against "Executive Order on Submission Information Packages" and found no errors in the product.
3. Extracting records from database (Data Set 3) - database with documents	July 2016	September 2016	SIARD2.0:  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 "eCH-0165 SIARD Format Specification 2.0"  SIARDDK:  100% extraction of all tables and their data. The DNA has validated against "Executive Order on Submission Information Packages" and found no errors in the end product.
4. Extracting records from database (Data Set 4) - database with documents (large)	August 2016	September 2016	SIARD2.0:  100% extraction of all tables and their data. The DNA has manually validated the SIARD-package up against the "eCH-0165 SIARD Format Specification 2.0". There is no automatic tool for this yet.  SIARDDK:  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	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 Additional scenario	Experiments with Database Visualization Toolkit
Data (input / output)	Movie database
Performance	Good
Issues	No issues found
Wishes	Users recommend showing technical information about the database on a separate page.
Comments	
Experiences and recommended	

practices	
p	

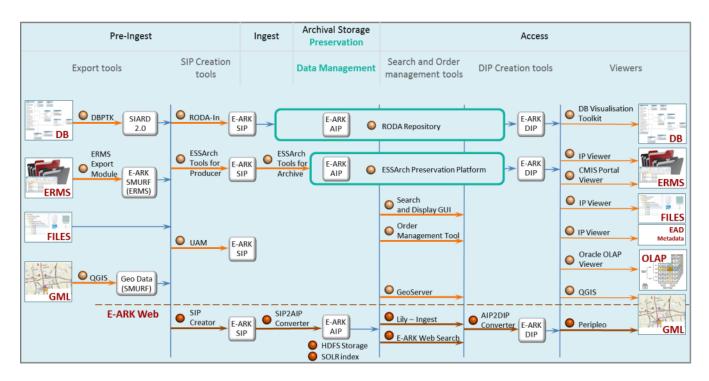
E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
ERMS Export Module	
Used in Additional scenario	Extract records with ERMS Export Module and ingest into Preservica
	(Joint scenario with NAE)
Data (input / output)	ERMS system of The Danish School of Media and Journalism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

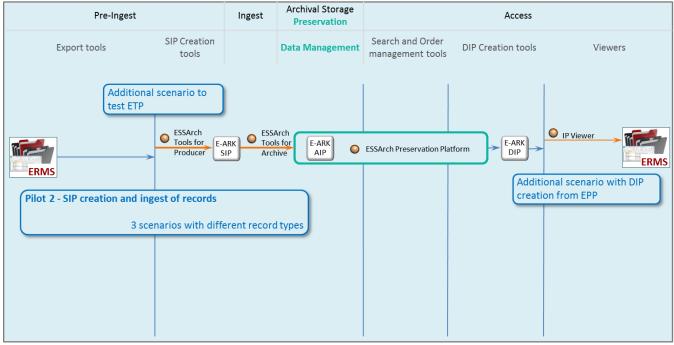
Recommended practices and further recommendations

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	ion a	and i	nges	t of r	ecor	ds																
Task leader	Nati	ional	Arch	nives	of N	orwa	У																	
Supported by	ESS	Solu	tions																					
Scope	Not	less	than	2 tra	nsfe	rs of	unst	ructı	ured	reco	rds w	ith r	nixec	rest	ricte	d an	d un	restr	icted	mat	erial,	and	not	less
					struc																			
Object													r stru									_	SArcl	h
			_						-				reser											
Short description		ne main part of the pilot includes the export of electronic records and their metadata from EDRM systems																						
_		nd databases of Norwegian public sector institutions, transfer and ingest them to the NAN digital repository.																						
Contacts	Nan	Iame (Title) E-mail Skype																						
Contact Person				Grov									verk											
Pilot staff member				en-D	ahl					_			<u>/verk</u>		<u>)</u>									
Pilot staff member		Hau	_										erket											
Pilot staff member	Jørg	en Ø	. Vik	-Stra	ndli					jorv	ik@a	rkiv	<u>rerke</u>	t.no										
OAIS Relevance		l	Pre-l	nges	t			Ing	est -	Stor	age						Sto	rage	– Ac	cess				1
E-ARK Formats			E	-AR	SIP	Х			E	-ARK	AIP	Х									E	-ARK	DIP	Х
			S	SIARD	2.0				SMU	RF EI	RMS	X			SMI	JRF S	SFSB					Geo	data	
E-ARK Tools				_			_																	
	ķi			ET			ETA			Platform										kit				
	00			er (	anle		) e/			attc				_						00				
	Database Preservation Toolkit	a)		ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	<u> </u>		n P			5	Order Management Tool					=	Database Visualization Toolkit			-	
	ati	ERMS Export Module		Pro	ng l	<b>&gt;</b>	. Ar	SIP2AIP (E-ARK Web)		ESSArch Preservation			Search and Display GUI	ant				Ę	AIP2DIP (E-ARK Web)	zati			Oracle (OLAP Viewer)	/er
	serv	Mo		for	hivi	-AR	for	× ∨	tory	ıva			sple	eme				arc	×	ıaliz			Vie	iev
	Pre	ort		loc	Arcl	r (Ē	sloc	Α̈́	osit	rese	age	×	Θ	nag	يد			b Se	-AR	Visu			ΑP	al/v
	se	Exp	<u>_</u>	h T	sal,	ato	h T	(E	Rep	h Pı	tor	nde	an	Mai	sagı	ver		Wel	P (E	se	ver	o	0	ort
	abe	NS I	RODA-In	Arc	ver	cre	Arc	ZAII	RODA Repository	Arc	HDFS-Storage	SOLR Index	rch	ler	Lily - Ingest	Geoserver	<u>s</u>	E-ARK Web Search	2DI	abe	IP Viewer	Peripleo	cle	CMIS portal/viewer
	Dat	ERI	RO	ESS	Uni	SIP	ESS	SIP	RO	ESS	H	SOI	Sea	ő	Lily	Geo	QGIS	E-A	AP	Dat	ē	Per	Ora	S
				Х			Х			Х											Х			
Scenario 1	SIP Creation and Ingest of unstructured records (Data Set 1)																							
Scenario 2	SIP Creation and Ingest of unstructured records (Data Set 2)																							
Scenario 3	SIP (	Creat	tion a	and I	nges	t of s	truct	ured	reco	ords (	Data	Set	3)											
Additional scenario	Crea	ating	SIP v	with	ESSA	rch T	ool f	or Pr	odu	cer														
Additional scenario	Gen	enerating E-ARK DIP from ESSArch Preservation Platform																						

#### Scenarios

Scenario 1	SIP	Crea	tion a	and I	nges	t of ι	ınstr	uctu	red r	ecor	ds (D	ata S	et 1)	)										
Description	Extr	act u	ınstrı	ıctur	ed re	cord	s fro	m EC	RMS	base	ed on	the	Norv	vegia	ın NC	ARK	4 sta	anda	rd. C	reate	SIP	using		
	ESSA	Arch	Tools	s. Ing	est tl	he SI	P to t	the r	epos	itory	using	g ESS	Arch	Pres	erva	ion F	Platfo	orm,	for f	urthe	r eva	luati	on.	
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd In	gest	ERM	IS red	cords	(sim	ilar t	о Мо	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	), ESS	SArch	Pres	erva	tion	Platf	orm					
Data	Noa	rk 4	outp	ut fro	m El	DRM	S																	
Description	EDR	MS c	data f	rom	publ	ic pro	oduc	er co	nver	ted ir	nto N	oark	4 οι	ıtput	(real	proc	ductio	on da	ata)					
Data type	Noa	Noark 5 XML file, documents in PDF/A (or a few other specified formats), in TAR file																						
Metadata format	XMI	XML: METS, PREMIS, ADDML (local)																						
Quantity	20G	20GB																						
OAIS Relevance		Pre-Ingest Ingest - Storage Storage – Access																						
E-ARK Format			E	-ARK	SIP	Х			Е	-ARK	AIP	Х									Е	-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														

Scenario 2	SIP Creation and Ingest	of u	ınstructured records (D	ata S	et 2)								
Description	Extract unstructured red	cord	s from EDRMS based or	the	Norwegian NOARK 5 sta	anda	rd. Create SIP using						
	ESSArch Tools. Ingest th	e SI	P to the repository using	g ESS	Arch Preservation Platfo	orm,	for further evaluation						
OIAS relevance	Pre-Ingest, Ingest												
Use-case	Extract and Ingest ERMS records (similar to MoReq2010)												
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 EDRMS												
Description	EDRMS data public prod	duce	r converted into Noark	5 out	put (real production da	ta)							
Data type	Noark 5 XML file, docun	nent	s in PDF/A (or a few oth	er sp	pecified formats), in TAR	R file							
Metadata format	XML: METS, PREMIS, AD	DDM	L (local)										
Quantity	5 GB												
OAIS Relevance	Pre-Ingest		Ingest - Storage		Sto	rage	- Access						
E-ARK Format	E-ARK SIP	Х	E-ARK AIP	P X E-ARK DI									
specifications	SIARD 2.0		SMURF ERMS	Х	SMURF SFSB Geodata								

	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)
Х	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 3	SIP (	Creat	tion a	and I	nges	t of s	truct	tured	d rec	ords	(Data	Set	3)											
Description	Extr	act d	ata f	rom	old d	atab	ase c	utpu	ıt, cr	eate	SIPs 1	or st	ructi	ured	recoi	rds u	sing	ESSA	rch T	ools,	inge	st th	e SIP	s to
	the	repo	sitor	y usir	ng ES	SArc	h Pre	serv	ation	Plat	form	, for	furth	er ev	⁄alua¹	tion.								
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	act a	nd In	igest	ERM	IS red	cords	(sim	ilar t	о Мо	Req	2010	)											
E-ARK specifications	E-AF	RK-SI	Р																					
E-ARK Tools	ESSA	Arch	Tool	Prod	lucer	(ETP	), ES	SArch	n Toc	l Arc	hive	(ETA	), ESS	SArch	Pres	erva	tion	Platf	orm					
Data	Old	data	base	(CSV	<b>'</b> )																			
Description	The	data	set l	nere	is the	nat	ional	regi	stry c	of lice	encec	l hun	iters	cont	ainin	g dat	a fro	m th	e pei	riod 1	1985	-1999	Э.	
Data type	CSV	form	nat (i	nput)	), tar	file																		
Metadata format	XML	.: ME	TS, P	REM	IIS, A	DDM	IL (lo	cal)																
Quantity	Con	ontaining 338.500 registered persons. 105 MB																						
OAIS Relevance		Pre-Ingest Ingest – Storage Storage - Access																						
E-ARK Format			E	-ARK	SIP	Х			E	-ARK	AIP	Х									Е	-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	İ		Х			Х				İ					İ					ı

Please note that more details with screenshots on scenario execution are available in the deliverable <u>D2.4 Pilot 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	ı full-	scale	e scer	nario	s) in	the v	/irtua	ıl env	/ironi	ment													
OIAS relevance	Pre-	Inge	st																					
Use-case	Extr	act a	nd In	igest	ERM	S red	cords	s (sim	ilar t	о Мо	Req	2010	)											
E-ARK specifications	E-AF	RK-SI	Р																					
E-ARK Tools	ESSA	Arch	Tool	Prod	lucer	(ETP	')																	
Data																								
Description	Loca	Local test data																						
Data type	Microsoft and pdf documents																							
Metadata format	Not relevant																							
Quantity	small																							
OAIS Relevance	Pre-Ingest Ingest – Storage Storage – Access																							
E-ARK Format	E-ARK SIP X E-ARK AIP X E-ARK DIP																							
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			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

Additional scenario	Gen	erati	ing E	-ARK	DIP	from	ESS	Arch	Pres	erva	tion I	Platfo	orm											
Description	The	EssA	rch F	rese	rvati	on Pl	atfor	m (E	PP) i	s full	y E-A	RK co	ompa	atible	. In t	his a	dditio	onal	scen	ario a	an E-	ARK I	OIP is	5
	gene	erate	d fro	m EF	P. TI	ne sc	enari	io co	uld n	ot be	e yet	com	plete	d be	cause	of t	he st	rict I	Vorw	egiar	n dat	a har	ndlin	g
	regu	ılatio	ns m	nake i	it ver	y dif	ficult	to u	se ar	chive	d 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	SSArch Preservation Platform (EPP)																						
Data	Sele	elected archived data																						
Description	Diffe	Different kinds of letters and documents																						
Data type	Mic	Microsoft and pdf documents																						
Metadata format	Not	lot relevant																						
Quantity	sma	mall																						
OAIS Relevance		Pre-Ingest Ingest – Storage Storage – Access																						
E-ARK Format			Е	-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	data	
E-ARK Tools	kit			ЕТР)			(AT			Ē										kit				
	Database Preservation Toolkit			ESSArch Tool for Producer (ETP)	Universal Archiving Module	(q	ESSArch Tools for Archive (ETA)			ESSArch Preservation Platform			_	00						ר Toolkit				
	/atior	dule		Prod	ng M	creator (E-ARK Web)	Arch	Veb)		rtion			Search and Display GUI	Order Management Tool				ج	Veb)	Database Visualization			Oracle (OLAP Viewer)	Je.
	ser	ERMS Export Module RODA-In ESSArch Tool for Prod Universal Archiving IV SIP creator (E-ARK Ww. ESSArch Tools for Arc ESSArch Tools for Arc ESSArch Tools for Arc ESSArch Tools for Arc ESSArch Tools for Arc ESSArch Preservation HDFS-Storage SOLR Index Geoserver Geoserver Geoserver Geoserver AGIS E-ARK Web Search AIP2DIP (E-ARK Web) Database Visualizatio IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer																						
	e Pre e Pre con (																							
	bas	S Ex	A-In	rch	ersa	reat	rg	AIP	A Re	rg	S-Sto	lnd	ch a	Σ	Ing	erv		×	DIP	bas	ewe	oelo	) el	od s
	Data	ERM	RODA-In	ESSA	Univ	SIP c	ESSA	SIP2AIP	ROD,	ESSA	HDFS-Storage	SOLR Index	Sear	Orde	Lily - Ingest	Geoserver	QGIS	E-AR	AIP2	Data	IP Viewer	Peripleo	Orac	CMIS
										X											X			

#### **Execution report**

Scenario	Started	Completed	Summary
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.
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" full-scale 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

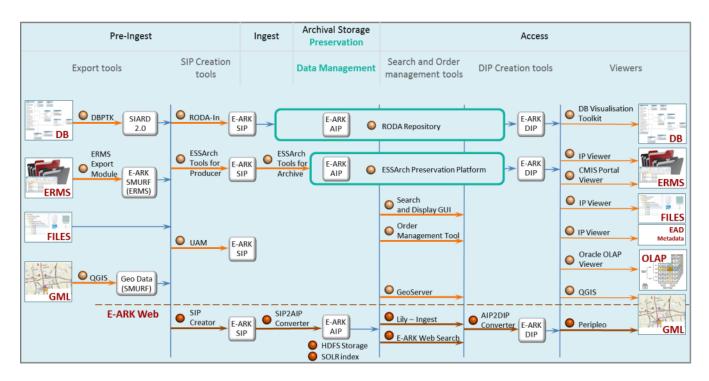
The following table contains the recommended practices and further development suggestions collected during pilot execution and evaluation.

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	ETA	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	ETP, ETA, EPP	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

#### D2.5 Recommended Practices and Final Public Report on 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	gove	nme	nt a	genci	ies																
Task leader	Nat	ional	Arch	nives	of Es	toni	a																	
Supported by																								
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																						
							_																	an
				_									_	al ED			he ag	gency	/); Th	e wh	ole s	et w	ill	
														cy of										
Object		itive EDRMS at a governmental agency (Alfresco DELTA), records preparation tool (UAM), digital																						
	-	eservation and access systems (Preservica, AIS)  e main part of the proposed pilot includes the export of electronic records and their metadata from EDRM																						
Short description																						from	EDF	₹M
		stems of Estonian public sector institutions, transfer and ingest to the NAE digital repository.  addition Estonian agencies have the responsibility to make public electronic records with no access																						
					_							-											المسائلة	
														he pi							_	tanc	iardiz	zea
Contacts	_			s me	tnou	s tria	t are	шр	ieme			ie ag	encie	es' di	gitai	IIIII d	Struc	ture	_		е			
Contacts	Ivan	ame (Title) E-mail Skype																						
Contact Person	Kar	(arin Oolu <u>karin.oolu@ttu.ee</u> karinoolu																						
Pilot staff member	Tar	Tarvo Kärberg tarvo.karberg@ra.ee tarvo.karberg																						
OAIS Relevance		Pre-Ingest Ingest - Storage Storage – Access																						
E-ARK Formats		E-ARK SIP X E-ARK AIP E-ARK DIP X																						
		SIARD 2.0 SMURF ERMS X SMURF SFSB Geodata																						
E-ARK Tools				)																				
	ķi			ETP			ETA			rr										kit				
	0			er (	) ji	_	) e/			atte				_						00				
	Database Preservation Toolkit	a		ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP creator (E-ARK Web)	ESSArch Tools for Archive (ETA)	2		ESSArch Preservation Platform			Ξ	Order Management Tool					-	Database Visualization Toolkit			5	
	/ati	ERMS Export Module		Pro	ng I	×	Ā	(E-ARK Web)		tio			Search and Display GUI	ent				ج	AIP2DIP (E-ARK Web)	zati			Oracle (OLAP Viewer)	/er
	Ser	Μ		for	hi.	-AR	ţo.	X V	RODA Repository	eva			spla	eme				E-ARK Web Search	× ×	ıali			Vie	CMIS portal/viewer
	Pres	Į,		o	Arc	r (E	soc	-AF	osit	ese	age	×	ΙŌΕ	Jago	٠,			p Se	-AR	Visu			AP	<u>₹</u>
	se	dx:	2	h Tc	sal/	atol	h Tc	(E	Зер	h Pr	tor	g	anc	Mar	ges	ver		Wel	P (E	se \	/er	0	0	ort
	apa	AS E	RODA-In	Arc	ver	cre	Arc	SIP2AIP	A	Arc	HDFS-Storage	SOLR Index	rch	er I	Lily - Ingest	Geoserver	<u>s</u>	R.	2DI	apa	IP Viewer	Peripleo	cle	IS p
	Dat	ER	RO	ESS	O.	SIP	ESS	SIP	80	ESS	무	SOI	Sea	Örc	Lily	Geo	QGIS	E-A	AIP	Dat	<u>a</u>	Per	Ora	S
		X			Х																			Х
Scenario 1	Extr	Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica (Data set 1)																						
Scenario 2	Pro	Provide access to records from governmental institution through RESTful services (Data set 1)																						
Scenario 3	Extr	Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica (Data set 2)																						
Scenario 4	Pro	Provide access to records from governmental institution through RESTful services (Data set 2)																						
Additional scenario	Extr	Extract records with ERMS Export Module and ingest into Preservica (Joint scenario with NAE)																						
Additional scenario	ERN	RMS Export Module scenario with local ERMS system DELTA																						

#### Scenarios

Scenario 1	Extr	act r	ecor	ds fro	om E	DRM	(of a	a gov	ernn	nenta	al ins	titut	ion),	crea	te SII	and	d inge	est to	o Pre	servi	са			
Description	Expo	xport public records from an EDRM system of a governmental agency, create SIP, and ingest to the Preservica																						
	syste	em a	t the	Nati	onal	Arch	ives	of Es	tonia	١.														
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extra	ract and Ingest ERMS records based on MoReq2010																						
	(Alfr	esco	is no	ot Mo	req-	com	plian	t sys	tem)															
E-ARK specifications	E-AF	ARK-SIP, SMURF																						
E-ARK Tools	Univ	niversal Archiving Module (UAM)																						
Data	Reco	ecords and metadata exported from native ERMS (DELTA) Export Module at Ministry of Justice of Estonia																						
Description	Data	ata set consists of different documents of Ministry of Justice from 6 series with different retention period.																						
Data type	ddo	doc, docx, PDF, TIFF																						
Metadata format	SMU	MURF ERMS																						
Quantity	15 fi	L5 files																						
OAIS Relevance		ı	Pre-li	ngest	:			Ing	est -	Stora	age						Sto	rage	– Acc	cess				
E-ARK Format			E	-ARK	SIP	Х			Е	-ARK	AIP										Е	-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	P Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
	Δ	ū	Ŗ	ŭ	X	S	ű	S	ď	Ü	I	Š	Š	0		9	ď	Ú	4	D	<u>-</u>	P	0	ō

Scenario 2	Provide access to records from governmental institution through RESTful services										
Description	Estonian agencies have	the	responsibility to make p	ublio	electronic records with	no a	access restrictions available				
	on their web sites, which	h m	eans that the pilot will a	ilso e	enable this through stand	lard	ized linking/access				
	methods that are implemented in the agencies' digital infrastructure / web site.										
OIAS relevance	Access										
Use-case	Access single ERMS reco	rds	via CMIS Browser								
	(To be consolidated with	To be consolidated with a CMIS interface access solution)									
E-ARK specifications	SMURF										
E-ARK Tools	CMIS Browser										
Data	Records and metadata e	хрс	orted from native ERMS	(DEL	TA) Export Module at Mi	nist	ry of Justice of Estonia				
Description	Data set consists of diffe	eren	t documents of Ministr	y of J	ustice from 6 series with	diff	erent retention period.				
Data type	ddoc, docx, PDF, TIFF										
Metadata format	SMURF ERMS										
Quantity	15 files										
OAIS Relevance	Pre-Ingest Ingest - Storage Storage - Access										
E-ARK Format	E-ARK SIP		E-ARK AIP				E-ARK DIP X				
specifications	SIARD 2.0	RD 2.0 SMURF ERMS X SMURF SFSB Ge									

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)  X CMIS portal/viewer	E-ARK Tools
ERMS Export Module  RODA-In  ESSArch Tool for Producer (ETP  Universal Archiving Module  SIP creator (E-ARK Web)  ESSArch Tools for Archive (ETA)  SIPZAIP (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  AIPZDIP (E-ARK Web)  Database Visualization Toolkit  IP Viewer  Peripleo  Oracle (OLAP Viewer)	atabase Preservation Too
ESSArch Tool for Producer (ETP Universal Archiving Module SIP creator (E-ARK Web) ESSArch Tools for Archive (ETA) SIPZAIP (E-ARK Web) RODA Repository ESSArch Preservation Platform HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolkit IP Viewer Peripleo Oracle (OLAP Viewer)	RMS Export Modul
ESSArch Tool for Producer (ETP  Universal Archiving Module  SIP creator (E-ARK Web)  ESSArch Tools for Archive (ETA)  SIPZAIP (E-ARK Web)  RODA Repository  ESSArch Preservation Platform  HDFS-Storage  SOLR Index  Search and Display GUI  Order Management Tool  Lily - Ingest  Geoserver  Geoserver  QGIS  E-ARK Web Search  AIPZDIP (E-ARK Web)  Database Visualization Toolkit  IP Viewer  Peripleo  Oracle (OLAP Viewer)	
Universal Archiving Module SIP creator (E-ARK Web) ESSArch Tools for Archive (ET. SIP2AIP (E-ARK Web) RODA Repository ESSArch Preservation Platforr 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 Toolki IP Viewer Peripleo Oracle (OLAP Viewer)	SSArch Tool for Producer (ETP
SIP creator (E-ARK Web)  ESSArch Tools for Archive (ET.  SIP2AIP (E-ARK Web)  RODA Repository ESSArch Preservation Platforr 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 Toolki IP Viewer Peripleo Oracle (OLAP Viewer)	niversal Archiving Modu
ESSArch Tools for Archive (ET SIPZAIP (E-ARK Web)  RODA Repository ESSArch Preservation Platforr HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolki IP Viewer Peripleo Oracle (OLAP Viewer)	IP creator (E-ARK W
SIPZAIP (E-ARK Web)  RODA Repository ESSArch Preservation Platfor HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	Tools for Archive (ET,
ESSArch Preservation Platfor HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	IP2AIP (E-ARK W
ESSArch Preservation Platfor HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	ODA Reposit
HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	SSArch Preservation Platf
SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	
Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	OLR Ind
Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	earch and Display
Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	rder Management To
Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	
eark Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	osen
E-ARK Web Search AIPZDIP (E-ARK Web) Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	QGIS
AIPZDIP (E-ARK Web)  Database Visualization Toolk IP Viewer  Peripleo Oracle (OLAP Viewer)	-ARK Web Sea
Database Visualization Toolk IP Viewer Peripleo Oracle (OLAP Viewer)	IP2DIP (E-ARK
IP Viewer Peripleo Oracle (OLAP Viewe	atabase Visualization Toolk
Peripleo Oracle (OLAP Viewe	Viewe
Oracle (OLAP Viewe	Peripleo
< CMIS portal/view	racle (OLAP Viewe
	MIS portal/view

Scenario 3	Extr	act r	ecor	ds fro	om E	DRM	(of a	gov	ernn	nenta	al ins	titut	ion),	crea	te SII	and	d ing	est to	o Pre	servi	ca	-		
Description	Expo	oport public records from an EDRM system of a governmental agency, create SIP, and ingest to the Preservica																						
	syst	rstem at the National Archives of Estonia.																						
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Extr	ract and Ingest ERMS records based on MoReq2010																						
	(Alfr	esco	is no	ot Mo	oreq-	com	plian	t sys	tem)															
E-ARK specifications	E-AF	RK-SIP, SMURF																						
E-ARK Tools	Univ	iversal Archiving Module (UAM)																						
Data	Reco	ecords and metadata exported from native ERMS (via DELTA) at Ministry of Justice of Estonia																						
Description	Data	ata set consists of different documents of Ministry of Justice from different series.																						
Data type	DDC	DOC (a file format holding Estonian digital signature information), DOCX, PDF, TIFF																						
Metadata format	SMU	MURF ERMS																						
Quantity	200	00 files																						
OAIS Relevance			Pre-li	ngest	:			Ing	est -	Stora	age						Sto	rage	– Acc	cess				
E-ARK Format			Е	-ARK	SIP	Х			Е	-ARK	AIP										Е	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geo	data	
E-ARK Tools	٠,			TP)			æ			Ε										ij				
	oolk			ır (E	nle		e (ET			tfor										Toolkit				
	on Te	•		duce	Mod	(eb)	chive	_		ı Pla			GUI	Tool					_					
	vatic	dule		Pro	ing [	×	r Ar	Web	_	atio			ay G	ent				ج	۷eb	zati			wer	ver
	ser	: Mo		for	chivi	E-AR	ls fo	(E-ARK Web)	itor	erve	a)		ldsi	gem				ear	RK.	uali			Vie	viev
	e Pre	port		Тоо	l Ar	or (I	T00	(E-A	sod	Pres	rag	ex	l pu	anag	est	-E		eb S	(E-A	e Vis	٦		IA	rtal/
	bas	S Ex	A-In	rch	ersa	creator (E-ARK Web)	rg		A Re	rg.	S-Stc	Ind	ch a	r M	Ing	erve		×	DIP	bas	ewe	oelo	) el	lod 9
	Database Preservation Toolkit	ERMS Export Module	RODA-In	ESSArch Tool for Producer (ETP)	Universal Archiving Module	SIP a	ESSArch Tools for Archive (ETA)	SIP2AIP	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	SOLR Index	Search and Display	Order Management Tool	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	AIP2DIP (E-ARK Web)	Database Visualization	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
					X																			

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		- 1	Pre-l	ngest	t		Ingest - Storage						Storage – Access											
E-ARK Format			Е	-ARK	SIP				Е	-ARK	AIP										Е	-ARK	DIP	Х
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			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

Please note that you can find more details with screenshots on scenario execution in the previous deliverable  $\underline{D2.4}$  Pilot Documentation.

#### **Additional scenarios**

Additional scenario	Extract records with ERMS Export Module and ingest into Preservica (Joint scenario with NAE)											
Description	The National Archives o	f Est	tonia was supposed to ι	se th	ne ERMS Export Module to	export records from ERMS	S					
	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 wit	h th	e newly deployed ERMS	Ехр	ort Module and sent the SN	JURF ERMS file to NAE						
	where a SIP was created	d, ar	nd ingested to Preservic	a. Wi	th this additional scenario e	every step that was origin	ally					
	planned to be tested in	Pilo	t 3 has been successfull	y tes	ted.							
OIAS relevance	Pre-Ingest, Ingest											
Use-case	Extract and Ingest ERMS records based on MoReq2010											
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) (DMJX	()					
Description	Different kinds of letter	s an	d documents									
Data type	Records from Alfresco E	RM	S									
Metadata format	EAD											
Quantity	121 files, 17 MB											
OAIS Relevance	Pre-Ingest	est Ingest - Storage Storage - 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	IS Ex	port	Mod	ule s	cena	rio w	ith lo	ocal E	RMS	syst	em D	DELTA	4					ERMS Export Module scenario with local ERMS system DELTA This additional pilot combines several tools and tests the E-ARK workflow in full from the beginning to the end.													
Description	This	addi	itiona	al pilo	ot co	mbin	ies se	evera	l too	ls and	d tes	ts the	e E-A	RK w	orkfl	ow ii	n full	fron	n the	begi	nnin	g to t	he e	nd.								
	Reco	ords <sup>1</sup>	from	the I	local	DELT	ΓA sy	stem	wer	e exp	orte	d wit	h ER	MS E	xpor	t Mo	dule	then	a SIF	was	crea	ated	and									
	inge	sted	into	Prese	ervic	a. Fir	nally	the a	cces	s was	prov	/ided	by (	CMIS	Port	al Vie	ewer.															
OIAS relevance	Pre-	Inge	st, In	gest																												
Use-case	Extra	act a	nd In	ngest	ERM	S rec	cords	base	ed on	Mol	Req2	010																				
E-ARK specifications	SMU	JRF E	RMS	,																												
E-ARK Tools	ERM	IS Ex	port	Mod	ule																											
Data	Sele	cted	reco	rds fr	rom l	DELT	A ERMS system from partner company Wisercat																									
Description	Diffe	erent	kind	ds of o	docu	men	ts	<u> </u>																								
Data type	Reco	ords '	from	DELT	TA EF	₹MS																										
Metadata format	Not	relev	<i>r</i> ant																													
Quantity	A sm	nall a	mou	ınt of	reco	rds																										
OAIS Relevance		ſ	Pre-li	ngest				Ingest - Storage Storage – Access																								
E-ARK Format			Е	-ARK	SIP	Х			E	-ARK	AIP										Е	-ARK	ARK DIP									
specifications			S	IARD	2.0				SMU	RF EF	RMS	Х			SMU	JRF S	FSB					Geo	data									
E-ARK Tools	ţ			[A]			<b>₹</b>			٦										يز												
	oolki			ır (E	nle		e (ET			Platform										Toolkit												
	on To	0		ance	Mod	(ep)	chive						5	Tool					_	on T												
	vatio	ğ		Pro	l Bu	×	r Ar	Neb		atio			ay G	ent				Ę,	Neb	zati			wei	ver								
		0		<b>&gt;</b>	.≥	Æ	ę ę	×	ton	erva	a)		ispl	me.				ear	×	ıali			Vie	'iev								
	ser	Σ		Ę	동		<u> </u>																	_								
	Preser	port M		Tool fc	l Arch	or (E-	Tools	(E-AI	posi	Pres	rage	ex	D D	anag	sst	Į.		eb Se	(E-A	Vist	L		LAP	tal/v								
	base Preser	S Export M	4-In	rch Tool fc	ersal Arch	reator (E-	rch Tools	AIP (E-ARK Web)	A Reposi	rch Pres	-Storage	Index	ch and D	r Manag	Ingest	erver		K Web So	DIP (E-AF	base Visu	ewer	leo	e (OLAP	portal/v								
	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-AI	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 Visualization	P Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer								

#### **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
CMIS Portal Browser	
Used in tasks	Access
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	

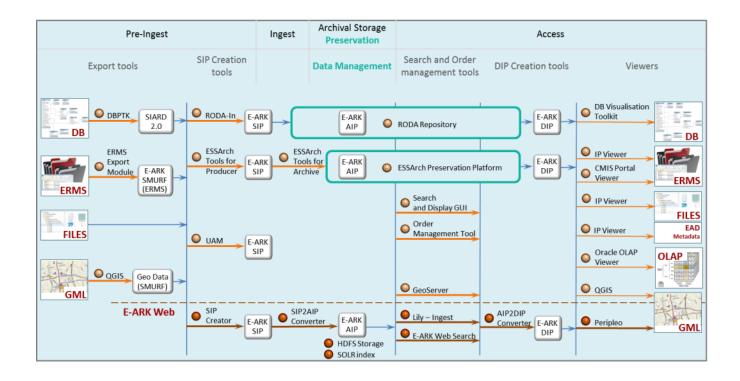
Although the tools and specifications proved to be usable, we are still planning to look for more possibilities to reduce the human factor and automate the workflow in the steps where it is possible in order to make the process even more scalable in the future.

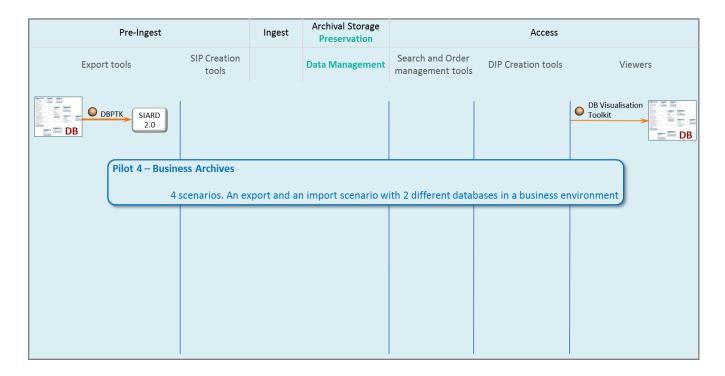
## Recommended practices and further recommendations

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		<ul> <li>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;</li> </ul>
		- 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	<ul> <li>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;</li> </ul>
		- 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	Busi	iness	Arcl	hives	;																			
Task leader	Nati	ional	Arch	nives	of Es	tonia	Э																	
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	solu	ition	in a	busir	iess a	archi	ve	
Object	bes	ooke	busi	ness	syste	em th	nat co	ontai	ns da	ataba	se re	cord	ls											
Short description	com	pany ords.	y is co The	ompı busir	Arch rised ness a to the	of pr	ivate ves p	bus bilot i	iness n the	ses ir	Esto RK pi	nia f ojec	for an	rchiv I focu	ing a us on	nd p tran	reser sfer	vatio	n of	both	рар	er an	d dig	
Contacts	Nan	ne (T	itle)							E-m	ail								Skyp	ре				
Contact Person	Raiv	Raivo Ruusalepp raivo@eba.ee raivoruu																						
Pilot staff member	Ats	Ats Rand <u>ats.rand@eba.ee</u> atsrand																						
OAIS Relevance		Pre-Ingest Ingest - Storage Storage - Access																						
E-ARK Formats		E-ARK SIP E-ARK DIP																						
		SIARD 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
Scenario 1		ratio	n an	l d Ing	est o	f bus	ines	s rec	ords	from	besi	ooke	busi	ness	syste	em (I	L Data	set 1	.)		l	l		
Scenario 2	_				fron						1				,				-					
Scenario 3					est o						besi	ooke	busi	ness	syste	em (I	Data	set 2	.)					
Scenario 4					fron										•				•					

#### Scenarios

Mig	ratio	n an	d Ing	est o	f bu	sines	s rec	ords	from	bes	poke	bus	iness	syst	em								
Expc	ort bu	usine	ss re	cord	s fro	m be	spok	e bus	sines	s syst	em.	Inge	st to	local	arch	ival s	yste	m of	EBA.				
Pre-l	Inge	st, In	gest																				
Extra	act a	nd In	gest	relat	iona	l data	abas	e bas	ed o	า SIA	RD 2	.0											
E-AR	RK SII	P, SIA	RD 2	2.0																			
Data	base	e Pre	serva	tion	Tool	kit																	
Reco	ords	from	besp	oke	busii	ness:	syste	m															
Busi	ness	syste	em w	ith 1	4 tak	oles.	The o	datak	ase o	conta	ins a	ppro	xima	tely	12 00	00 re	cord	s.					
MS-S	SQL a	as m	df																				
none	9																						
more																							
Pre-Ingest Ingest - Storage Storage – Access																							
		Е	-ARK	SIP		E-ARK AIP E-ARK DI											DIP						
		S	IARD	2.0	Х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	lata	
➤ 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
	Patapase Preservation Toolkit  E-AE Busi MS-: none mor	Export be Pre-Inge: Extract a E-ARK SI: Database Records Records Business MS-SQL: none more that	Export busine Pre-Ingest, In Extract and In E-ARK SIP, SIA Database Pre Records from Business syste MS-SQL as me none more than 12 Pre-In ES	Export business re Pre-Ingest, Ingest Extract and Ingest E-ARK SIP, SIARD 2 Database Preserva Records from besp Business system w MS-SQL as mdf none more than 12 000 Pre-Ingest E-ARK SIARD  F-ARK SIARD  E-ARK SIARD	Export business record Pre-Ingest, Ingest Extract and Ingest relat E-ARK SIP, 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  Injude a large of the product of t	Export business records from Pre-Ingest, Ingest Extract and Ingest relationa E-ARK SIP, 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  (au)  Au)  Au)  Au)  Au)  Au)  Au)  Au)	Export business records from be Pre-Ingest, Ingest Extract and Ingest relational data E-ARK SIP, 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  (Y)  SIARD 2.0 X  (EXAMPLE AND A CONTROL OF THE PROPERTY OF	Export business records from bespok Pre-Ingest, Ingest Extract and Ingest relational database E-ARK SIP, SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The of MS-SQL as mdf none more than 12 000 rows Pre-Ingest Ing E-ARK SIP SIARD 2.0 X  Pre-Ingest Ing E-ARK SIP SIARD 2.0 X  ESSArch Tools for Archive (ETA)  SIARM (ETA)	Export business records from bespoke business records from bespoke business records from bespoke business records from bespoke business system  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 14 tables. The database business system with 14 tables. The database business system with 14 tables. The database business system with 14 tables. The database business system with 14 tables. The database business system with 14 tables. The database business system with 14 tables. The database business system with 14 tables. The database business system business system with 14 tables. The database business system with 14 tables. The database business system business system with 14 tables. The database business system business system business system business system business system with 14 tables. The database business system busines	Export business records from bespoke business Pre-Ingest, Ingest  Extract and Ingest relational database based of E-ARK SIP, 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  E-ARK SIP  E-ARK SIP  E-ARK Mep)  E-	Export business records from bespoke business system Pre-Ingest, Ingest Extract and Ingest relational database based on SIA E-ARK SIP, SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 14 tables. The database conta MS-SQL as mdf none more than 12 000 rows  Pre-Ingest Ingest - Storage E-ARK SIP SIARD 2.0  SMURF ERMS Pre-ARK AIP SIARD 2.0  SMURF ERMS  E-ARK AIP SIARD 2.0  BUSINGS  BUSIN	Export business records from bespoke business system.  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2  E-ARK SIP, 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  E-ARK SIP  SIARD 2.0  SMURF ERMS  Pre-ARK AIP  SIARD 2.0  SMURF ERMS  AND SMURF ERMS  SOIR Index  SOIR	Export business records from bespoke business system. Inger Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, 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  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 3.0  SIARD	Export business records from bespoke business system. Ingest to Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, 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  SIARD 2.0  SIARD 2.	Export business records from bespoke business system. Ingest to local Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 14 tables. The database contains approximately 3  MS-SQL as mdf  none  more than 12 000 rows  Pre-Ingest  SIARD 2.0  Signary and Disblay GnI  Onder Management Tool  Onder Management Tool  Onder Management Tool  Onder Management Tool  Pre-Ingest  Solution	Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, 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  SIARD 2.0  SIARD 2.	Export business records from bespoke business system. Ingest to local archival some pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 14 tables. The database contains approximately 12 000 records from bespoke business system  Business system with 14 tables. The database contains approximately 12 000 records from bespoke business system  Business system with 12 tables. The database contains approximately 12 000 records from bespoke business system  Business system with 14 tables. The database contains approximately 12 000 records from bespoke business system  Business system with 14 tables. The database contains approximately 12 000 records from bespoke business system  Business system with 14 tables. The database contains approximately 12 000 records from bespoke business system.  Business system with 14 tables. The database contains approximately 12 000 records from bespoke business system.  Business system with 14 tables. The database contains approximately 12 000 records from bespoke business system.  Business system with 14 tables. The database contains approximately 12 000 records from bespoke business system.  Business system with 14 tables. The database contains approximately 12 000 records from business system.  Business system. Ingest - Storage	Export business records from bespoke business system. Ingest to local archival syste  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, 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  SIARD 2.0  Tiply - Ingest  Search and Disblay GnI  F-ARK SIP  SIARD 2.0  Search and Disblay GnI  F-ARK Mep  SIARD 2.0  Tiply - Ingest  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  F-ARK Mep  Search and Disblay GnI  F-ARK Mep  F-ARK Me	Export business records from bespoke business system. Ingest to local archival system of Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, 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  E-ARK Mep  SIARD 2.0  SIARD 3.0  SIARD 3	Export business records from bespoke business system. Ingest to local archival system of EBA.  Pre-Ingest, Ingest Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, 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  SIARD	Export business records from bespoke business system. Ingest to local archival system of EBA.  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, 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 SIARD 2.0 X SMURF ERMS SMURF STORAGE SIARD 2.0 X SMURF ERMS SMURF STORAGE SMURF STORAGE STORAGE SMURF SMURF	Export business records from bespoke business system. Ingest to local archival system of EBA.  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, 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 SIARD 2.0  Tigh - Ingest  E-ARK SIARD 2.0  Tigh - Ingest  E-ARK MeP  Database Nicalization Toolkit  Database Nicalization Toolkit  Database Nicalization Toolkit  Database Preservation Toolkit  ESSArch Lool Iou Accident  ESSArch Lool Iou Accident  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System Well 2000 records  Business System	Export business records from bespoke business system. Ingest to local archival system of EBA.  Pre-Ingest, Ingest  Extract and Ingest relational database based on 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  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  SIARD 2.0  Beripleo  Beri

Mig	ratio	n an	d Ing	est o	f bu	sines	s rec	ords	from	bes	poke	bus	iness	syst	em								
Expo	ort b	usine	ss re	cord	s fro	m be	spok	e bus	sines	s sys	tem.	Inge	st to	local	arch	ival s	yste	m of	EBA.				
Pre-	Inge	st, In	gest																				
Extra	act a	nd In	gest	relat	iona	l data	abas	e bas	ed o	n SIA	RD 2	.0											
E-AR	RK SII	P, SIA	RD 2	2.0																			
Data	abase	e Pre	serva	tion	Tool	kit																	
Reco	ords	from	besp	oke	busii	ness	syste	m															
Busi	ness	syste	em w	ith 6	3 tak	oles (	+sev	eral l	nisto	y an	d sup	port	tabl	es th	at ar	e not	nee	ded f	or a	com	olete		
structure of the working database). The database contains approximately 200 000 records.  MS-SQL as mdf																							
MS-S	SQL	as m	df																				
none	е																						
mor	e tha	n 20	0 000	) rov	٧S																		
		Pre-li	ngest				Ing	est -	Stora	age						Sto	age	– Acc	cess				
		E	-ARK	SIP				Е	-ARK	AIP										Е	-ARK	DIP	
		S	IARD	2.0	Х			SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	lata	
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
	Expo Pre- Extra E-AF Data Recco Busi struc MS- none mor	Export be Pre-Inger Extract a E-ARK SII Database Records Business structure MS-SQL a none more that	Export busine Pre-Ingest, In Extract and In E-ARK SIP, SIA Database Pre Records from Business syste structure of t MS-SQL as me none more than 20 Pre-In E S	Export business re Pre-Ingest, Ingest Extract and Ingest E-ARK SIP, SIARD 2 Database Preserva Records from besp Business system w structure of the w MS-SQL as mdf none more than 200 000  Pre-Ingest E-ARK SIARD	Export business record Pre-Ingest, Ingest Extract and Ingest relat E-ARK SIP, SIARD 2.0 Database Preservation Records from bespoke Business system with 6 structure of the workin MS-SQL as mdf none more than 200 000 row Pre-Ingest E-ARK SIP SIARD 2.0	Export business records from Pre-Ingest, Ingest Extract and Ingest relationa E-ARK SIP, SIARD 2.0 Database Preservation Tool Records from bespoke busin Business system with 63 tak structure of the working da MS-SQL as mdf none more than 200 000 rows  Pre-Ingest E-ARK SIP SIARD 2.0 X	Export business records from be Pre-Ingest, Ingest Extract and Ingest relational data E-ARK SIP, SIARD 2.0 Database Preservation Toolkit Records from bespoke business Business system with 63 tables ( structure of the working database MS-SQL as mdf none more than 200 000 rows  Pre-Ingest E-ARK SIP SIARD 2.0 X	Export business records from bespok Pre-Ingest, Ingest Extract and Ingest relational database E-ARK SIP, SIARD 2.0 Database Preservation Toolkit Records from bespoke business system Business system with 63 tables (+sev structure of the working database). T MS-SQL as mdf none more than 200 000 rows  Pre-Ingest Ing E-ARK SIP SIARD 2.0 X  I (E-ARK SIP) SIARD 2.0 X  I (E-ARK SIP) I	Export business records from bespoke business records from bespoke business. Ingest  Extract and Ingest relational database base E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several Itstructure of the working database). The districture of the working database). The districture of the working database of the wo	Export business records from bespoke business  Pre-Ingest, Ingest  Extract and Ingest relational database based of E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several histor structure of the working database). The database MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Store  E-ARK SIP E-ARK  SIARD 2.0 X SMURF EF  SIAR	Export business records from bespoke business system  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIA  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and structure of the working database). The database of MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage  E-ARK SIP E-ARK AIP  SIARD 2.0 X SMURF ERMS  Pre-Ingest Ingest - Storage  E-ARK AIP  SIARD 2.0 X SMURF ERMS	Export business records from bespoke business system.  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and supstructure of the working database). The database contains MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage  E-ARK SIP E-ARK AIP  SIARD 2.0 X SMURF ERMS  SIARD 2.0 X SMURF ERMS  F-ARK AIP  SIARD 2.0 X SMURF ERMS	Export business records from bespoke business system. Inge  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support structure of the working database). The database contains all MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage  E-ARK SIP E-ARK AIP  SIARD 2.0 X SMURF ERMS  SIARD 2.0 X SMURF ERMS  Output  Ingest - Storage  E-ARK SIP E-ARK AIP  SIARD 2.0 X SMURF ERMS  Output  Ingest - Storage	Export business records from bespoke business system. Ingest to Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support table structure of the working database). The database contains approx MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP  SIARD 2.0 X SMURF ERMS  SIARD 2.0 X SMURF ERMS  Ingest - Storage E-ARK AIP Ingest - Storage Ingest - Storag	Export business records from bespoke business system. Ingest to local Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables the structure of the working database). The database contains approximat MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage E-ARK SIP E-ARK AIP  SIARD 2.0 X SMURF ERMS SMU  SIARD 2.0 X SMURF ERMS SMURF INGER SIARD 2.0 SMURF ERMS SMURF INGER SIARD 2.0 SMURF ERMS SMURF INGER SIARD 2.0 SMURF ERMS SMURF INGER SIARD 2.0 SMURF ERMS SMURF INGER SIARD 2.0 SMURF ERMS SMURF INGER SIARD 2.0 SMURF ERMS SMURF INGER SIARD 2.0 SM	Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables that an structure of the working database). The database contains approximately 2  MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage  E-ARK SIP E-ARK AIP  SIARD 2.0 X SMURF ERMS SMURF SMURF SIARD SMURF	Export business records from bespoke business system. Ingest to local archival struct and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables that are not structure of the working database). The database contains approximately 200 of MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage Storage Storage SIARD 2.0 X SMURF ERMS SMURF SFSB  SIARD 2.0 X SMURF ERMS SMURF SFSB  SIARD 2.0 X SMURF ERMS SMURF SFSB	Export business records from bespoke business system. Ingest to local archival syste  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables that are not nee structure of the working database). The database contains approximately 200 000 re  MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage Storage  E-ARK SIP E-ARK AIP  SIARD 2.0 X SMURF ERMS SMURF SFSB  SIARD 2.0 X SMURF ERMS SMURF SFSB  SIARD 2.0 X SMURF ERMS SMURF SFSB	Export business records from bespoke business system. Ingest to local archival system of Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables that are not needed to structure of the working database). The database contains approximately 200 000 record MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage Storage Acc  E-ARK SIP E-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB  SIARD 2.0 X SMURF ERMS SMURF SFSB  E-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB  SIARD 2.0 X SMURF ERMS SMURF SFSB  BE-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB  BE-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB  BE-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB	Export business records from bespoke business system. Ingest to local archival system of EBA.  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables that are not needed for a structure of the working database). The database contains approximately 200 000 records.  MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage Storage Access  E-ARK SIP E-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  E-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  B-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  B-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  B-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  B-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  B-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  B-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  B-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access  B-ARK SIP SIARD 2.0 X SMURF ERMS SMURF SFSB Ingest - Storage Access SMUR	Export business records from bespoke business system. Ingest to local archival system of EBA.  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables that are not needed for a compstructure of the working database). The database contains approximately 200 000 records.  MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage Storage - Access  E-ARK SIP E-ARK AIP E  SIARD 2.0 X SMURF ERMS SMURF SFSB E  SIARD 2.0 X SMURF ERMS SMURF SFSB E  SIARD 2.0 X SMURF ERMS SMURF SFSB E  SIARD 2.0 IN SMURF ERMS SMURF SFSB SMURF SFSB E  SIARD 2.0 IN SMURF ERMS SMURF SFSB SMURF SFSB SIARD S	Export business records from bespoke business system. Ingest to local archival system of EBA.  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables that are not needed for a complete structure of the working database). The database contains approximately 200 000 records.  MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage Storage - Access  E-ARK SIP E-ARK AIP E-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB Geod  SIARD 2.0 X SMURF ERMS SMURF SFSB Geod  SIARD 2.0 X SMURF ERMS SMURF SFSB Geod  SIARD 2.0 X SMURF ERMS SMURF SFSB Geod  SIARD 2.0 X SMURF ERMS SMURF SFSB Geod  SIARD 2.0 X SMURF ERMS SMURF SFSB Geod  Business system of EBA.  Pre-Ingest Ingest - Storage Storage - Access  E-ARK SIP E-ARK AIP E-ARK  SIARD 2.0 X SMURF ERMS SMURF SFSB Geod  SMURF SFSB GEOD  S	Export business records from bespoke business system. Ingest to local archival system of EBA.  Pre-Ingest, Ingest  Extract and Ingest relational database based on SIARD 2.0  E-ARK SIP, SIARD 2.0  Database Preservation Toolkit  Records from bespoke business system  Business system with 63 tables (+several history and support tables that are not needed for a complete structure of the working database). The database contains approximately 200 000 records.  MS-SQL as mdf  none  more than 200 000 rows  Pre-Ingest Ingest - Storage Storage - Access  E-ARK AIP SIARD 2.0 X SMURF ERMS SMURF SFSB Geodata  E-ARK DIP SIARD 2.0 X SMURF ERMS SMURF SFSB Geodata

Scenario 4	Extracting records from date	tabase										
Description	Extracting records from dat	abase containing no doc	ume	nts.								
OIAS relevance	Access (not DIPs involved or	nly restoring data from S	SIARE	D packages)								
Use-case	Access databases via DBVT	( (SQL)										
E-ARK specifications	SIARD 2.0											
E-ARK Tools	Database Preservation Toolkit											
Data	Records from bespoke business system											
Description	Business system with 63 tables (+several history and support tables that are not needed for a complete											
	structure of the working da	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 Storage – Access											
E-ARK Format	E-ARK SIP E-ARK AIP E-ARK DIP X											
specifications	SIARD 2.0	SMURF ERMS	Х	SMURF SFSB		Geodata						

se Preservation Trickport Module In h Tool for Produce sal Archiving Modil ator (E-ARK Web) h Tools for Archiving (E-ARK Web) Repository h Preservation Platorage and Display GUI Management Tool gest ver Meb Search Neb Search rer o O OLAP Viewer) ortal/viewer		E-ARK Tools
ERMS Export Module RODA-In ESSArch Tool for Produce Universal Archiving Mod SIP creator (E-ARK Web) SIP ZAIP (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 AIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		atabase Preservation
ESSArch Tool for Produce Universal Archiving Mod SIP creator (E-ARK Web) ESSArch Tools for Archivo SIPZAIP (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 AIPZDIP (E-ARK Web) AIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		RMS Export Modul
ESSArch Tool for Produce Universal Archiving Mod SIP creator (E-ARK Web) ESSArch Tools for Archivo SIP ZAIP (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 AIPZDIP (E-ARK Web) AIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		Ā
Universal Archiving Mod SIP creator (E-ARK Web) ESSArch Tools for Archive SIPZAIP (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 AIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		SSArch Tool for Pro
SIP creator (E-ARK Web) ESSArch Tools for Archiw 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) AIP2DIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		niversal Archiving Mo
ESSArch Tools for Archives SIPZAIP (E-ARK Web) RODA Repository ESSArch Preservation Plates and Display GUI Order Management Tool Lily - Ingest Geoserver Geoserver QGIS E-ARK Web) AIPZDIP (E-ARK Web) AIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		IP creator (E-ARK W
SIPZAIP (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) AIPZDIP (E-ARK Web) AIPZDIP (E-ARK Web) AIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		Tools for
ESSArch Preservation Pla HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) AIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		2AIP (E-ARK W
ESSArch Preservation Pla HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) RIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		ODA Reposit
Solr Index Solr Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) AIPZDIP (E-ARK Web) Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		SSArch Preservati
SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) C Database Visualization T IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		
Search and Display GUI Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) C Database Visualization T IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		OLR
Order Management Tool Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) C Database Visualization T IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		earch and Display
Lily - Ingest Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) C Database Visualization T IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		rder Management To
Geoserver QGIS E-ARK Web Search AIPZDIP (E-ARK Web) C Database Visualization T IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		7
E-ARK Web Search AIPZDIP (E-ARK Web) C Database Visualization T IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		osen
E-ARK Web Search AIPZDIP (E-ARK Web) C Database Visualization T IP Viewer Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		QGIS
AIPZDIP (E-ARK Web)  C Database Visualization T  IP Viewer  Peripleo  Oracle (OLAP Viewer)  CMIS portal/viewer		-ARK Web S
Peripleo Oracle (OLAP Viewer) CMIS portal/viewer		IP (E-ARK
Viewer eripleo racle (OLAP Viewe	Χ	atabase Visualization
eripleo racle (OLAP Viewe MIS portal/viewer		P Viewe
racle (OLAP Viewe		o.
MIS portal/view		racle (OLAP Viewe
		MIS portal/view

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
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	

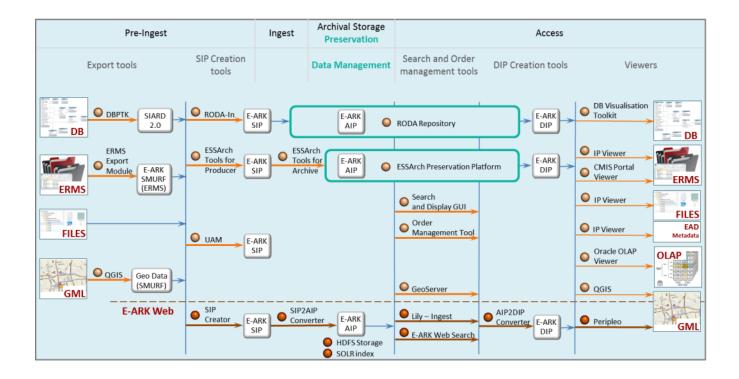
E-ARK Tool – Version	Issues (bugs, wishes, comments)
	Experiences / Recommended practices
Database Visualization Toolkit	
Used in task	Access – in scenario 2 and 4
Data (input / output)	Input: SIARD 2.0 packages
	Output: Restored DB tables
Performance	Good
Issues	No issues found
Wishes	None
Comments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

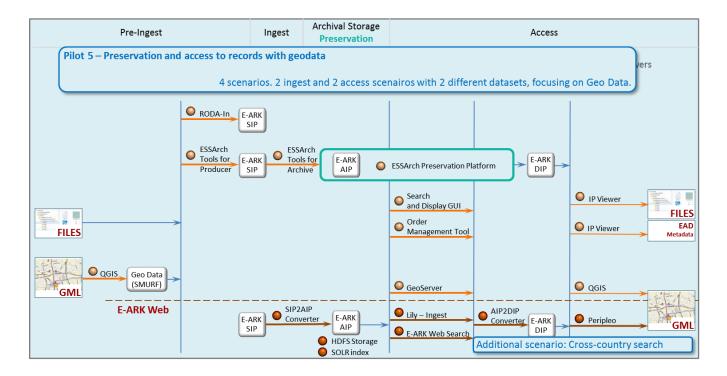
## Recommended practices and further recommendations

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	erva	tion	and	acce	ss to	reco	rds v	with	geod	lata													
Task leader	Nati	onal	Arch	nives	of Sl	oven	ia																	
Supported by	Dan	ish N	latio	nal A	rchiv	es																		
Scope	Pilot	t will	prov	e th	at the	e SIP	and	DIP i	mple	men	tatio	ns fu	ılfill s	peci	fic re	quire	emer	nts fo	r the	erec	ords	conta	ainin	3
					instr											_	_				_			
					ise of						а оре	en da	ita m	etho	d, di	rect	acces	ss in	the a	archiv	es a	nd us	se GIS	5
					teria					•														
Object										_														t for
					ng up rable				ironr	nent	OT SE	electe	ea E-	AKK	arcn	ivai t	OOIS	prov	iae r	earı	re ex	ampi	es no	)W
Short description					proje				dize	d me	thod	for i	nges	ting	gen (	lata i	will h	ne de	velo	ned	This	willa	llow	the
Short description		_			geoda								_	_	_									
		ne ar		_	,						-   ,					-,				,				
Contacts	Nan	lame (Title)  E-mail  Skype  Gregor Završnik ()  gregor.zavrsnik@gov.si  gregor.zavrsnik																						
Contact Person	Gre																							
Pilot staff member	Ale	lenka Starman () <u>alenka.starman@gov.si</u>																						
Pilot staff member	Anja	nja Paulič ()  Anja.Paulic@gov.si																						
Pilot staff member	Joze																							
OAIS Relevance		pre-Ingest     Ingest - Storage     Storage - Access																						
E-ARK Formats			E	-ARk	( SIP	X			Ε	-ARK	AIP	X									Е	-ARK	DIP	X
			S	IARD	2.0				SMU	RF EF	RMS				SMI	JRF S	FSB	X				Geo	data	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 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						
	CID.		. X	X	<u> </u>		X	X		<u> </u>		<b>X</b>	X	X	X	X	X	Х	X		Х	X		
Scenario 1					ngest																			
Scenario 2		Search and Access information using Geodata (Data set 1-2) SIP Creation and Ingest of records with Geodata (Data set 3)																						
Scenario 3																								
Scenario 4	Sear	ch a	nd A	ccess	info	rmat	ion ι	ısing	Geo	data	(Dat	a set	3)											
Additional scenario	Cros	ss-co	untry	y sea	rch w	ith E	-ARK	( We	b (jo	int sc	enar	io wi	ith N	AH)										

#### Scenarios

Scenario 1	SIP (	Creat	tion a	and Ir	nges	t of r	ecor	ds w	ith G	eoda	ata													
Description	Crea	ite SI	P fro	m re	cord	s and	d met	tadat	a exp	orte	d fro	m Gl	JRS (	The S	Surve	ying	and	Мар	ping	Auth	ority	of th	1e	
	Repu	ublic	of SI	oveni	ia).																			
	SIP c	creat	ion a	nd in	gest	of at	t leas	t one	e sma	all ve	ctor	geod	ata s	et wi	th le	ss th	an 10	00 re	cord	s and	one	with	mor	e
	than	100 ا	0 rec	ords.	. Arc	hivis	t crea	ates a	a Sub	miss	ion a	gree	ment	t for	SIP cı	eatio	on, a	ccord	ding t	to E-A	٩RK ۽	guide	lines	for
	geod	data	SIP c	reatio	on. P	rodu	icer c	reate	es a S	SIP co	ontaiı	ning (	geod	ata, a	accor	ding	to S	ubmi	issior	n agre	eeme	nt, b	ased	on
				cificat								-								_				:-
	ARK	guid	eline	s for	geo	data	SIP c	reati	on. A	rchiv	ist c	onfiri	ms, t	hat c	onte	nt va	lidat	ion o	of the	subr	mitte	d SIF	1	
	pack	(age	was	perfo	rme	d. Ar	AIP	is ge	nera	ted fi	rom t	he S	IP an	d get	s ing	este	d into	the	arch	ival r	epos	itory	•	
OIAS relevance	Pre-	Inge	st, Ing	gest																				
Use-case	Othe	er (SI	SIP Creation and Ingest of records with Geodata) SIP, E-ARK AIP (with GeoData)																					
E-ARK specifications	E-AR	≀K SIF	P, E-A	\RK A	AIP (v	vith (	GeoD	ata)																
E-ARK Tools	ROD	A-In,	, ESS/	Arch	Tool	s Arc	hive	(ETA	), SIP	2AIP	(E-A	RK V	Veb)	, ESS	Arch	Prese	ervat	ion F	Platfo	rm, I	EAD	Edito	r, QG	ilS
Data	Two	sets	from	1 the	Surv	ols Archive (ETA), SIP2AIP (E-ARK Web), ESSArch Preservation Platform, EAD Editor, QG rveying and Mapping Authority of the Republic of Slovenia:																		
	1.) R	1.) Records and metadata of municipalities as valid until 1994, exported from GURS, database																						
	2.) Records and metadata of administrative units until 1994, exported from GURS																							
Description	Reco	ords a	and r	netac	data	of m	aps v	with (	Geod	lata														
Data type	GML	doc	ume	nt wi	th m	etad	ata iı	n XM	L for	mat,	ESRI	Shap	efile	, csv										
Metadata format	ISO :	1911	5 (IN	SPIRE	E)																			
Quantity	62 re	ecord	ds (cc	ca. 3N	ИВ) -	120	)4 red	cords	(сса	. 12,	4 MB	)												
OAIS Relevance		F	Pre-Ir	ngest				Ing	est -	Stora	age						Sto	rage	– Acc	cess				
E-ARK Format			Е	-ARK	SIP	Х			Е	-ARK	AIP	Х									Е	-ARK	DIP	Х
specifications			S	IARD	2.0				SMU	RF E	RMS				SMU	JRF S	FSB	Х				Geo	lata	Х
E-ARK Tools				(a			7			_														
	okit			ET	le		(ET/			forn										oki				
	10			ncer	odu	(q	ive			Plat			=	loo						ιTo				
	tio	ule		rod	g M	We	Arch	(qə		ion			, GU	nt To				_	eb)	ation			/er)	<u>.</u>
	erv	Mod		or P	ivin	ARK	for	<b>≥</b>	ory	rvat			pla	me				arch	<b>≥</b>	alize			/iev	iewe
	e Preservatio  tport Module  Tool for Prod al Archiving IV tor (E-ARK Web)  Preservation orage dex and Display Gl anagement T est er  E-ARK Web) ev Search (E-ARK Web) ev Search est er  CE-ARK Web) ev Sivalizatio												al/v											
	ase I	Expo	阜	된	sal /	atoı	h T	P (E	Rep	h Pı	Stor	nde	anc	Mar	ges	rver		Wel	IP (E	ase 1	ver	Q	ō	ort
	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 Geoserver AIP2DIP (E-ARK Web) Database Visualization Toolkit IP Viewer Peripleo Oracle (OLAP Viewer)											CMIS portal/viewer												
	Da	ER		ES.	Ž	SIF			RG.		呈		Se	ŏ	Ë	Ge		E-4	₹	Da	₫	Pe	ŏ	S
			X				X	Х		X		X					X							

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:

OAIS Relevance Pre-Ingest Ingest - Storage Storage - Access  E-ARK Format	Description Data type Metadata format Quantity	2.) Record Record GML de ISO 19	cord ds ar docui 9115	nd meta ment w (INSPII	netad adata vith m RE)	ata o of m etad	of adn aps v ata ir	ninist vith G	trativ Geod L forr	e uni ata mat, E	ts un	ntil 19	994,	expo					, data	abas	e			
specifications SIARD 2.0 SMURF ERMS SMURF SFSB X Geodata E-ARK Tools 및 및 및 및 및 및 및 및 및 및 및 및 및 및 및 및 및 및 및	,			<u> </u>	<u> </u>				<u> </u>								Sto	rage	- Acc	ess				
E-ARK Tools	E-ARK Format			E-AR	K SIP				E-	-ARK	AIP	Х									Е	-ARK	DIP	Х
	specifications			SIAR	2.0			S	SMUF	RF ER	MS				SML	JRF S	FSB	Х				Geod	lata	Х
ERMS Export Module  RODA-In  ESSArch Tool for Producer (I  Universal Archiving Module  SIP creator (E-ARK Web)  ESSArch Tools for Archive (E  SIPZAIP (E-ARK Web)  RODA Repository  ESSArch Preservation Platfo  ESSArch Tools for Archive (I  SIPZAIP (E-ARK Web)  RODA Repository  A Search and Display GUI  X Search and Display GUI  X Crider Management Tool  X Cacserver  X Geoserver  X Geoserver  X Geoserver  X Geoserver  X Geoserver  X Peripleo  Oracle (OLAP Viewer)	E-ARK Tools	oolkit			dule		ve (ETA)			atform				<del>-</del> 0						Toolkit				

Scenario 3	SIP (	Crea	tion a	and I	nges	t of r	ecor	ds w	ith G	eoda	ata													
Description	Crea	ate SI	P fro	m re	cord	s and	d met	tadat	a exp	oorte	d fro	m Al	RSO (	Envi	ronm	enta	l Age	ncy	of Re	publi	ic of	Slove	nia).	
	SIP	creat	ion a	nd in	gest	of at	t leas	t one	e vec	tor g	eoda	ta wi	ith at	leas	t 250	reco	ords.	Data	is ex	port	ed di	irectl	y fro	m
	thei	r ow	n sys	tem i	nto (	GML	form	at. A	nd th	neir s	yster	n als	o exp	oorts	INSF	PIRE r	neta	data						
	Arch	nivist	crea	tes a	Sub	missi	on a	greei	nent	for S	SIP cr	eatic	n, ac	cord	ing t	o E-A	RK g	uide	lines	for g	eoda	ita SI	Р	
	crea	ition.	Prod	ducer	crea	ates a	a SIP	cont	ainin	g geo	odata	, acc	ordi	ng to	Subr	nissi	on ag	greer	nent,	base	ed or	i EAR	K SIF	)
	spec	cifica	tions	for g	geod	ata. A	Archi	vist t	echn	ically	/ valid	dates	the	subn	nitte	d SIP	pack	age,	acco	rding	g to E	-ARK		
	guid	leline	es for	geo	data	SIP c	reati	on. A	rchi	∕ist c	onfir	ms, t	hat c	onte	nt va	lidat	ion o	of the	subi	mitte	d SIF	pac	kage	
	was	perf	orme	ed. Aı	n AIP	is ge	enera	ited 1	from	the S	SIP ar	nd ge	ts in	geste	d int	o the	e arcl	hival	repo	sitor	у.			
OIAS relevance	Pre-	Inge	st, In	gest		and Ingest of records with Goodata)																		
Use-case	Oth	er (SI	P Cre	eatio	n and	and Ingest of records with Geodata)																		
E-ARK specifications	E-AF	RK SII	P, E- <i>F</i>	ARK A	AIP (v	P (with GeoData)																		
E-ARK Tools	ESSA	Arch	Tools	s Pro	duce	ucer (ETP), ESSArch Tools Archive (ETA), ESSArch Preservation Platform, EAD Editor, QGIS										;								
Data	Reco	ords	and r	neta	data	ata of Natura 2000 areas created in 2004, exported from ARSO database																		
Description	Records and metadata of maps with Geodata																							
Data type	GML document with metadata in XML format, ESRI Shapefile																							
Metadata format	INSF	PIRE																						
Quantity	286	reco	rds (	cca. S	9,6 N	1B)																		
OAIS Relevance			Pre-I	ngest				Ing	est -	Stor	age						Sto	rage	– Acc	cess				
E-ARK Format			Е	-ARK	SIP				Е	-ARK	AIP										Е	-ARK	DIP	
specifications			S	IARD	2.0	Х			SMU	RF EI	RMS				SMU	JRF S	SFSB					Geo	lata	
E-ARK Tools	ير			(P)			€			۶										يز				
	olki			r (E1	əlr		ET			ffor										oolki				
	n Tc			luce	lodı	(qa	hive	_		Pla			5	00					_	n Tc			_	
	atio	dule		Prod	N B∩	×	Arc	(E-ARK Web)		tion			9	nt T				ء	/eb)	atio			wer)	ē
	serv	Moc		for I	hiviı	-ARI	for	× ×	tory	erva			spla	eme				earc	×	raliz			Vie	/iew
	Pre	ort		00.	Arc	r (E	Si Oo	E-AF	oosit	rese	rage	X	ΘÞ	nag	st	_		S q	E-AF	Vist			ΙAΡ	la 🗸
	ase	Exp	-In	ch T	rsal	eatc	ch T		Rel	ch P	Sto	pul	h an	Ma	nge	ırve		×	J die	ase	wer	eo	0) a	pod
	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
	Õ	Ef	Ŗ	X	n	SI	X	N X	ž	X	I	X	Š	0	L	9	ď	Ē	A	Ö	I P	P.	0	ס
				^			^	^		^		^												Щ.

Scenario 4	Sear	rch a	nd A	ccess	info	rma	tion	using	Gea	dota	1													
Description	Crea	ate D	IP fro	m A	IP со	ntair	ning r	ecor	d wit	h Ge	odat	a. Pre	esent	t Geo	data	info	rmat	ion v	vith (	QGIS	alon	g wit	h	
	cont	content and metadata from DIP.																						
	A da	A data object containing geodata can be identified by using search criteria as specified by E-ARK Tool																						
	requ	uiren	nent :	speci	ficat	ion a	fter	searc	h inc	lex w	as up	odate	ed fro	om ai	า AIP	Sele	ected	l data	a obj	ects a	are se	elect	ed ar	nd
	orde	er is i	issue	d. DII	is p	repa	red a	ccor	ding	to or	der s	peci	ficati	on a	nd er	d us	er cr	eden	itials.	DIP	file s	truct	ure v	vith
			•																a fro				ın be	
																			acce	essing	g PRE	MIS		
	doc	umei	ntatio	n. A	ccess	to D	IP is	docu	ımen	ited a	and c	aptu	red r	neta	data	can b	e ex	porte	ed.					
OIAS relevance	Acce																							
Use-case	Oth	er (A	ccess	of re	ecord	ds wi	th Ge	eodat	ta)															
E-ARK specifications			P, E-																					
E-ARK Tools									-			•	_			n Pre	serv	ation	ı Plat	form	, E-A	ARK V	Veb	
		(Search), AIP2DIP (E-ARK Web), IP Viewer, QGIS, Geoserver, Peripleo																						
Data											reate	d in 2	2004	, exp	ortec	fror	n AR	SO d	ataba	ase				
Description	Reco	ords	and r	neta	data	of m	aps v	with (	Geod	lata														
Data type	GMI	L doc	ume	nt wi	th m	etad	ata ii	n XM	L for	mat,	ESRI	Shap	efile	!										
Metadata format	INSF	PIRE																						
Quantity	286	reco	rds (	cca. S	9,6 N	1B)																		
OAIS Relevance			Pre-li	ngest				Ing	est -	Stor	age						Sto	rage	– Ac	cess				
E-ARK Format			Е	-ARK	SIP				E	-ARK	AIP	Х									Е	-ARK	DIP	Х
specifications			S	IARD	2.0				SMU	RF E	RMS				SMU	JRF S	FSB	Х				Geo	data	Х
E-ARK Tools				(A			8			_														
	olki			. (ЕТ	le		(ET,			form										oki				
	n To			nceı	npo	(q	nive			Plat			=	00						n To				
	atio	ule		rod	MB	We	Arc	(eb)		ion			א פר	ntT				_	(qə	atio			ver)	-a
	erv	Mod		for F	ivin	ARK	for	\ <u>\</u>	ory	rvat			spla	me				arch	≥ ≥	aliz			Viev	iew
	Pres	ort		ool 1	Arch	r (E-	sloo	(E-ARK Web)	osit	rese	age	×	βDi	nage	it			b Se	-AR	Visu			AP	al/v
	ase	Exp	투	ch To	rsal	ato	유	) d	Rep	ch P	Stor	nde	an(	Ма	nges	Zer		We	IP (E	ase	ver	o <sub>a</sub>	0	oort
	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
	Da	ER	RC	ES	בֿ	SIF	ES	SIF	RC	ES	불			ŏ				Ē,		Da			ō	บ
			İ					l				Х	Х	Х	Х	Х	Х	Х	Х		Х	Х		l

Additional scenario	Cross-country search with E	E-ARK Web (joint scenario w	vith NAH)							
Description	The SOLR index and E-ARK V	The SOLR index and E-ARK Web infrastructure theoretically makes it possible to perform a federated search								
	over more than one archive	over more than one archive. When the SOLR index of the other archival institution can be "seen" by the search								
	engine (e.g. one institution	engine (e.g. one institution has access rights to the others SOLR) then it can make a common list of the result.								
	The National Archives of Slo	venia and the National Arch	ives of Hungary both have an E-ARK implementation at							
	their pilot sites. This scenari	o is a simple feasibility study	y 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 a	archives will be theoretically	connected in this sceanrio							
Data type	Not relevant									
Metadata format	Not relevant	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		SMURF ERMS SMURF SFSB Geoda			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						

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	OK
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	OK
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	OK
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
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.

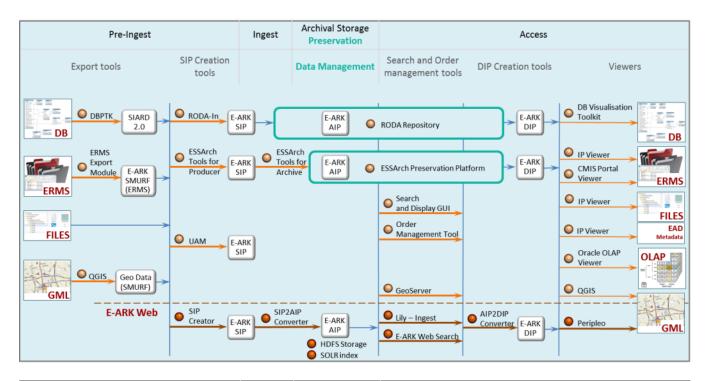
#### Full text search brings the archival experience closer to our users

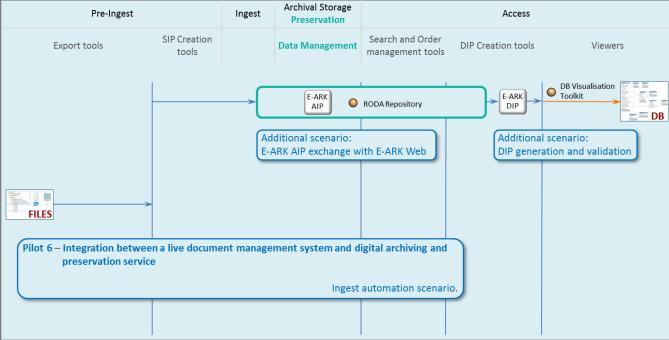
E-ARK Web based SOLR index with the Magenta Search interface brought us a new experience - full text search. Previously the only search option was using the catalogue. This brings our users an experience similar to the way of searching that they are used to already using (Google, Bing...). This provides 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 ma	nagement :	sy	stem and digital archiving a	nd preservation service								
Task leader	KEEP SOLUTIONS (KEEPS)														
Supported by	Instituto Superior Técnico (IST)														
Scope	The goal of this pilot is	two-	fold. On one h	nand, KEEP S	SO	LUTIONS will demonstrate t	hat the pan-European SIP								
	structure designed in t	structure designed in the WP3 is adequate to support the media types found in today's Electronic Records													
	Management Systems	(e.g.	text documen	ts, video, au	ud	io, images, etc) and, on the	other hand, that the most								
	adequate and scalable	form	of ingest is to	automate	th	e SIP creation and delivery p	rocess 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 pilot will also demonstrate the capabilities of the preservation services that follow the transfer of data to														
						ins to access Dissemination I s served by the preservation									
Short description	·					Information Package Specific									
Short description	*					- · ·									
	metadata and data should be packaged in order to move records between the three stages of records keeping - active, semi-active and inactive.														
				o be archive	ed	usually falls into one these t	three "ages":								
	- Active - when the metadata and data are "live" being used and modified regularly Semi-active - when the metadata and data are archived for a short period – say up to 5 years.														
	- Inactive - when the metadata and data are moved to a long-term repository for permanent conservation.														
	The pilot aims to do ensure the seamless transference of information between the semi-active and the														
	inactive stages in a way that no relevant data or metadata is lost in the process. To accomplish this goal, a														
	special integration tool has been developed that implements the package specifications and orchestrates 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														
		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 is 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														
	or semi-active archival														
Contacts	Name (Title)	<u>,                                     </u>		E-mail		· · ·	Skype								
Contact Person	Miguel Ferreira			mferreira@	k	eep.pt	jmaferreira								
Pilot staff member	Luís Faria			Ifaria@kee	p.	<u>pt</u>	luis100								
Pilot staff member	Hélder Silva			hsilva@kee	p.	.pt	hsilva_keep								
Pilot staff member	Sebastien Leroux			sleroux@ke	ee	<u>p.pt</u>	slerouxatkeep								
Pilot staff member	Rui Rodrigues			rrodrigues(	<u>ه</u>	<u>keep.pt</u>	rui.tiago.mr								
Pilot staff member	Ricardo Vieira			rjcv@tecnio			ricardojoao.vieira								
Pilot staff member	João Cardoso			joao.m.f.ca	rd	loso@tecnico.ulisboa.pt	joao.m.f.cardoso								
OAIS Relevance	Pre-Ingest			Storage		Storage	– Access								
E-ARK Formats	E-ARK SIP	X			X		E-ARK DIP X								
	SIARD 2.0		SMU	RF ERMS		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
									Х															
Scenario 1	Aut	Automatic ingest of records from a semi-active archival management system																						
Additional scenario	Inte	Integration with OMT via E-ARK DIP																						
Additional scenario	Bon	Repository succession via E-ARK AIP (E-ARK AIP exchange experiments)																						

## Scenarios

Scenario 1	Automatic ingest of records from a semi-active archival management system												
Description	This scenario aims to demonstrate the ability to seamlessly transfer data from a semi-active records												
	management system to	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. The												
	scenario enhances the	estal	olished practice by addi	ng an	additional component to	o its a	architecture that will b	эe					
	responsible for the long	g-ter	m preservation of histor	ical r	ecords once they reach t	heir	inactive age. The long-	;-					
	term preservation repo	sitor	ry runs as a back-end se	vice	of the Archival Managen	nent	System and aims to						
	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 Pipeline (RIP), RODA Repository												
Data	Historical records												
Description	Data used in this pilot scenario was comprised of a collection of digitised books related to the Peninsular												
	dating from 1778 to 1834. The collection is composed of 964 records stored in a relational database following												
	the semantic elements of EAD. The dataset also contains a total of 34.600 pages of documentation in												
	uncompressed TIFF files at 300 dpi. The total amount of data is around 1.2 TB. This collection can be inspected												
	at its original location at http://arquivo.cm-mafra.pt/details?id=173037.												
Data type	300 dpi uncompressed TIFF files												
Metadata format	EAD												
Quantity	964 records described	964 records described in EAD containing a total of 34.600 pages of 300 dpi uncompressed TIFF files. The total											
,	amount of data is around 1.19 TB.												
OAIS Relevance	Pre-Ingest Ingest - Storage Storage — Access												
E-ARK Format	E-ARK SIP	Х	E-ARK AIP	Х			E-ARK DIP						
specifications	SIARD 2.0		SMURF ERMS		SMURF SFSB	Х	Geodata	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)
Х	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	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 (	comb	oinat	ion o	f the	Ord	er M	anag	emer	nt To	ol (O	MT)	and	E-AR	K IP \	/iewe	er to	prov	ide a	ccess	s to
	exist	ting o	digita	ıl obj	ects	to its	user	rs. In	orde	r to	articu	ılate	the I	RODA	A rep	osito	ry sy	stem	with	the:	se to	ols, a	new	,
	proc	ess l	nas b	een d	devel	ope	d for	ROD	A tha	it en	ables	an a	rchiv	ist to	crea	ate E	-ARK	com	plian	t DIP	s. Th	ese f	iles c	an
	then	be o	dowr	load	ed ai	nd ac	dded	to th	ne ON	∕IT w	orkfl	ows i	n or	der to	be s	serve	d to	the e	end-u	ıser.				
				wor			_								_								-	
				oade																				
				lso b		rsult	ed us	ing t	he R	ODA'	s RES	ST AP	I, foi	exa	mple	, to s	uppo	ort a	more	adv	ance	d sys	tems	;
			on ap	proa	ıch.																			
OIAS relevance	Acce	ess																						
Use-case																								
E-ARK specifications	E-AR	RK DI	Р																					
E-ARK Tools	ROD	A Re	posit	tory,	Orde	er Ma	anage	emer	nt To	ol														
Data	Test	est data Different kinds of letters and decuments																						
Description	Diffe	Different kinds of letters and documents																						
Data type	Not	relev	/ant																					
Metadata format	Not	relev	/ant																					
Quantity	sma	II																						
OAIS Relevance		ı	Pre-Ii	ngest				Ing	est -	Stor	age						Sto	rage	- Acc	ess				
E-ARK Format			Е	-ARK	SIP				Е	-ARK	AIP										Е	-ARK	DIP	Х
specifications			S	IARD	2.0				SMU	RF E	RMS				SMU	JRF S	FSB					Geod	lata	
E-ARK Tools	it			TP)			(¥			Ε										ī				
	S S			ır (E	nle		ET			ttor										N N				
	n T			ance	Лод	eb)	hive			Pla			5	00						n T				
	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) SIP2AIP (E-ARK Web) SIP2AIP (E-ARK Web) COLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver Geoserver QGIS E-ARK Web Search AIP2DIP (E-ARK Web) Database Visualization Toolkit IP Viewer Peripleo Oracle (OLAP Viewer)																							
	sen	Μ		for	hivi	-AR	s fo	RK \	ton	erva	a)		ispla	em (				earc	RK V	uali			Vie	view
	Pre	oort		[00]	Arc	or (E	00	(E-A	posi	Pres	rage	ex	O br	anag	st	<u>_</u>		eb S	E-A	Vis	L		LAP	tal/
	base	S Ex	₄-In	. u	ersa	eat	G		A Re	함	-Sto	Ind	h ar	ž	lnge	erve		Š	) dic	base	we	leo	e (0	por
	ata	RM	RODA-In	SSA	Jnive	IP CI	SSA	SIP2AIP	RODA Repository	SSA	HDFS-Storage	SOLR Index	ear	Orde	Lily - Ingest	Geoserver	QGIS	E-ARK Web Search	IIP2	ata	IP Viewer	Peripleo	Jrac	CMIS portal/viewer
		Ш	æ	Ш	٦	S	Ш	S	X	В	_	S	S	Х	_	0	J	Ш	٨		=	Ь	U	U

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

	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																							
		_	•																			•		
			-	idex i												,	(		, .					,
										ction	n of A	IPs v	vill b	e trai	nsfer	red f	rom	the F	RODA	repo	osito	rv svs	stem	to
											vious													
							•				in th					-								Ps,
						_					em w													
OIAS relevance			Stora																					
Use-case																								
E-ARK specifications	E-AF	RK AI	Р																					
E-ARK Tools	ROD	A Re	posi	tory,	E-AR	K W	eb																	
Data	Test	data	3																					
Description	Diffe	erent	kind	ls of I	lette	rs an	d do	cume	ents															
Data type	Not	relev	/ant																					
Metadata format	Not	Not relevant																						
Quantity	sma	II																						
OAIS Relevance			Pre-li	ngest				Ing	est -	Stora	age						Sto	rage	- Acc	ess				
E-ARK Format			E	-ARK	SIP				E	-ARK	AIP	Х									Е	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EF	RMS				SMU	JRF S	FSB					Geod	data	
E-ARK Tools				(a			2			_														
	olkit			(ЕТ	e		(ETA			form										olkit				
	Toc			ıcer	Inpo	(q	ive			Plat			_	loc						To				
	tior	ule		rodi	g M	week) (GUU   tt To   t																		
	erva	ERMS Export Module ESSArch Tool for Producer (ETP) Universal Archiving Module SIP creator (E-ARK Web) ESSArch Tools for Archive (ETA) SIPZAIP (E-ARK Web) SIPZAIP (E-ARK Web) ESSArch Preservation Platform HDFS-Storage SOLR Index Search and Display GUI Order Management Tool Lily - Ingest Geoserver Geoserver AIPZDIP (E-ARK Web) Database Visualization Toolkit IP Viewer Peripleo Oracle (OLAP Viewer)																						
	Pres	ort l		ool f	Arch	r (E-	sloc	-AR	osit	rese	age	×	Dis	nage	,,			b Se	-AR	Visu			AP \	al/vi
	ase	Exp	Ė	ch T	'sal	ato	h T		Rep	h P	Stor	nde	au	Маі	ıges	rver		We	IP (E	ase	ver	oa	(01	oort
	Database Preservation Toolkit	<b>SMS</b>	RODA-In	SAr	ive	P cre	SAr	SIP2AIP	RODA Repository	SAr	HDFS-Storage	SOLR Index	arch	der	Lily - Ingest	Geoserver	QGIS	ARK	P2D	tab	P Viewer	Peripleo	acle	AIIS I
	۵	ER	RC	ES	Ď	SII	ES	SII		ES			Se	ō	5	Ğ	ŏ	Ę.	Ā	۵	П	Pe	Ō	บ
									Х		Х	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 semi-active 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

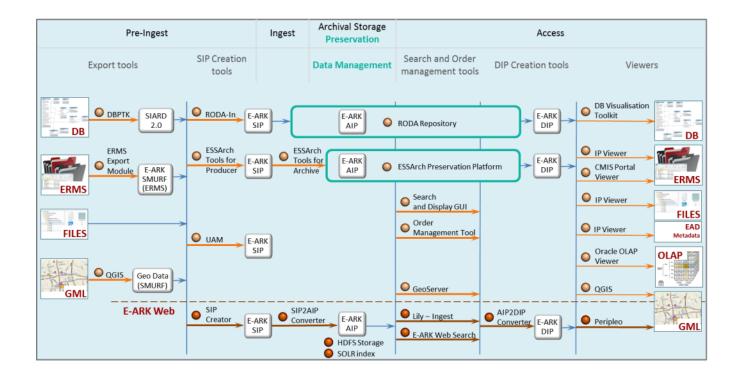
The RIP application follows a well-established software design pattern called "Pipes and Filters". This pattern makes use of a sequence of tasks (called "filters") that handle part of the entire processing workflow. Each filter is programmed to be simple and stateless. Streaming of data is used whenever 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.

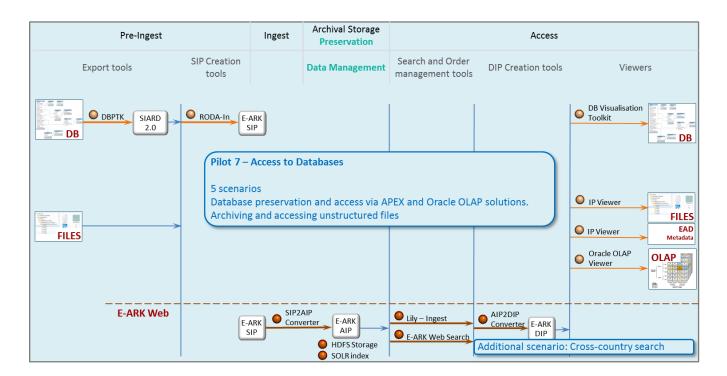
### Systems integration is the way forward

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

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	Dat	tabas	ses																			
Task leader	Nati	onal	Arch	nives	of H	unga	ry																	
Supported by	Dan	ish N	atio	nal A	rchiv	es																		
Scope	Rep	reser	ntatio	on of	not	less t	han	2 dat	tabas	ses o	f diff	eren	t size	s and	d con	nplex	ities	with	rest	ricte	d and	d ope	en	
	cont	ent.																						
Object													SIPs										_	
											-	_	ne ES											
Short description													abase							-				
						•							n an											
		_										-	inge					_	prot	otyp	e for	acce	ess w	ill
				ndly	web-	-base	ed ap	plica	tion			the	DIP s	pecit	icatio	on of	WP:	5	C.I					
Contacts	Nam	ne (T	itie)							E-m	aıı								Skyp	e				
Contact Person	Zolt	an L	ux							lux.	zolta	n@	mnl.	gov.	<u>hu</u>				lux.:	zolta	n1			
Pilot staff member	Józs	ef N	1eze	ei						mez	zei.jo	zset	f@m	nl.go	ov.h	<u>u</u>			jme	zei_	92			
OAIS Relevance		ı	Pre-l	nges	t			Ing	est -	Stor	age						Sto	rage	– Ac	cess				
E-ARK Formats		E-ARK SIP X E-ARK AIP X E-ARK DIP X																						
			S	IARD	2.0	Х			SMU	RF EI	RMS				SMU	JRF S	FSB	Х				Geo	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)	Database Visualization Toolkit	IP Viewer	Peripleo	Oracle (OLAP Viewer)	CMIS portal/viewer
Scenario 1	SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format																							
Scenario 2	SIP (	Creat	ion a	and I	ngest	of u	nstru	ıctur	ed fi	les														
Scenario 3	"Ext	ract	SIAR	D Pa	ckage	e fror	n Pre	eserv	ica/E	-ARI	( AIP													
Scenario 4	(APE	X/O	racle	BI a	ccess	)"																		
Scenario 5	"Sea	rch a	and p	orese	nt SI	ARD	base	d inf	orma	ation	with	E-A	RK ac	cess	tool	5								
Additional scenario	Cros	oss-country search with E-ARK Web (joint scenario with NAH)																						

# 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, R	RODA	-In, S	SIP2A	IP (E	-ARK	Web	b), H[	OFS-S	tora	ge												
Data	Hun	garia	ın Pro	osecu	ution	Offic	ce da	taba	se															
Description	Old	(not	norn	nalize	ed) da	ataba	ase ir	n CSV	expo	orts o	of DB	ASE 1	files.											
Data type	CSV	files																						
Metadata format	non	е																						
Quantity	mor	e the	en 30	0.00	0 cas	es ar	nd 50	0.00	0 nar	ne. (	1,6 G	B)												
OAIS Relevance		ı	Pre-li	ngest	:	Ingest - Storage Storage – Access																		
E-ARK Format			E	-ARK	SIP	X E-ARK AIP X E-ARK DIP																		
specifications	SIARD 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		X			X																		

Scenario 2	SIP Creation and Ingest of unstructured files  Create SIP from scanned documents of the Meeting minutes of the Central Coimmettee of the Hungarian																							
Description	Crea	ate S	IP fro	m sc	anne	d do	cum	ents	of th	е Ме	eting	g min	utes	of th	e Ce	ntral	Coin	nmet	tee o	of the	Hur	garia	ın	
	Soci	alist	Party	/. The	e ima	ge fi	les ar	e in	PDF 1	form	at wi	th EA	AD m	etada	ata. C	Creat	e bo	th E-	ARK a	and l	ocal S	SIPs a	ınd	
	inge	st th	ıem iı	nto B	27an	d Pr	eserv	ica a	rchiv	al re	posit	ory.	Both	E-AF	≀K an	d loc	al Al	Ps ar	e ger	nerat	ed di	uring	the	
	inge	est.																						
OIAS relevance	Pre-	Inge	st, In	gest																				
Use-case	Oth	er (E	xtrac	t and	Inge	st cc	mpu	ter f	iles f	rom :	simp	le file	e-syst	tem)										
E-ARK specifications	E-AF	RK SI	P, E-	ARK A	ΝIP																			
E-ARK Tools	ROD	)A-In	, SIP2	2AIP	(E-AF	K W	eb), I	HDFS	-Sto	age														
Data	Scar	nned	mee	ting ı	minu	tes o	f the	Cen	tral (	Comr	nitte	e of t	he H	unga	rian	Socia	alist F	arty						
Description	Scar	nned	docu	ımen	ts in	file s	yste	ms ir	) PDF	file	and c	orre	spon	ding	meta	data	(EA	) )						
Data type	<b>!</b>		files																					
Metadata format	EAD	)					-																	
Quantity	123.225 files. (101 GB)																							
OAIS Relevance			Pre-li	ngest				Ing	est -	Stora	age						Sto	rage	– Ac	cess				
E-ARK Format			E	-ARK	SIP	Х			E	-ARK	AIP	Х									Е	-ARK	DIP	
specifications			S	IARD	2.0				SMU	RF EI	RMS				SMU	JRF S	SFSB	Х				Geod	lata	
E-ARK Tools				Ъ)			۷)			_														
	olki			. (ЕТ	le		(ET/			form										olki				
	10			rcer	oqn	(q	ive			Plat			_	0						10				İ
	ıtior	n e		rodi	g M	We	Arch	eb)		ion			J9 /	μŢ					eb)	tior			rer)	<u>بر</u>
	Itodu Itodu																							
	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) SIPZAIP (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 AIPZDIP (E-ARK Web) Database Visualization Toolkit IP Viewer Peripleo Oracle (OLAP Viewer)																							
	se F	odx:	⊆	h To	sal A	ator	h To	) (E	Sep	h Pr	tore	ge	and	Man	- Ingest	ver		Web	P (E	se V	'er	0	(OL	orta
	taba	MS E	RODA-In	Arc	iver	cre	Arc	2AIF	DA	Arc	HDFS-Storage	SOLR Index	ırch	der F	' - In	Geoserver	IS	RK )	2DI	taba	IP Viewer	Peripleo	acle	IS p
	Dat	ERI	RO	ESS	Uni		ESS		RO	ESS		SO	Sea	Orc	Lily	Ge	QGIS	E-A	AIP	Dat	۱P۱	Per	Ora	S
			Х			X		Х			Х													

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.

OIAS relevance	Acc	ess																						
Use-case	Oth	er (A	r (Access database via APEX and Oracle BI)																					
E-ARK specifications	E-Al	RK AI	P, E-	ARK I	DIP																			
E-ARK Tools	HDF	S-Sto	orage	, Lily	/ – In	gest,	, E-Al	RK W	'eb (S	Searc	h), A	P2D	IP (E-	ARK	Web	)	,	DBV	TK					
Data	Hun	garia	n Pr	osecı	ution	Offic	ce da	itaba	se															
Description	Old	(not	norn	nalize	ed) da	ataba	ase ir	า CSV	exp	orts o	of DB	ASE 1	files.											
Data type	CSV	files																						
Metadata format	non	е																						
Quantity	mor	e the	e then 300.000 cases and 500.000 name. (1,6 GB)																					
OAIS Relevance		Pre-Ingest Ingest - Storage Storage – Access																						
E-ARK Format		E-ARK SIP E-ARK AIP X E-ARK DIP X							Х															
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	Х			Х			Х	Х		Х			

Scenario 4	Sea	arch and present SIARD based information with E-ARK access tools																						
Description	Acce	ess database information of the Hungarian Prosecution Office in SIARD format using HADOOP based search																						
	and	access with HIVE Lily Presentation in local environment.																						
OIAS relevance	Acce	cess																						
Use-case	Acce	ess d	ata w	ith C	)LAP	via o	racle	<del>)</del>																
E-ARK specifications	E-AF	RK AI	P, E-	ARK [	OIP																			
E-ARK Tools	HDF	S-Sto	orage	, Lily	/ – In	gest,	, E-AF	RK W	'eb (S	Searc	h), A	IP2DI	IP (E-	ARK	Web	)	,	DBV	TK					
Data	Hun	garia	n Pro	osecu	ution	Offic	ce da	taba	se															
Description	Old	(not	not normalized) database in CSV exports of DBASE files.																					
Data type	CSV	files	iles																					
Metadata format	non	e	:																					
Quantity	mor	ore then 300.000 cases and 500.000 name. (1,6 GB)																						
OAIS Relevance		Pre-Ingest Ingest - Storage Storage – Access																						
E-ARK Format			Е	-ARK	SIP				Е	-ARK	AIP	X									Е	-ARK	DIP	Х
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	Х			Х			Х	Х				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

		cialist Party. The image files are in PDF format with EAD metadata in E-ARK Web HDFS storage and eservica. Create both E-ARK and local DIPs.																						
OIAS relevance	Acce		.a. Ci	eate	ווטט	I E-A	NN di	10 10	Cai D	IPS.														
Use-case			atab:	ases	via S	OLD /	'no c	۵۱۱																
USE-Case								• •	-toro	~~ ~	d fra		aala	cc.t.o	C	: n ا		d fo			f	II +ov	امما د	٥.,
								טרט:	stora	ge ar	iu irc	)111 10	Cais	syste	111. 30	JLKI	s use	20 10	r seai	CH U	ie iu	ıı tex	t ina	ex
E ADV annaitiantiana	_			the c		nent	5.																	
E-ARK specifications				ARKI		/= 4	DI/ 14							1 /6		١		C-1 1						
E-ARK Tools																			iewr					
Data			ned meeting minutes of the Central Committee of the Hungarian Socialist Party																					
Description			ned documents in file systems in PDF file and corresponding metadata (EAD)																					
Data type	PDF,	/JPG files (representations)																						
Metadata format	EAD	D																						
Quantity	123.	23.225 files. (101 GB)																						
OAIS Relevance		- 1	Pre-l	ngest	t			Ing	est -	Stora	age						Sto	rage	– Acc	cess				
E-ARK Format			Е	-ARK	SIP				Е	-ARK	AIP	Х									E-ARK DIP X			
specifications			S	SIARD	2.0				SMU	RF EF	RMS		SMURF SFSB X Geodata											
E-ARK Tools	يز			TP)			<u>8</u>			E										ij				
	n Toolk			ucer (E	lodule	(qe	hive (E1			Platform			5	100						n Toolkit				
	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)  SOAR 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)	al/viewer																
	Database	ERMS Exp	RODA-In	ESSArch To	Universal,	SIP creato	ESSArch To	SIP2AIP (E	RODA Repository	ESSArch Pi	HDFS-Storage	SOLR Index	Search and	Order Mar	Lily - Ingest	Geoserver	ggis	E-ARK Web Search	AIP2DIP (E	Database '	IP Viewer	Peripleo	Orade (OL	CMIS portal/viewer
											Х	Х			Χ			Х	Х		Χ			

Additional scenario	Cross-country search with	E-ARK Web (joint scenario v	with NAS)								
Description	The SOLR index and E-ARK \	The SOLR index and E-ARK Web infrastructure theoretically makes it possible to perform a federated search									
	over more than one archive	over more than one archive. When the SOLR index of the other archival institution can be "seen" by the search									
	engine (e.g. one institution	engine (e.g. one institution has access rights to the others SOLR) then it can make a common list of the result.									
	The National Archives of Slo	ovenia and the National Arcl	hives of Hungary both have	an E-ARK implementation at							
	their pilot sites. This scenar	heir pilot sites. This scenario is a simple feasibility study of cross-country search.									
OIAS relevance	ccess										
Use-case	earch 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 DIP										
specifications	SIARD 2.0 SMURF ERMS SMURF SFSB Geodata										

|--|

## **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
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.
			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	OK
Issues	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	OK
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
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

#### 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.

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

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



NARA & SIARD 2

### **VERIFICATION OF FLAT FILE RECORDS**

#### **Fixed-Length Verification**

Counts the number of records in which the width of the value is less than the width of that field

#### Range Verification

Matches the values in a field against a consecutive set of values

#### **Distinct Values Verification**

Counts the Frequency of all values in a field Code List Verification

Matches the values in a Field against a list of acceptable values

#### Min/Max Verification

Identifies the lowest and highest value in a field

#### **Mandatory Verification**

Counts the number of records in which the field is blank

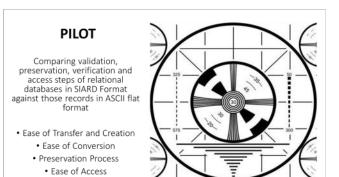
#### **Primary Key Verification**

Identifies the number of non-unique values in field or combination of fields

#### **Datatype Verification**

Shows the values in a field that do not match the data type designated in the AERIC Record Layout





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):

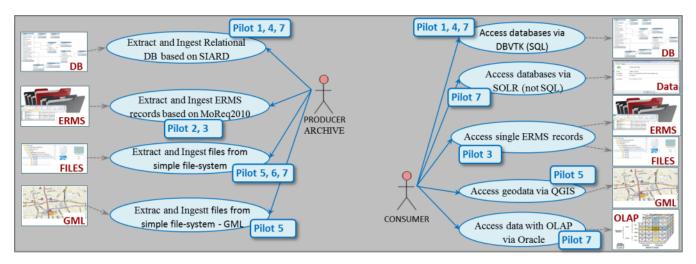
The overall objective of this work package is to ensure that the scenarios implemented at 7 identified pilot sites are both realistic and relevant. That is, that they bring together a meaningful subset at each site of the use cases that define establish a general model of the E-ARK service.

# Project level pilot success evaluation

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

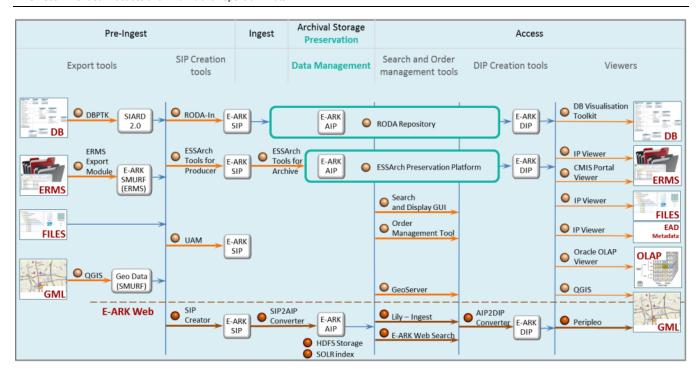
No#	Requirement	MoSCoW	Comment
7.2	The whole E-ARK full-scale pilot is successful if all the high-level E-ARK use cases are piloted in at least one of the pilots	М	
7.3	The whole E-ARK full-scale pilot is successful if all of the core E-ARK tools are piloted in at least one of the pilots	М	
7.4	The whole 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	•
		Pilot 7	Scenario 1	
		External	NARA,	
		evaluation	MinHAP, SFA	
	Extract and Ingest ERMS records based on MoReq2010	Pilot 2	Scenario 1-3	./
		Pilot 3	Scenario 1,3	•
		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	/
	- 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	/
		Pilot 5	Scenario 1,3	•
		Pilot 6	Scenario 1	
		Pilot 7	Scenario 1-2	
Access	Access databases via DBVTK (sql)	Pilot 4	Scenario 1-4	/
		Pilot 1	Additional sc.	•
	Access databases via SOLR (no-sql)	Pilot 5	Scenario 3	/
		Pilot 7	Scenario 3-5	<b>V</b>
	Access single ERMS records	Pilot 3	Scenario 2,4	./
		Pilot 2	Additional sc.	
	Access geodata via qgis	Pilot 5	Scenario 2,4	<b>/</b>
	Access data with OLAP via oracle	Pilot 7	Sceanrio 4	<b>/</b>

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	•
		Pilot 7	Scenario 1	
		External	NARA,	
		evaluation	MinHAP, SFA	
	ERMS Export Module	Pilot 1	Additional sc.	1
		Pilot 3	Additional sc.	•
	RODA-In	Pilot 5	Scenario 1	1
		Pilot 7	Scenario 1,2	•
	ESSArch Tool Producer (ETP)	Pilot 2	Scenario 1-3	1
	- Redesigned UI, E-ARK compatible version	Pilot 2	Additional sc.	•
		Pilot 5	Scenario 3	
	Universal Archiving Module	Pilot 3	Scenario 1,3	<b>√</b>
	SIP creator (E-ARK Web)	Pilot 7	Scenario 2	<b>✓</b>
Ingest	ESSArch Tools Archive (ETA)	Pilot 3	Scenario 1,3	1
		Pilot 5	Scenario 2	<b>~</b>
	SIP2AIP (E-ARK Web)	Pilot 5	Scenario 1,2	1
		Pilot 7	Scenario 1,2	<b>V</b>
	RODA Repository	Pilot 6	Scenario 1	<b>√</b>
	ESSArch Preservation Platform	Pilot 3	Scenario 1,3	<b>✓</b>
	HDFS-Storage	Pilot 7	Scenario 1-5	<b>/</b>

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

	Use Case	Pilot	Scenario	Successful?
Information		Pilot 2	Scenario 1-3	./
Package format	E-ARK SIP	Pilot 3	Scenario 1,2	•
specification		Pilot 5	Scenario 1,2	
	(Supplier Information Package)	Pilot 6	Scenario 1	
	Į.	Pilot 7	Scenario 1,2	
		Pilot 2	Scenario 1-3	/
	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	/
		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.	<b>√</b>
E-ARK SMURF SFSB	Pilot 5 Pilot 6 Pilot 7	Scenario 1-4 Scenario 1 Scenario 2,5+D14	<b>√</b>
E-ARK SMURF Geodata	Pilot 5	Scenario 1-4	<b>/</b>

# 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	<b>√</b>
Scenario 1	Extract records from MS SQL Server database containing 50-60 tables and about 90.000 records. (95% success rate)	J
Scenario 2	Extract records from MySQL database about 5 million records.(95% success rate)	<b>√</b>
Scenario 3	Extract records from MS SQL Server database containing documents. (95% success rate)	<b>√</b>
Scenario 4	Extract records from MS SQL Server database containing documents. (95% success rate)	<b>√</b>
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).	<b>√</b>
	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	<b>√</b>
Scenario 2	Ingest around 5 GBs of EDRMS data from public producer converted into Noark 4 output	J
Scenario 3	Ingest around 335.000 registered persons (105 MB) from the national registry of licenced hunters.	J
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)	<b>√</b>
	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.	<b>√</b>

Scenario 1	Extract records from EDRM, create and ingest SIP of different documents of Ministry of Justice with different retention period (95% success rate)	<b>√</b>
Scenario 2	Provide access to archived records of Ministry of Justice (95% success rate)	<b>√</b>
Scenario 3	Extract records from EDRM, create and ingest SIP of different documents of Ministry of Justice with different retention period (95% success rate)	<b>√</b>
Scenario 4	Provide access to archived records of Ministry of Justice (95% success rate)	<b>&gt;</b>
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.	<b>✓</b>
Scenario 1	Exporting records from database for more than 12 000 business records from bespoke business system	<b>&gt;</b>
Scenario 2	Importing records to database for more than 12 000 business records from bespoke business system	<b>√</b>
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)	<b>&gt;</b>
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)	<b>✓</b>
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.	
Scenario 1	SIP creation, verification and ingest of more than 1000 records with a vector geodata layer. (90% success rate)	<b>√</b>
Scenario 2	Finding, accessing, modifying and exporting a DIP containing a vector geodata layer of more than 1000 records. (90% success rate)	<b>√</b>
Scenario 3	SIP creation, verification and ingest of more than 200 records with a vector geodata layer. (90% success rate)	<b>\</b>
Scenario 4	Finding, accessing, modifying and exporting a DIP containing a vector geodata layer of more than 200 records. (90% success rate)	<b>√</b>
Pilot 6	Test the E-ARK compatible RODA Repository in a pilot environment. (Yes/No)	<u> </u>
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)	<b>✓</b>
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
	<del></del>	

	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)	<b>&gt;</b>
Scenario 1	Create SIP and Ingest more than 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)	<
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)	<b>/</b>
Scenario 3	Provide access for more than 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)	<b>√</b>
Scenario 4	Provide access for more than 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)	<b>/</b>
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)	<b>/</b>

# **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: <a href="http://eark-project.com/resources/project-deliverables/87-d24docs-p1-1">http://eark-project.com/resources/project-deliverables/87-d24docs-p1-1</a>

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: <a href="http://eark-project.com/resources/general-model">http://eark-project.com/resources/general-model</a>

# 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.

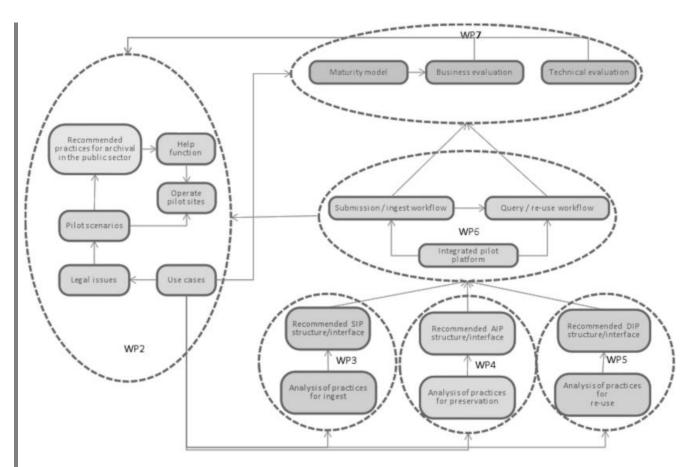


Figure 1: E-ARK - Overall Approach

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.

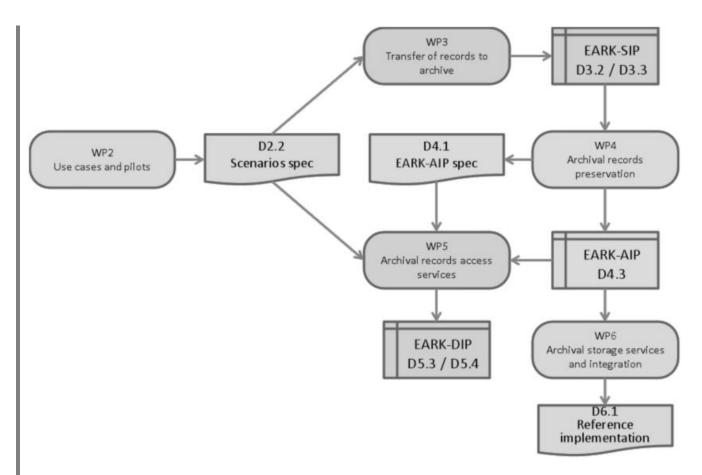


Figure 2: E-ARK Technical Integration

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.

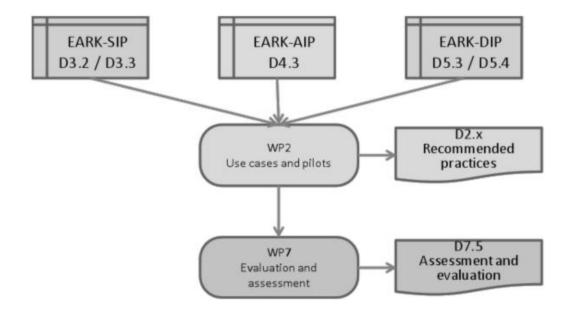


Figure 3: Pilot Workflows

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, D03.3) 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 user-friendly 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