D.2.3 Detailed Pilots Specification

DOI: 10.5281/zenodo.1170609

Grant Agreement Number:	620998
Project Title:	European Archival Records and Knowledge Preservation
Release Date:	10th February 2018

Contributors

Name	Affiliation
István Alföldi	National Archives of Hungary
István Réthy	National Archives of Hungary
Sven Schlarb	Austrian Institute of Technology
Clive Billenness	University of Brighton
David Anderson	University of Brighton
Janet Anderson	University of Brighton
Anders Bo Nielsen	Danish National Archive
Phillip Mike Tømmerholt	Danish National Archive
Alex Thirifays	Danish National Archive
Hans Fredrik Berg	National Archives of Norway
Terje Pettersen-Dahl	National Archives of Norway
Arne-Kristian Groven	National Archives of Norway
Tarvo Kärberg	National Archives of Estonia
Karin Oolu	National Archives of Estonia
Raivo Ruusalepp	Estonian Business Archive
Ats Rand	Estonian Business Archive
Gregor Završnik	National Archives of Slovenia
Joze Skofljanec	National Archives of Slovenia
Miguel Ferreira	Keep Solutions
Zoltán Lux	National Archives of Hungary
Mezei József	National Archives of Hungary

Table of Contents

1. 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 activities 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.

Project Timescales

The E-ARK Project started on 1 February 2014 (M1) and is scheduled to be completed on 31 January 2017 (M36)

In this document, references are made to Project Months (Mnn), and this table shows how these relate to calendar months:

M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36
Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16	Oct-16	Nov-16	Dec-16	Jan-17

Work Package 2

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
  - 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 the 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 which scenario which level of activity is needed in order to bridge the gap of the currently existing solutions (e.g. integration, software development, interface definition).

In order to make the E-ARK service as widely as possible to demonstrate the functionality of the service built on D2.1 from the pilot will be finalized around the pilot sites. In order to plan ahead for a pilot project previously identified three 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 directly not 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.

This document corresponds to the deliverable:

D2.3) Detailed pilot requirements
This report will detail the technical and other requirements for each of the pilot sites. [month 26]

Objective of this deliverable

The objective of this document is to define the pilots and the requirements in a way that that those responsible for each pilot and staff involved know in detail

- what task to do in the corresponding steps,
- according to what schedule,
- using which tools,
- providing what kind of results
- and what type of documentation.

Note that this document covers pilot activities belonging to the *E-ARK Full-scale pilots (1)*. The *Additional project pilots (2)* and *External validation activities (3)* will be defined in a separate document.

2. Pilot planning

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.

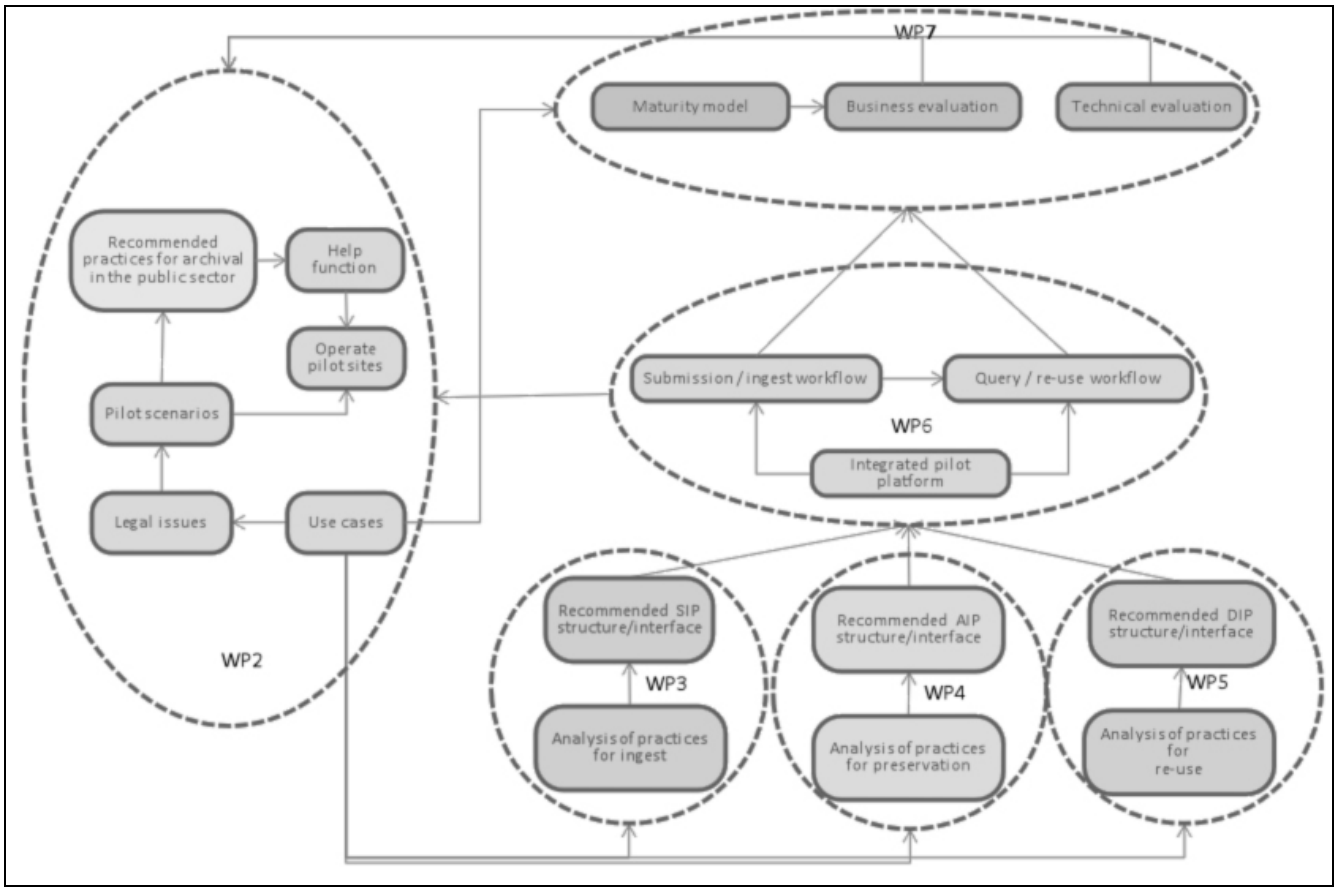

Figure 1: E-ARK – Overall Approach

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 re-use 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.

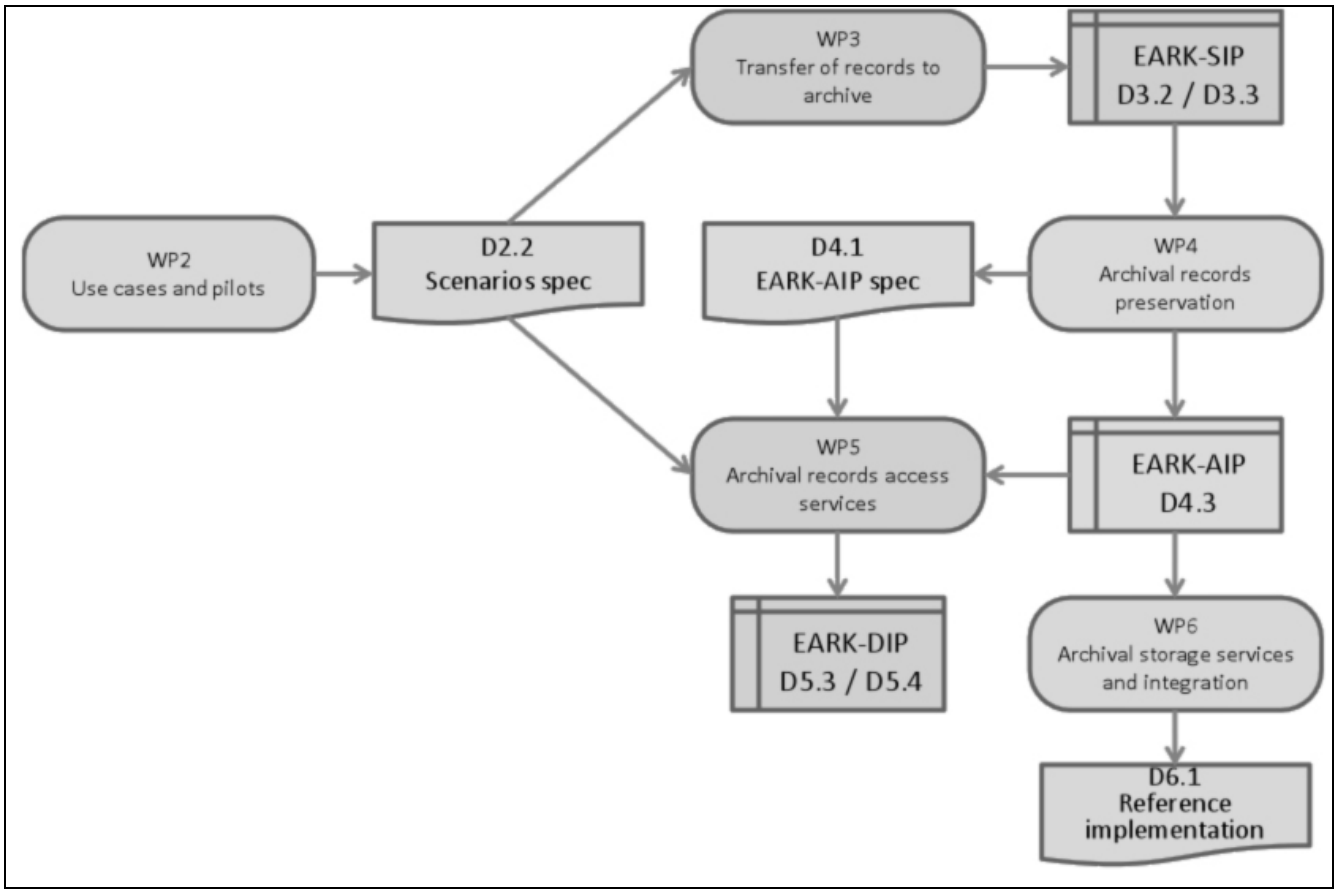

Figure 2: E-ARK Technical Integration

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.

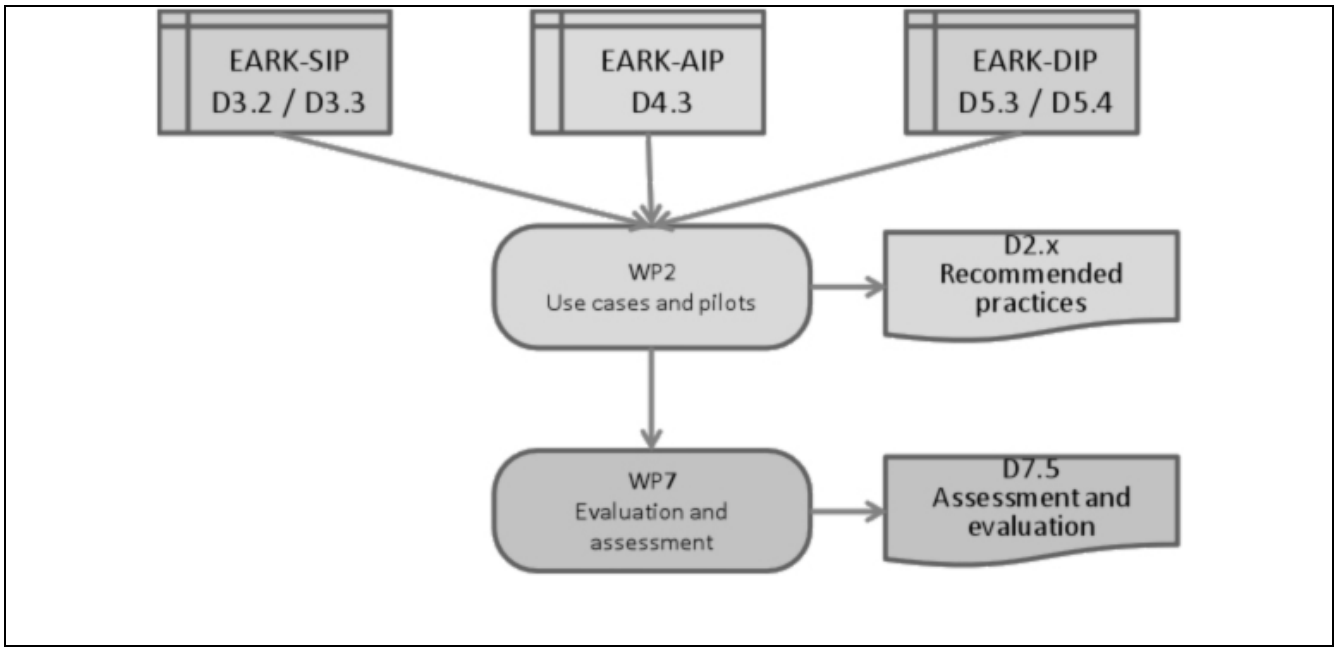

Figure 3: Pilot Workflows

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.

2.1 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 transfer-ingest 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

2.2 Schedule

The planning of the pilots goes in parallel with the design and implementation of E-ARK tools and interfaces. For this reason the pilot planning has two phases:

- the pilot definition phase, and
- the requirements phase.

The pilot definition phase aims to set a basis for the pilots while the requirements gathering phase adds further information, finalizes the pilot infrastructure and activities and defines the requirements by and towards the pilots.

In the pilot definition phase the plan is to define

- the scope and objective,
- the pilot site responsible and other pilot members,
- pilot scenarios,
- E-ARK use-cases that shall be tested,
- OAIS processes the pilot shall implement,
- information package types (SIP, AIP, DIP) the pilot shall implement,
- E-ARK tools and interfaces used,
- and pilot data (if available).

The definition phase concluded at the E-ARK Technical Meeting, 14-16. December 2015.

The requirements gathering phase will

- finalize pilot definitions with detailed process step and data description,
- set requirements by the project towards the pilots (like documentation requirements),
- set requirements by the pilots towards other project members (like user manual and helpdesk availability),
- define schedule of the planned tasks.

This document is the result of the cooperation and work of pilot responsible and tool developers in both phases.

3. Structure of this Document

The document consists of two main parts

- Pilot Definition
  This part defines the pilot scenarios, use-cases, pilot data, technical infrastructure and pilot processes in detail.
- Further Pilot Requirements

  This part describes further requirements by the project towards the pilots (like documentation requirements) and requirements by the pilot sites towards the other project members (like requirements about user manuals and help-desk availability).

Requirement format

- No#
- Requirement (short and below detailed description if necessary)
- MoSCoW
- Acceptance criteria

MoSCoW (http://en.wikipedia.org/wiki/MoSCoW_method)

For the purpose of the E-ARK project and this internal deliverable, T3.1 decided to use the MoSCoW method for indicating the requirement levels. A brief explanation of the meaning of each term is below.

Letter	Meaning	Description
M	MUST	Describes a requirement that must be satisfied in the final solution for the solution to be considered a success.
S	SHOULD	Represents a high-priority item that should be included in the solution if it is possible. This is often a critical requirement but one which can be satisfied in other ways if strictly necessary.
C	COULD	Describes a requirement which is considered desirable but not necessary. This will be included if time and resources permit.
W	WON'T	Represents a requirement that stakeholders have agreed will not be implemented in a given release, but may be considered for the future. (note: occasionally the word "Would" is substituted for "Won't" to give a clearer understanding of this choice).

Example

No #	Requirement	MoSCoW	Comment
6.1	Each pilot site should execute the pilot according to the following schedule	M	

4. Pilot Overview

4.1 Pilots and the E-ARK General Model

4.1.1 OAIS Relevance

The following table provides OAIS process – Pilot cross-reference information. This table is part of the EARK General Model v2.0.

Pilot – OAIS Process cross reference table	
E-ARK	General Model v2.0

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

■ Focus of the pilot

■ Elements also used/tried within the pilot

4.1.2 Pilot Scenarios and E-ARK Use cases

WP3-4 (Pre-Ingest/Ingest) and WP5 (Access) have been focusing on the following business use-case scenarios in the scope of the E-ARK project:

Pre-Ingest/Ingest use-cases

- Extract and Ingest relational database based on SIARD 2.0
- Extract and Ingest ERMS records based on MoReq2010
- Extract and Ingest computer files from simple file-system – GML
- Extract and Ingest computer files from simple file-system

Access use-cases

- Access databases via Sofia (SQL)
- Access databases via SOLR (not SQL)
- Access single ERMS records via Alfresco CMS
- Access geodata via QGIS
- Access data with OLAP via Oracle

The following table provides cross-reference information about scenarios defined in the full-scale pilots and the above listed E-ARK use-cases. This table is part of the EARK General Model v2.0.

Pilot Scenario – E-ARK Use-cases cross reference table	
E-ARK	**General Model v2.0**

	Pre-Ingest / Ingest					Access					
Pilot Scenario (Pilot #. Scenario #)	**Extract and Ingest relational database based on SIARD 2.0**	**Extract and Ingest ERMS records based on MoReq2010**	**Extract and Ingest computer files from simple file-system – GML**	**Extract and Ingest computer files from simple file-system**	**Other**	**Access databases via Sofia (SQL)**	**Access databases via SOLR (not SQL)**	**Access single ERMS records via Alfresco CMS**	**Access geodata via QGIS**	**Access data with OLAP via Oracle**	**Other**
1.1 Extracting records from database	■										
1.2 Extracting records from database	■										
1.3 Extracting records from database	■										
1.4 Extracting records from database	■										
2.1 SIP Creation and Ingest of unstructured records		■									
2.2 SIP Creation and Ingest of unstructured records		■									
2.3 SIP Creation and Ingest of structured records					■						
3.1 Extract records from EDRM and ingest into Preservica		■									
3.2 Extract records from EDRM and ingest into Preservica		■									
3.3 Provide access to records from governmental institution through CMIS interface								■			■
3.4 Provide access to records from governmental institution through CMIS interface											■
4.1 SIP Creation and Ingest of business records from bespoke business system	■										
5.1 SIP Creation and Ingest of records with Geodata			■		■						

5.2 SIP Creation and Ingest of records with Geodata									■		■
5.3 Search and Access information using Geadota			■		■						
5.4 Search and Access information using Geadota									■		■
6.1 Automatic ingest of records from a semi-active archival management system					■						
6.2 Automatic ingest of records from a semi-active archival management system		■									
7.1 SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format	■										
7.2 SIP Creation and Ingest of unstructured files				■							
7.3 Extract SIARD Package from Preservica / E-ARK AIP (APEX/OWB access)										■	
7.4 Search and present SIARD based information with E-ARK access tools (HADOOP, HIVE Presentation)						■					
7.5 Access information from unstructured files							■				

4.1.3 Pilots and E-ARK Tools

The following table provides cross-reference information about scenarios defined in the full-scale pilots and E-ARK tools developed by WP3-WP5 or in WP6 as part of the integrated prototype. This table is part of the EARK General Model v2.0.

Scenarios - Tools

E-ARK **General Model v2.0**

Pilot	Scenario	Pre-Ingest							Ingest - Storage				Storage - Access																				
		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	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Orade (OLAP Viwer)	CMIS portal/viewer			
Pilot 1 (DNA)	1. Extracting records from database	■																														
	2. Extracting records from database	■																														
	3. Extracting records from database	■																														
	4. Extracting records from database	■																														
Pilot 2 (NAN)	1. SIP Creation and Ingest of unstructured records				■	■					■																					
	2. SIP Creation and Ingest of unstructured records				■	■					■																					
	3. SIP Creation and Ingest of structured records				■	■					■																					
Pilot 3 (NAE)	3.1 Extract records from EDRM and ingest into Preservica		■				■																									
	3.2 Extract records from EDRM and ingest into Preservica		■				■																									
	3.3 Provide access to records from governmental institution through CMIS interface																							■								■
	3.4 Provide access to records from governmental institution through CMIS interface																							■								■
Pilot 4 (EBA)	4.1 SIP Creation and Ingest of business records from bespoke business system	■		■																												
Pilot 5 (SNA)	5.1 SIP Creation and Ingest of records with Geodata				■	■					■																					
	5.2 SIP Creation and Ingest of records with Geodata				■	■					■																					
	5.3 Search and Access information using Geadota												■	■	■	■	■	■	■			■				■	■	■				
	5.4 Search and Access information using Geadota												■	■	■	■	■	■	■			■				■	■	■				
Pilot 6 (KEEPS)	6.1 Automatic ingest of records from a semi-active archival management system									■																						
	6.2 Automatic ingest of records from a semi-active archival management system									■																						
Pilot 7 (NAH)	7.1 SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format	■		■					■			■																				
	7.2 SIP Creation and Ingest of unstructured files			■					■			■																				
	7.3 Extract SIARD Package from Preservica / E-ARK AIP (APEX/OWB access)											■					■		■	■	■		■									
	7.4 Search and present SIARD based information with E-ARK access tools											■					■		■	■	■									■		
	7.5 Access information from unstructured files											■					■		■	■		■			■							

4.2 Pilot Card concept

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 to reach detailed pilot information descriptions and corresponding documents.

Pilot Card example

Pilot #1	SIP Creation on relational databases
Status	• Pilot defined √ • Installation started √ • Installation ready - • Pilot execution started - • Pilot execution completed -

OAIS relevance	Pre-Ingest	Ingest	E-ARK SIP	E-ARK AIP	Preservation	Data Management	E-ARK DIP	Access
	X							

Task leader	Danish National Archives		
Supported by	Magenta		
Contacts	Name (Title)	e-mail	Skype
Contact Person	Phillip Mike Tømmerholt	pmt@sa.dk	philliptommerholt_rigsarkivet
Contact Person	Anders Bo Nielsen	abn@sa.dk	
Pilot staff members			
Scope	The scope of this Pilot is to test the E-ARK SIP Creation tool with not less than 4 databases of different sizes and complexities (one contains several million records)		
Object	Creating SIPs for relational databases using the tool created in WP3, T3.3: SIP Creation Tools, for further evaluation		
Short description	The goal of the pilot is to create SIPs in EARK-SIP format of each selected database with the DBextract tool. After quality assurance on each SIP, a feedback will be given to WP3		
Timeframe	M28-M33		
Preconditions	M03.3, M03.4 (DoW)		

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	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)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
X																												

Pilot Scenarios	
Scenario 1	Extracting records from database (Data Set 1) - database with no documents
Scenario 2	Extracting records from database (Data Set 2) - database with no documents (large)
Scenario 3	Extracting records from database (Data Set 3) - database with documents
Scenario 4	Extracting records from database (Data Set 4) - database with documents (large)
Links	Process and use case information / Pilot definition
	Test data specification / Pilot documentation

All the information a Pilot Card holds is included in the Pilot Definition tables in more detail therefore we do not include the Pilot Cards in this document.

5. Pilot Definition

This section defines the pilots. Pilot definition includes a full description of pilot scope, scenarios, pilot data, infrastructure and activities for each pilot.

The Pilot Definition description follows this structure

Pilot

- **Scenario**
  - Business use-case (from General Model)
  - Used Information packages
  - 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
      - E-ARK (and local) tools usage details

Requirements related to pilot definition

5.1	Each pilot site should execute the pilot according to the following pilot definition	M	
5.2	Each pilot should perform the scenarios defined in its pilot definition	M	
5.3	Each pilot should create and use the E-ARK information packages (E-ARK SIP, AIP, DIP) selected in its pilot definition	M	
5.4	Each pilot should perform a scenario according to the E-ARK high-level use case selected in the pilot definition.	M	Minor deviations acceptable.
5.5	Each pilot should use the E-ARK and E-ARK web tools selected in the pilot definition.	M	Minor deviations acceptable.
5.6	Each pilot should perform the process steps defined in the pilot definition.	M	Minor deviations acceptable.

Note that some of the datasets are not yet fully defined either because the archive is waiting for the official approval of the data provider or because the data set to be used is one of the production data sets available from a producer when the pilot is actually executed.

The source of the following pilot definition tables are the Excel files that were created and worked on in collaboration with the pilot responsible. The tables included in this document correspond to version 1.10 of the seven pilot definition excel files.

5.1 Pilot 1 - SIP Creation on relational databases

Pilot 1	**SIP Creation on relational databases**
Task leader	Danish National Archives
Supported by	Magenta
Scope	The scope of this Pilot is to test the E-ARK SIP Creation tool with not less than 4 databases of different sizes and complexities (one contains several million records)
Object	Creating SIPs for relational databases using the tool created in WP3, T3.3: SIP Creation Tools, for further evaluation
Short description	The goal of the pilot is to make four successful data extractions from live authentic databases into the SIARD 2.0 format.
Timeframe	M28-M33
Preconditions	M03.3, M03.4 (DoW)

Contacts	**Name (Title)**	**E-mail**	Skype
Contact Person	Anders Bo Nielsen	abn@sa.dk	
Pilot staff member	Phillip Mike Tømmerholt	pmt@sa.dk	philliptommerholt_rigsarkivet
Pilot staff member			

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
	E-ARK SIP							E-ARK AIP				E-ARK DIP															
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Orade (OLAP Viwer)	CMIS portal/viewer
	X																											

Pilot Scenarios	
Scenario 1	Extracting records from database (Data Set 1) - database with no documents
Scenario 2	Extracting records from database (Data Set 2) - database with no documents (large)
Scenario 3	Extracting records from database (Data Set 3) - database with documents
Scenario 4	Extracting records from database (Data Set 4) - database with documents (large)
Success Criteria	
General	
Scenario 1	Extract records from Ms SQL database more then 100 GB. (95% success rate)
Scenario 2	Extract records fromlarge database containing documents. (To be finalized after data provider approval)
Scenario 3	Extract records from Ms SQL database containing 50-60 tables and about 90.000 records. (95% success rate)
Scenario 4	Extract records from MySQL database about 5 million records.(To be finalized after data provider approval)

Pilot 1	Scenarios
Scenario 1	**Extracting records from database (Data Set 1)**
Description	Extracting records from database containing documents. The DNA will go to the producers 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.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
	E-ARK SIP							E-ARK AIP				E-ARK DIP																	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
	X																											
Use-case	**Extract and Ingest relational database based on SIARD 2.0**																											
Note																												

Scenario 2	**Extracting records from database (Data Set 2)**
Description	Extracting records from database containing documents. The DNA will go to the producers 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.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
	E-ARK SIP							E-ARK AIP				E-ARK DIP																	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
	X																											
Use-case	**Extract and Ingest relational database based on SIARD 2.0**																											
Note																												

Scenario 3	**Extracting records from database (Data Set 3)**
Description	Extracting records from database containing documents. The DNA will go to the producers 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.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																		
	E-ARK SIP							E-ARK AIP				E-ARK DIP																		
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
	X																												

Use-case	**Extract and Ingest relational database based on SIARD 2.0**
Note	

Scenario 4	**Extracting records from database (Data Set 4)**
Description	Extracting records from database containing documents. The DNA will go to the producers 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.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																		
	E-ARK SIP							E-ARK AIP				E-ARK DIP																		
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
	X																												

Use-case	**Extract and Ingest relational database based on SIARD 2.0**
Note	

Pilot 1	Pilot Data

Information Packages (IP)	IP		Note
	E-ARK SIP	X	Not the full SIP, only the SIARD2.0-file
	non E-ARK SIP		
	E-ARK AIP		
	non E-ARK AIP		
	E-ARK DIP		
	non E-ARK DIP		

Pilot data description	
Data Set 1	**Administrative system from The Danish National Archives**
Description	Database containing information about incoming research data, and public deliveries of research data.
Data type	Microsoft SQL Server
Metadata format	Not relevant
Quantity	small
Data Set 2	**(Under confirmation by the data Provider)**
Description	Database containing documents (large)Since this pilot operates in the existing pre-ingest processes in t
Data type	
Metadata format	Not relevant
Quantity	large
Data Set 3	**Health system from The Danish National Serum Institute**
Description	Database containing information from reported diseases at a national level. 50-60 tables and about 90.
Data type	Microsoft SQL Server 2008
Metadata format	Not relevant
Quantity	small
Data Set 4	**Commercial system (Under confirmation by the data Provider)**
Description	Database containing 5 million records.
Data type	MySQL
Metadata format	Not relevant
Quantity	large

OAIS Process				Pre-Ingest					
Main Process Steps	**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	**Extracting records from database (Data Set 1) - database with no documents**								
Used E-ARK tool				Database Preservation Toolkit					
Used local tools			Manual						
Perfomer (actor)			Archivist	SIP provider + DNA + Magenta					
Prelimineries and Start condition									
Input			Data Source	Selected DB					
Output			DB selected	SIARD2.0 file					
Scenario 2	**Extracting records from database (Data Set 2) - database with no documents (large)**								
Used E-ARK tool				Database Preservation Toolkit					
Used local tools			Manual						
Perfomer (actor)			Archivist	SIP provider + DNA + Magenta					
Prelimineries and Start condition									
Input			Data Source	Selected DB					
Output			DB selected	SIARD2.0 file					
Scenario 3	**Extracting records from database (Data Set 3) - database with documents**								
Used E-ARK tool				Database Preservation Toolkit					
Used local tools			Manual						
Perfomer (actor)			Archivist	SIP provider + DNA + Magenta					
Prelimineries and Start condition									
Input			Data Source	Selected DB					
Output			DB selected	SIARD2.0 file					
Scenario 4	**Extracting records from database (Data Set 4) - database with documents (large)**								
Used E-ARK tool				Database Preservation Toolkit					
Used local tools			Manual						
Perfomer (actor)			Archivist	SIP provider + DNA + Magenta					
Prelimineries and Start condition									
Input			Data Source	Selected DB					
Output			DB selected	SIARD2.0 file					

5.2 Pilot 2 - SIP creation and ingest of records

Pilot 2	**SIP creation and ingest of records**
Task leader	National Archives of Norway
Supported by	
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
Short description	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.
Timeframe	M28-M33
Preconditions	M03.3, M03.4 (DoW)

Contacts	**Name (Title)**	**E-mail**	Skype
Contact Person	Arne-Kristian Groven	arngro@arkivverket.no	
Pilot staff member	Terje Pettersen-Dahl	tepe@arkivverket.no	
Pilot staff member			

OAIS Relevance	**Pre-Ingest**						E-ARK SIP	**Ingest - Storage**			E-ARK AIP	**Storage - Access**																	E-ARK DIP
							X				X																		
E-ARK Tools	**Database Preservation Toolkit**	**Alfrsco Export Module**	**RODA-in**	**ESSArch Tool Producer (ETP)**	**ESSArch Tools Archive (ETA)**	**UAM**	**SIP creator (E-ARK Web)**	**SIP2AIP (E-ARK Web)**	**RODA Repository**	**ESSArch Preservation Platform**	**HDFS-Storage**	**ICA-AtoM Catalogue**	**OMT - Search and Dsiplay GUI**	**Order Submission Service**	**OMT - Order Management Tool**	**Lily - Ingest**	**ESSArch Preservation Platform**	**E-ARK Web (Search)**	**AIP2DIP (E-ARK Web)**	**DBPTK**	**IP Viewer**	**DB Viewer (Sofia)**	**ERMS Viewer (Alfresco)**	**Single file Viewr**	**QGIS**	**Geoserver**	**Peripleo**	**Oracle (OLAP Viwer)**
				X	X					X																		

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

Success Criteria	
General	The following E-ARK tools will be tested in a pilot environment: ESSArch Tools Producer (ETP), ESSArch Tools Archive (ETA), ESSArch Preservation Platform (EPP). 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 (potentially) 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 dissminated during piloting. Another success criteria is if we succeed in specifying a conversion of EDRMS output produced in accordance with the national NOARK standard, into Moreq 2010. This is not directly linked to the ESSArch Tools which can handle content regardless of whether it is NOARK og Moreq 2010.
Scenario 1	SIP Creation and Ingest of unstructured records. (To be finalized after data provider approval)
Scenario 2	SIP Creation and Ingest of unstructured records. (To be finalized after data provider approval)
Scenario 3	SIP Creation and Ingest of unstructured records. (To be finalized after data provider approval)

Pilot 2	Scenarios
Scenario 1	**SIP Creation and Ingest of unstructured records (Data Set 1)**
Description	Extract unstructured records from EDRMS based on the Norwegian NOARK standard. Convert NOARK format to MoReq2010. Create SIP using ESSArch Tools. Ingest the SIP to the repository using ESSArch Preservation Platform, for further evaluation.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
							E-ARK SIP	X				E-ARK AIP		X														E-ARK DIP	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
				X	X					X																		

Use-case	**Extract and Ingest ERMS records based on MoReq2010**
Note	

Scenario 2	**SIP Creation and Ingest of unstructured records (Data Set 2)**
Description	Extract unstructured records from EDRMS based on the Norwegian NOARK standard. Convert NOARK format to MoReq2010. Create SIP using ESSArch Tools. Ingest the SIP to the repository using ESSArch Preservation Platform, for further evaluation

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
							E-ARK SIP	X				E-ARK AIP		X														E-ARK DIP	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
				X	X					X																		

Use-case	**Extract and Ingest ERMS records based on MoReq2010**
Note	

Scenario 3	**SIP Creation and Ingest of structured records (Data Set 3)**																																
Description	Extract data from database, create SIPs for structured records using ESSArch Tools, ingest the SIPs to the repository using ESSArch Preservation Platform, for further evaluation.																																
OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																						
							E-ARK SIP	X			E-ARK AIP	X																					E-ARK DIP	
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**	**ICA-AtoM Catalogue**	**OMT - Search and Dsiplay GUI**	**Order Submission Service**	**OMT - Order Management Tool**	**Lily - Ingest**	**ESSArch Preservation Platform**	**E-ARK Web (Search)**	**AIP2DIP (E-ARK Web)**	**DBPTK**	**IP Viewer**	**DB Viewer (Sofia)**	**ERMS Viewer (Alfresco)**	**Single file Viewr**	**QGIS**	**Geoserver**	**Peripleo**	**Oracle (OLAP Viwer)**						
				X	X					X																							
Use-case	**Other (please specify content type and presentation tool)**																																
Note	**Extract rows and tables from DB and create SIP.**																																

Pilot 2	**Pilot Data**		
Information Packages (IP)	**IP**		**Note**
	E-ARK SIP	X	
	non E-ARK SIP		
	E-ARK AIP	X	
	non E-ARK AIP		
	E-ARK DIP		
	non E-ARK DIP		

Pilot data description	
Data Set 1	**(Under confirmation by the data Provider)**
Description	This EDRMS data set will be real production data, picked at a later stage.
Data type	
Metadata format	
Quantity	

Data Set 2	**(Under confirmation by the data Provider)**
Description	This EDRMS data set will be real production data, picked at a later stage.
Data type	
Metadata format	
Quantity	

Data Set 3	**(Under confirmation by the data Provider)**
Description	The data set here is the national registry of licenced hunters containing data from the period 1985-1999.
Data type	
Metadata format	
Quantity	Containing 338.500 registred persons.

OAIS Process					Pre-Ingest				
Main Process Stepps	**Content definition Technical feasibility Legal issues etc.**	**Create/Review transfer agreement**	**Select data**	**Data Extraction**	**Manual compilation of non ERMS content**	**Metadata mapping**	**Create SIP**	**Post-packaging quality control**	**Submit SIP**
Scenario 3	**Structured data, not ERMS**								
Used E-ARK tool						ETP	ETP	ETP	ETP
Used local tools									
Perfomer (actor)						Producer, Archivist	Producer, Archivist	Producer, Archivist	Producer, Archivist
Prelemineries and Start condition									
Input							Metadata	E-ARK SIP	E-ARK SIP
Output						Metadata	E-ARK SIP	Quality report	
Scenario 1,2	**SIP Creation and Ingest of ERMS files**								
Used E-ARK tool						ETP	ETP	ETP	ETP
Used local tools			NOARK 5 extraction tool	NOARK 5 extraction tool					
Perfomer (actor)			Producer	Producer		Producer, Archivist	Producer, Archivist	Producer, Archivist	Producer, Archivist
Prelemineries and Start condition									
Input				Selected data		Extracted data	Metadata	E-ARK SIP	E-ARK SIP
Output			Selected data	Extracted data		Metadata	E-ARK SIP	Quality report	

OAIS Process				Ingest				
Main Process Stepps	**Receive SIP**	**Security check**	**Integrity check**	**Sintax check**	**Format check**	**Review and Approve SIP**	**Modify SIP**	**Generate and Store AIP**
Scenario 3	**Structured data, not ERMS**							
Used E-ARK tool	ETA		ETA, EPP			EPP		EPP
Used local tools		Local tools		Arkade 4	Arkade 4			
Perfomer (actor)	Archive	Archive	Archive	Archive	Archive	Archive		Archive
Prelemineries and Start condition								
Input	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP		E-ARK SIP
Output		check report	check report	check report	check report	Approved SIP		E-ARK AIP
Scenario 1,2	**SIP Creation and Ingest of ERMS files**							
Used E-ARK tool	ETA		ETA, EPP			EPP		EPP
Used local tools		(yes)		Arkade 4, ArkN4	Arkade 4			
Perfomer (actor)	Archive	Archive	Archive	Archive	Archive	Archive		Archive
Prelemineries and Start condition								
Input	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP		E-ARK SIP
Output		check report	check report	check report	check report	Approved SIP		E-ARK AIP

5.3 Pilot 3 - Ingest from government agencies

Pilot 3	**Ingest from government agencies**
Task leader	National Archives of Estonia
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 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).
Object	EDRMS at a governmental agency (Alfresco), records preparation tool (UAM), digital preservation and access systems (Preservica, AIS)
Short description	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
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 (DoW)

Contacts	**Name (Title)**	**E-mail**	Skype
Contact Person	Karin Oolu	karin.oolu@ttu.ee	karinoolu
Pilot staff member	Tarvo Kärberg	tarvo.karberg@ra.ee	tarvo.karberg
Pilot staff member			

OAIS Relevance

E-ARK Tools

Pre-Ingest						E-ARK SIP	**Ingest - Storage**			E-ARK AIP	**Storage - Access**																E-ARK DIP
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**	**ICA-AtoM Catalogue**	**OMT - Search and Dsiplay GUI**	**Order Submission Service**	**OMT - Order Management Tool**	**Lily - Ingest**	**ESSArch Preservation Platform**	**E-ARK Web (Search)**	**AIP2DIP (E-ARK Web)**	**DBPTK**	**IP Viewer**	**DB Viewer (Sofia)**	**ERMS Viewer (Alfresco)**	**Single file Viewr**	**QGIS**	**Geoserver**	**Peripleo**	**Oracle (OLAP Viwer)**	**CMIS portal/viewer**
	X				X																	X					X

Pilot Scenarios	
Scenario 1	Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica (Data set 1)
Scenario 2	Provide access to records from governmental institution through RESTful services (Data set 2)
Scenario 3	Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica (Data set 3)
Scenario 4	Provide access to records from governmental institution through RESTful services (Data set 4)
Success Criteria	
General	
Scenario 1	Extract records from EDRM, create and ingest SIP of more then 650 records of "Decrees General" and "Board Meeting Minutes" documents. (95% success rate)
Scenario 2	Extract records from EDRM, create and ingest SIP of more then 3400 cases of the Estonian Unemployment Insurance Fund (Töötukassa). (95% success rate)
Scenario 3	Access information from more then 650 records "Decrees General" and "Board Meeting Minutes" documents. (95% success rate)
Scenario 4	Access information of more then 3400 cases of the Estonian Unemployment Insurance Fund (Töötukassa). (95% success rate)

Pilot 3	Scenarios
Scenario 1	**Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica**
Description	Export public records from an EDRM system of a governmental agency, create SIP, and ingest to the Preservica system at the National Archives of Estonia.

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																		
							E-ARK SIP	X			E-ARK AIP																	E-ARK DIP	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
		X				X																							

Use-case	**Extract and Ingest ERMS records based on MoReq2010**
Note	Alfresco is not Moreq-compliant system
Scenario 2	**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 standardised linking/access methods that are implemented in the agencies' digital infrastructure / web site.

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																		
							E-ARK SIP				E-ARK AIP																	E-ARK DIP	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
																							X						X

Use-case	**Access single ERMS records via alfresco cms**
Note	**To be consolidated with a CMIS interface access solution**
Scenario 3	**Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica**
Description	Scenario 3 will repeat the scenario 1 with the Data set 2
Scenario 4	**Provide access to records from governmental institution through RESTful services**
Description	Scenario 4 will repeat the scenario 2 with the Data-set 2

Pilot 3	Pilot Data		
Information Packages (IP)	**IP**		**Note**
	E-ARK SIP	X	
	non E-ARK SIP		
	E-ARK AIP		
	non E-ARK AIP	X	Preservica XIP
	E-ARK DIP		
	non E-ARK DIP	X	computer files sent directly to access

Pilot data description	
Data Set 1	**Records and metadata exported from Alfresco ERMS at Estonian Unemployment Insurance Fund (Töötukassa)**
Description	Data set consists of documents within two different series "Decrees General" and "Board Meeting Minutes" (retention period: permanent)
Data type	ddoc (word); bdoc (word) digitally signed
Metadata format	E-ARK Extended EAD
Quantity	Total 676 files from two series: 1. "Decrees General" total 325 files (2010 - 66; 2011 – 58; 2012 – 46; 2013 – 50; 2014 – 52; 2015 – 53) 2. "Board Meeting Minutes" total 351 files (2010 – 74; 2011 – 59; 2012 – 63; 2013 – 62; 2014 – 60; 2015 – 33.)
Data Set 2	**Records and metadata exported from Alfresco ERMS at Estonian Unemployment Insurance Fund (Töötukassa)**
Description	Data set consists of cases within the series "Practice contracts" from year 2015 including different formats of documents (retention period: 10 years)
Data type	pdf, doc, msg, bdoc, ddoc (digitally signed)
Metadata format	E-ARK Extended EAD
Quantity	Total 3404 cases incl different files as pdf, doc, msg, bdoc, ddoc

OAIS Process				Pre-Ingest					
Main Process Stepps	**Content definition Technical feasibility Legal issues etc.**	**Create/Review transfer agreement**	**Select data**	**Data Extraction**	**Manual compilation of non ERMS content**	**Metadata mapping**	**Create SIP**	**Post-packaging quality control**	**Submit SIP**
Scenario 1, 3	**Extract records from EDRM (of a governmental institution), create SIP and ingest to Preservica**								
Used E-ARK tool			Alfresco Export Module	Alfresco Export Module		Alfresco Export Module			
Used local tools							UAM		UAM
Perfomer (actor)			Producer	Producer		Producer	Producer		Producer
Prelemineries and Start condition									
Input				Alfresco Native Export Format		SMURF	E-ARK Extended EAD		E-ARK SIP
Output			Alfresco Native Export Format	SMURF		E-ARK Extended EAD	E-ARK SIP		SIP submitted

OAIS Process				Ingest					
Main Process Stepps	**Receive SIP**	**Security check**	**Integrity check**	**Sintax check**	**Format check**	**Review and Approve SIP**	**Modify SIP**	**Generate and Store AIP**	
Scenario 1, 3									
Used E-ARK tool									
Used local tools	KLEIO	KLEIO/Preservica	KLEIO/Preservica	KLEIO/Preservica	KLEIO/Preservica	KLEIO/Preservica	KLEIO/Preservica	KLEIO/Preservica	
Perfomer (actor)	archive	archive	archive	archive	archive	archive	archive	archive	
Prelemineries and Start condition									
Input	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	
Output	E-ARK SIP	check report	check report	check report	check report	check report	E-ARK SIP	Preservica XIP	

OAIS Process			Access				
Main Process Stepps	**Search & Select Records**	**Manage Order**	**Prepare DIP**	**Prepare DIP delivery**	**Deliver DIP**	**Access DIP**	**DIP Management**
Scenario 2, 4	**Provide access to records from governmental institution through RESTful services**						
Used E-ARK tool	CMIS portal/viewer		CMIS portal/viewer			CMIS portal/viewer	
Used local tools	Preservica		Preservica			Preservica	
Perfomer (actor)	user, archivist		user, archivist			user, archivist	
Prelemineries and Start condition	Portal is populated with metadata about records, user searches locally if finds relevant record, it is queried from Preservica						
Input			Selected records			DIP	
Output	Selected records		DIP				

5.4 Pilot 4 - Business Archives

Pilot 4	**Business Archives**		
Task leader	National Archives of Estonia		
Supported by	Estonian Business Archives		
Scope	Pre-ingest preparation and transfer of business records to a digital archive solution in a business archive		
Object	bespoke business system that contains database records		
Short description	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 database records from a private company to the digital archive solution of the Estonian Business Archives.		
Timeframe	M25-M27: setting up the pilot sites; M28-M31: running the pilots; M32-M33: testing and reporting.		
Preconditions	03.3, M03.4, M04.2, M05.4, M05.6 (DoW)		
Contacts	**Name (Title)**	**E-mail**	Skype
Contact Person	Raivo Ruusalepp	raivo@eba.ee	raivoruu
Pilot staff member	Ats Rand	ats.rand@eba.ee	atsrand
Pilot staff member			

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																		
							E-ARK SIP	X				E-ARK AIP																E-ARK DIP	
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**	**ICA-AtoM Catalogue**	**OMT - Search and Dsiplay GUI**	**Order Submission Service**	**OMT - Order Management Tool**	**Lily - Ingest**	**ESSArch Preservation Platform**	**E-ARK Web (Search)**	**AIP2DIP (E-ARK Web)**	**DBPTK**	**IP Viewer**	**DB Viewer (Sofia)**	**ERMS Viewer (Alfresco)**	**Single file Viewr**	**QGIS**	**Geoserver**	**Peripleo**	**Oracle (OLAP Viwer)**	**CMIS portal/viewer**
	X		X																									

Pilot Scenarios	
Scenario 1	Migration and Ingest of business data from bespoke business system
Success Criteria	
General	
Scenario 1	Migration and Ingest of more than 200 000 business records from bespoke business system (success rate 85% due complicated database architecture)

Pilot 4	Scenarios
Scenario 1	**Migration and Ingest of business records from bespoke business system**
Description	Export business records from bespoke business system. Ingest to local archival system of EBA.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																		
	E-ARK SIP						X	E-ARK AIP				E-ARK DIP																		
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	CMIS portal/viewer
	X		X																										
Use-case	**Extract and Ingest relational database based on SIARD 2.0**																												
Note																													

Pilot 4	Pilot Data

Information Packages (IP)	**IP**		**Note**
	E-ARK SIP	X	
	non E-ARK SIP	X	Format of local archival system of EBA
	E-ARK AIP		
	non E-ARK AIP	X	Format of local archival system of EBA
	E-ARK DIP		
	non E-ARK DIP		

Pilot data description	
Data Set 1	**(Under confirmation by the data Provider)**
Description	Business system with 12 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.
Data type	MS-SQL as mdf
Metadata format	none
Quantity	more than 200 000 rows

OAIS Process				Pre-Ingest					
Main Process Stepps	**Content definition Technical feasibility Legal issues etc.**	**Create/Review transfer agreement**	**Select data**	**Data Extraction**	**Manual compilation of non ERMS content**	**Metadata mapping**	**Create SIP**	**Post-packaging quality control**	**Submit SIP**
Scenario 1	**Migration and Ingest of MS-SQL database in SIARD 2.0 format**								
Used E-ARK tool				Database Preservation Toolkit			RODA-in		
Used local tools	Gathering agreements to use data for testing purpose.								Local tools
Perfomer (actor)	Manager		Specialists	Specialist			Specialist		Specialist
Prelemineries and Start condition			Populate test database with production data						
Input	Contracts		Set of production database	Database			SIARD 2.0		E-ARK SIP Local SIP
Output	Permission		Database for testing	SIARD 2.0			E-ARK SIP Local SIP		

OAIS Process				Ingest				
Main Process Stepps	**Receive SIP**	**Security check**	**Integrity check**	**Sintax check**	**Format check**	**Review and Approve SIP**	**Modify SIP**	**Generate and Store AIP**
Scenario 1								
Used E-ARK tool								
Used local tools	Local archival system of EBA	Local archival system of EBA	Local archival system of EBA	Local archival system of EBA	Local archival system of EBA	Local archival system of EBA		Local archival system of EBA
Perfomer (actor)	Specialists	Specialists	Specialists	Specialists	Specialists	Specialists		Specialists
Prelemineries and Start condition								
Input	Local SIP	Local SIP	Local SIP	Local SIP	Local SIP	Local SIP		Approved SIP
Output		check report	check report	check report	check report	Approved SIP		Local AIP

5.5 Pilot 5 - Preservation and access to records with geodata

Pilot 5	**Preservation and access to records with geodata**
Task leader	National Archives of Slovenia
Supported by	Danish National Archives
Scope	Pilot will prove that the SIP and DIP implementations fulfill specific requirements for the records containing GIS data, test the instructions (for the producer and for the archive) regarding all phases of ingest, to prove that the archival use of GIS data is possible (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
Short description	During the e-ARK project the standardized method for ingesting geo data will be developed. This will allow the archives to offer geodata as a selection and display criteria of records by means of integration of current state of the art tools.
Timeframe	M25-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 (DoW)

Contacts	**Name (Title)**	**E-mail**	Skype
Contact Person	Gregor Završnik ()	gregor.zavrsnik@gov.si	gregor.zavrsnik
Pilot staff member	Alenka Starman ()	alenka.starman@gov.si	
Pilot staff member	Joze Skofljanec ()	joze.skofljanec@gov.si	

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																
							E-ARK SIP	X			E-ARK AIP	X															E-ARK DIP	X
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
				X	X					X		X	X	X	X	X	X	X			X				X	X	X	

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

Success Criteria	
General	The following E-ARK tools will be tested in a pilot environment: ESSArch Tools Producer (ETP), ESSArch Tools Archive (ETA), ESSArch Preservation Platform (EPP), OMT - Search and Dsiplay GUI, OMT - Order Management Tool , IP Viewer, along with components of the Integrated Protoype (E-ARK Web): Order Submission Service, Lily-Ingest, Geoserver, Peripleo, with the integration of an Archival EAD based Catalogue and QGIS (Yes/No)
Scenario 1	SIP creation, verification and ingest of more then 1000 records with a vector geodata layer. (90% success rate)
Scenario 2	Finding , accessing, modifying and exporting a DIP containing a vector geodata layer of more then 1000 records. (90% success rate)

Pilot 5	Scenarios
Scenario 1	**SIP Creation and Ingest of records with Geodata**
Description	Create SIP from records and metadata exported from GURS (The Surveying and Mapping Authority of the Republic of Slovenia). SIP creation and ingest of at least one vector geodata with at least 1000 records Archivist creates a Submission agreement for SIP creation, according to E-ARK guidelines for geodata SIP creation. Producer creates a SIP containing geodata, acording to Submission agreement, based on EARK SIP specifications for geodata. Archivist technicaly validates the submited SIP package, according to E-ARK guidelines for geodata SIP creation. Archivist confirms, that content validation of the submited SIP package was performed. An AIP is generated from the SIP and gets ingested into the archival repository.

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																	
							E-ARK SIP	X			E-ARK AIP	X																E-ARK DIP	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
				X	X					X																		

Use-case	**Extract and Ingest computer files from simple file-system - GML**
Note	**SIP Creation and Ingest of records with Geodata**

Scenario 2	**Search and Access information using Geadota**
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. 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 enduser. Geodata from the order can be accessed in the designated viewer (QGIS). The user checks authenticity of the DIP by accessing PREMIS documentation. Access to DIP is documented and captured metadata can be exported.

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																	
							E-ARK SIP				E-ARK AIP	X																E-ARK DIP	X
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
												X	X	X	X	X	X	X			X				X	X	X	

Use-case	**Access geodata via QGIS**
Note	Access records with Geodata and present geodata with QGIS

Scenario 3	SIP Creation and Ingest of records with Geodata
Description	Create SIP from records and metadata exported from ARSO (Environmental Agency of Republic of Slovenia). SIP creation and ingest of at least one vector geodata with at least 1000 records. Data iz exported directly from their own system into GML format. And their sistem also exports INSPIRE metadata. Archivist creates a Submission agreement for SIP creation, according to E-ARK guidelines for geodata SIP creation. Producer creates a SIP containing geodata, acording to Submission agreement, based on EARK SIP specifications for geodata. Archivist technicaly validates the submited SIP package, according to E-ARK guidelines for geodata SIP creation. Archivist confirms, that content validation of the submited SIP package was performed. An AIP is generated from the SIP and gets ingested into the archival repository.

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																		
							E-ARK SIP	X			E-ARK AIP	X																E-ARK DIP		
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	
				X	X					X																			
Use-case	**Extract and Ingest computer files from simple file-system - GML**																													
Note	**SIP Creation and Ingest of records with Geodata**																													

Scenario 4	Search and Access information using Geadota
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. 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 enduser. Geodata from the order can be accessed in the designated viewer (QGIS). The user checks authenticity of the DIP by accessing PREMIS documentation. Access to DIP is documented and captured metadata can be exported.

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																		
							E-ARK SIP				E-ARK AIP	X																E-ARK DIP		X
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)	
												X	X	X	X	X	X	X			X			X	X	X			
Use-case	**Access geodata via QGIS**																													
Note	Access records with Geodata and present geodata with QGIS																													

Pilot 3	Pilot Data		
Information Packages (IP)	**IP**		**Note**
	E-ARK SIP	X	Focusing on Geodata preservation
	non E-ARK SIP		
	E-ARK AIP	X	Focusing on Geodata preservation
	non E-ARK AIP		
	E-ARK DIP	X	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 maps with Geodata
Data type	GML document with metadata in XML format, ESRI Shapefile, csv
Metadata format	ISO 19115 (INSPIRE)
less	62 records (cca. 3MB)

Data Set 2	Records and metadata of Natura 2000 areas, exported from ARSO
Description	Records and metadata of maps with Geodata
Data type	GML document with metadata in XML format
Metadata format	ISO 19115 (or INSPIRE)
less	1209 records (cca. 10 MB)

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

OAIS Process				Ingest				
Main Process Stepps	**Receive SIP**	**Security check**	**Integrity check**	**Sintax check**	**Format check**	**Review and Approve SIP**	**Modify SIP**	**Generate and Store AIP**
Scenario 1	**SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format**							
Used E-ARK tool	ESS Arch ETA	ESS Arch EPP	ESS Arch EPP	ESS Arch EPP + GML schema validator	ESS Arch EPP	ESS Arch EPP and QGIS	EAD Editor, Archival Catalogue	ESSArc EPP + integrated platform + Peripleo + Lilly
Used local tools						Local SFBS file viewers		
Perfomer (actor)	Archivist	Archivist	Archivist	Archivist	Archivist	Archivist	Archivist	Archivist
Prelemineries and Start condition							SIP is approved	
Input	SIP from Producer	SIP from Producer	SIP from Producer	SIP from Producer	SIP from Producer	SIP from Producer	Aproved SIP	Aprooved SIP + Archival EAD
Output	Validated SIP + Report	Validated SIP + Report	Validated SIP + Report	Validated SIP + Report	Validated SIP + Report	Aproved SIP	SIP + Catalogue metadata	AIP + Search Indexes in Peripleo and Lilly
Scenario 3	**SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format**							
Used E-ARK tool	ESS Arch ETA	ESS Arch EPP	ESS Arch EPP	ESS Arch EPP + GML schema validator	ESS Arch EPP	ESS Arch EPP and QGIS	EAD Editor, Archival Catalogue	ESSArc EPP + integrated platform + Peripleo + Lilly
Used local tools						Local SFBS file viewers		
Perfomer (actor)	Archivist	Archivist	Archivist	Archivist	Archivist	Archivist	Archivist	Archivist
Prelemineries and Start condition							SIP is approved	
Input	SIP from Producer	SIP from Producer	SIP from Producer	SIP from Producer	SIP from Producer	SIP from Producer	Aproved SIP	Aprooved SIP + Archival EAD
Output	Validated SIP + Report	Validated SIP + Report	Validated SIP + Report	Validated SIP + Report	Validated SIP + Report	Aproved SIP	SIP + Catalogue metadata	AIP + Search Indexes in Peripleo and Lilly

OAIS Process			Access				
Main Process Stepps	**Search & Select Records**	**Manage Order**	**Prepare DIP**	**Prepare DIP delivery**	**Deliver DIP**	**Access DIP**	**DIP Management**
Scenario 2	**Access Geodata from DIP0**						
Used E-ARK tool	Search and display GUI, Peripleo, LILY, Archival Catalogue	OMT	QGIS	QGIS + IP Viewer	OMT	QGIS + IP Viewer	OMT
Used local tools			Local SFBS file tools				
Perfomer (actor)	User	Archivist	Archivist	Archivist		User	Archivist
Prelemineries and Start condition	User is verifyed and registred	User is allowed acces to requested records	User credentials				
Input	EAD Catalog, search Index	Order.xml	Order.xml	DIP0 /DIPp, Order.xml	DIPu	DIPu	DIPu
Output	Order.xml	Order.xml approved	DIP0 /DIPp	DIPu	Use and access metadata	Use and access metadata	DIPp, deleted DIPu
Scenario 4	**Access Geodata from DIP0**						
Used E-ARK tool	Search and display GUI, Peripleo, LILY, Archival Catalogue	OMT	QGIS	QGIS + IP Viewer	OMT	QGIS + IP Viewer	OMT
Used local tools			Local SFBS file tools				
Perfomer (actor)	User	Archivist	Archivist	Archivist		User	Archivist
Prelemineries and Start condition	User is verifyed and registred	User is allowed acces to requested records	User credentials				
Input	EAD Catalog, search Index	Order.xml	Order.xml	DIP0 /DIPp, Order.xml	DIPu	DIPu	DIPu
Output	Order.xml	Order.xml approved	DIP0 /DIPp	DIPu	Use and access metadata	Use and access metadata	DIPp, deleted DIPu

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

Pilot 6	Integration between a live document management system and digital archiving and preservation service
Task leader	KEEP SOLUTIONS (KEEPS)
Supported by	Instituto Superior Técnico (IST)
Scope	The goal of this pilot is two-fold. On one hand, KEEP SOLUTIONS will demonstrate that the pan-European SIP structure designed in the WP3 is adequate to support the media types found in today's Electronic Records Management Systems (e.g. text documents, video, audio, images, etc) and, on the other hand, that the most adequate and scalable form of ingest is to automate the SIP creation and delivery process to the preservation service.
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 repository, namely, ingest and access by providing means to access Dissemination Information Packages from the producers Electronic Records Management Systems served by the preservation service.
Short description	This pilot focus on enhancing the pre-ingest and ingest processes integrating active Electronic Records Management Systems managed by producer's to a preservation service supported by the RODA open source repository. The pilot aims to demonstrate the suitability of the E-ARK SIP format for packaging the various content types typically found in ERMS systems as well as all the necessary contextual information and semantics to adequately consume the ingested data. The pilot will also demonstrate the scalability that can be achieved by deploying an integrated seamless ingest process that requires little or no human intervention. The pilot is composed of two scenarios based on two distict data producers using two diferent ERMS systems, namely, eDocLink and GfiDoc.
Timeframe	Scenario 1 - Between M26–M27 the pilot will be deployed. Between M28–M33 the pre-ingest and ingest process will be run. Scenario 2 - Between M26–M27 the pilot will be deployed. Between M28–M33 the pre-ingest and ingest process will be run.
Preconditions	The pan-European SIP format defined (WP3). RODA enhanced to support the new SIP format (WP3). Automatic SIP creation tool/middleware developed to integrate with the ERMS with the long-term repository. The ERMS capable of exporting records in MoReq/XML.

Contacts	Name (Title)	E-mail	Skype
Contact Person	Miguel Ferreira	mferreira@keep.pt	jmaferreira
Pilot staff member	Luís Faria	lfaria@keep.pt	luis100
Pilot staff member	Hélder Silva	hsilva@keep.pt	hsilva_keep
Pilot staff member	Sebastien Leroux	sleroux@keep.pt	slerouxatkeep
Pilot staff member	Rui Rodrigues	rrodrigues@keep.pt	rui.tiago.mr
Pilot staff member	Ricardo Vieira	rjcv@tecnico.ulisboa.pt	ricardojoao.vieira
Pilot staff member	João Cardoso	joao.m.f.cardoso@tecnico.ulisboa	joao.m.f.cardoso

OAIS Relevance	Pre-Ingest							ngest - Storage				Storage - Access																	
							E-ARK SIP	X		E-ARK AIP																	E-ARK DIP		
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
									X																			
Pilot Scenarios																												
Scenario 1	Automatic ingest of records from a semi-active archival management system																											
Scenario 2	Automatic ingest of records from a live Enterprise Records Management System																											
Success Criteria																												
General	Test the E-ARK compatible RODA Repository in a pilot environment. (Yes/No)																											
Scenario 1	Ingest of no less that 900 historical records in E-ARK SIP format automatically generated by a specially developed integration tool																											
Scenario 2	Ingest of no less than 500 accounting records in E-ARK SIP format automatically generated by a specially developed integration tool																											

Pilot 6	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 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 established practice by adding an additional component to its architecture that will be responsible for the long-term preservation of historical records once they reach their inactive age. The long-term preservation repository runs as a back-end service of the Archival Management System and aims to support its data curation activities.
Organisations involved	The organizations involved in this pilot are: Mafra Municipality (the data provider) and KEEP SOLUTIONS (the vendor behind the semi-active and the inactive archival solutions). The Mafra Municipality (Portuguese pronunciation: [ˈmafrɐ]) is a city and a municipality in the district of Lisbon, on the west coast of Portugal, and part of the urban agglomeration of the Greater Lisbon subregion. The population in 2011 was 76,685 in an area of 291.66 km². It is mostly known for the sumptuous Mafra National Palace built in the baroque style. Other points of interest around the municipality include the Tapada Nacional de Mafra, an enclosed wildlife and game reserve, and Ericeira's World Surf Reserve, the 2º in the world.
Software products involved	The software products involved in this scenario are: Archeevo and RODA. Archeevo (http://www.keep.pt/en/produtos/archeevo/) is an Archival Management Software capable of handling millions of archival records and terabytes of digital assets. This software consists of 9 functional modules that meet the needs of the most experienced archival professional, those being management of finding aids, management of digital assets, online publication, conservation and restoration, intermediate archive, management of deposits, virtual reference room, administration, productivity management, and interoperable programmable interfaces.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
	E-ARK SIP						X	E-ARK AIP				E-ARK DIP																	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
									X																			
Use-case	**Other (please specify in the Note field)**																											
Note	**Ingest of Archival Management Records using the SMURF profile.**																											

Scenario 2	**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 active records management system to a long-term preservation repository with little or no human intervention. The scenario is based on real-world operations already in place at a public organization since mid-2015. The workflow in place automatically transfers data from an ERMS (that supports day-to-day business operations and workflows) to an Archival Management System responsible for managing records in their semi-active age. The scenario enhances the established workflow by adding an additional component to this architecture that will be responsible for the long-term preservation of these records once they reach their inactive age. The long-term preservation repository runs as a back-end service of the Archival Management System and aims to support its data curation activities.
Organisations involved	The organizations involved in this pilot are: Mafra Municipality (the data provider), Link Consulting (the vendor of the active records management system) and KEEP SOLUTIONS (the vendor behind the semi-active and the inactive archival solutions). The Mafra Municipality (Portuguese pronunciation: [ˈmafrɐ]) is a city and a municipality in the district of Lisbon, on the west coast of Portugal, and part of the urban agglomeration of the Greater Lisbon subregion. The population in 2011 was 76,685 in an area of 291.66 km². It is mostly known for the sumptuous Mafra National Palace built in the baroque style. Other points of interest around the municipality include the Tapada Nacional de Mafra, an enclosed wildlife and game reserve, and Ericeira's World Surf Reserve, the 2º in the world.
Software products involved	The software products involved in this scenario are: edoclink, Archeevo and RODA. edoclink (http://www.linkconsulting.com/blog/whatwedo/edoclink-2/) is a document management solution that consists of a base solution and a set of additional modules, which provide coverage and flexibility in the implementation of Enterprise Records Management Systems. These modules include a BPM designer, connectors for ERPs, CRM, Archival Management Systems, reporting, digitization, barcode support, meetings, etc. Archeevo (http://www.keep.pt/en/produtos/archeevo/) is an Archival Management Software capable of handling millions of archival records and terabytes of digital assets. This software consists of 9 functional modules that meet the needs of the most experienced archival professional, those being management of finding aids, management of digital assets, online publication, conservation and restoration, intermediate archive, management of deposits, virtual reference room, administration, productivity management, and interoperable programmable interfaces.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
							E-ARK SIP	X		E-ARK AIP																	E-ARK DIP	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
									X																			
Use-case	**Extract and Ingest ERMS records based on MoReq2010**																											
Note																												

Pilot 6	Pilot Data		
Information Packages (IP)	**IP**		**Note**
	E-ARK SIP	X	E-ARK SIP is the format used to transfer data between Archeevo and RODA.
	non E-ARK SIP		
	E-ARK AIP	X	RODA keeps data in the E-ARK AIP format
	non E-ARK AIP		
	E-ARK DIP	X	You may download E-ARK DIP directly from RODA
	non E-ARK DIP	X	Will also support RODA search and presentation

Pilot data description	
Data Set 1	Historical records
Description	A collection of digitised books related to the Peninsular War dating from 1778 to 1834
Data type	300 dpi uncompressed TIFF files
Metadata format	EAD
Quantity	925 records described in EAD containing a total of 34.600 pages of 300 dpi uncompressed TIFF files. The total amount of data is around 1.10 TB.

Pilot Data - Scenario 2			
Information Packages (IP)	**IP**		**Note**
	E-ARK SIP	X	E-ARK SIP is the format used to transfer data between Archeevo and RODA.
	non E-ARK SIP	X	There are 3 systems involved. The interchange format between eDocLink and Archeevo is BagIt.
	E-ARK AIP	X	RODA keeps data in the E-ARK AIP format
	non E-ARK AIP		
	E-ARK DIP	X	You may download E-ARK DIP directly from RODA
	non E-ARK DIP	X	Will also support RODA search and presentation

Pilot data description	
Data Set 2	Accounting records
Description	A collection of accounting records related to payment orders undertaken by the municipality in 2016
Data type	Mostly office formats. 85% of the files will be smaller than 1 MB, however, the rest of the 15% will be between 1 and 10 MB
Metadata format	EAD
Quantity	Varies from year to year. In 2015 there were around 3000 records a year

OAIS Process				Pre-Ingest					
Main Process Stepps	**Content definition Technical feasibility Legal issues etc.**	**Create/Review transfer agreement**	**Select data**	**Data Extraction**	**Manual compilation of non ERMS content**	**Metadata mapping**	**Create SIP**	**Post-packaging quality control**	**Submit SIP**
Scenario 1	**Automatic ingest of records from a semi-active archival management system**								
Used E-ARK tool	-	-	-	-	-	-	-	-	
Used local tools	-	-	A2R	A2R	-	A2R	A2R	-	A2R
Perfomer (actor)	Data provider	Data provider/Repository manager	Tool	Tool	-	Tool	Tool	-	Tool
Prelemineries and Start cond	-	-	-	-	-	-	-	-	
Input	-	Transfer agreement	Archeevo API	Archeevo API	-	Archeevo API	Archeevo API	-	E-ARK SIP
Output	-	Signed transfer agreement	E-ARK SIP	E-ARK SIP	-	E-ARK SIP	E-ARK SIP	-	RODA API
Scenario 2	**Automatic ingest of records from a live Enterprise Records Management System**								
Used E-ARK tool	-	-	-	-	-	-	-	-	
Used local tools	-	-	A2R	A2R	-	A2R	A2R	-	A2R
Perfomer (actor)	Data provider	Data provider/Repository manager	Tool	Tool	-	Tool	Tool	-	Tool
Prelemineries and Start cond	-	-	-	-	-	-	-	-	
Input	-	Transfer agreement	Archeevo API	Archeevo API	-	Archeevo API	Archeevo API	-	E-ARK SIP
Output	-	Signed transfer agreement	E-ARK SIP	E-ARK SIP	-	E-ARK SIP	E-ARK SIP	-	RODA API

OAIS Process		Ingest						
Main Process Stepps	**Receive SIP**	**Security check**	**Integrity check**	**Sintax check**	**Format check**	**Review and Approve SIP**	**Modify SIP**	**Generate and Store AIP**
Scenario 1	**Automatic ingest of records from a semi-active archival management system**							
Used E-ARK tool	RODA Repository REST API	RODA Ingest Workflow	RODA Ingest Workflow	RODA Ingest Workflow	RODA Ingest Workflow	RODA Ingest Workflow	x	
Used local tools	A2R (to create and submit SIPs)						x	
Perfomer (actor)	Tool	Tool	Tool	Tool	Tool	Tool	x	Tool
Prelemineries and Start condition	SIP Available	SIP received					x	
Input	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	x	E-ARK SIP
Output		Ingest report	Ingest report	Ingest report	Ingest report	Ingest report	x	E-ARK AIP
Scenario 2	**Automatic ingest of records from a semi-active archival management system**							
Used E-ARK tool	RODA Repository REST API	RODA Ingest Workflow	RODA Ingest Workflow	RODA Ingest Workflow	RODA Ingest Workflow	RODA Ingest Workflow	x	
Used local tools	A2R (to create and submit SIPs)						x	
Perfomer (actor)	Tool	Tool	Tool	Tool	Tool	Tool	x	Tool
Prelemineries and Start condition	SIP Available	SIP received					x	
Input	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	x	E-ARK SIP
Output		Ingest report	Ingest report	Ingest report	Ingest report	Ingest report	x	E-ARK AIP

5.7 Pilot 7 - Access to Databases

Pilot 7	**Access to Databases**
Task leader	National Archives of Hungary
Supported by	Danish National Archives
Scope	Representation of not less than 2 databases of different sizes and complexities with restricted and open content.
Object	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
Short description	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
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 (DoW)

Contacts	**Name (Title)**	**E-mail**	Skype
Contact Person	Zoltan Lux	lux.zoltan@mnl.gov.hu	
Pilot staff member	József Mezei	mezei21@icloud.com	
Pilot staff member			

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																
							E-ARK SIP	X			E-ARK AIP	X															E-ARK DIP	X
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
	X		X					X			X					X		X	X	X		X					X

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

Success Criteria	
General	The following E-ARK tools will be tested in a pilot environment: DBPTK, RODA-in using Oracle OLAP Viewer, along with components of the Integrated Protoype (E-ARK Web): SIP2AIP, HDFS-Storage, Lily-Igest, Search, AIP2DIP, with the integration of Oracle (OLAP Viwer). (Yes/No)
Scenario 1	Create SIP and Ingest more then 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)
Scenario 2	Create SIP and Ingest more then 30.000 pages of scanned pdf images of meeting minutes of the former Hungarian Socialist Party. (95% success rate)
Scenario 3	Provide access for more then 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)
Scenario 4	Provide access for more then 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)
Scenario 5	Provide access for more then 30.000 pages of scanned pdf images of meeting minutes of the former Hungarian Socialist Party. (95% success rate)

Pilot 7	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.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
							E-ARK SIP	X			E-ARK AIP	X																E-ARK DIP	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
	X		X					X			X																	

Use-case	**Relational database based on SIARD 2.0**
Note	
Scenario 2	**SIP Creation and Ingest of unstructured files**
Description	Create SIP from scanned documents of the Meeting minutes of the Central Coimmettee of the Hungarian Socialist Party. The image files are in PDF format with EAD metadata. Create both E-ARK and local SIPs and ingest them into B27and Preservica archival repository. Both E-ARK and local AIPs are generated during the ingest.

OAIS Relevance	**Pre-Ingest**							**Ingest - Storage**				**Storage - Access**																	
							E-ARK SIP	X			E-ARK AIP	X																E-ARK DIP	
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	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
			X					X			X																	

Use-case	**Extract and Ingest computer files from simple file-system - Other (please specify)**
Note	Scanned document images in PDF files

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

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																
							E-ARK SIP				E-ARK AIP	X															E-ARK DIP	X
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
											X					X		X	X	X		X					

Use-case	Access databases via Sofia (sql)
Note	

Scenario 4	Search and present SIARD based information with E-ARK access tools
Description	Access database infromation of the Hungarian Prosecution Office in SIARD format using HADOOP based search and access with HIVE Presentation in local environmen.

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																
							E-ARK SIP				E-ARK AIP	X															E-ARK DIP	X
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
											X					X		X	X	X							X

Use-case	Access data with OLAP via oracle
Note	

Scenario 5	Access information from unstructured files
Description	Create DIP from scanned documents of the Meeting minutes of the Central Coimmettee of the Hungarian Socialist Party. The image files are in PDF format with EAD metadata in E-ARK Web HDFS storage and Preservica. Create both E-ARK and local DIPs.

OAIS Relevance	Pre-Ingest							Ingest - Storage				Storage - Access																
	E-ARK SIP							E-ARK AIP			X	E-ARK DIP																X
E-ARK Tools	Database Preservation Toolkit	Alfrsco Export Module	RODA-In	ESSArch Tool Producer (ETP)	ESSArch Tools Archive (ETA)	UAM	SIP creator (E-ARK Web)	SIP2AIP (E-ARK Web)	RODA Repository	ESSArch Preservation Platform	HDFS-Storage	ICA-AtoM Catalogue	OMT - Search and Dsiplay GUI	Order Submission Service	OMT - Order Management Tool	Lily - Ingest	ESSArch Preservation Platform	E-ARK Web (Search)	AIP2DIP (E-ARK Web)	DBPTK	IP Viewer	DB Viewer (Sofia)	ERMS Viewer (Alfresco)	Single file Viewr	QGIS	Geoserver	Peripleo	Oracle (OLAP Viwer)
											X					X		X	X		X			X				

Use-case	Access databases via SOLR (no-sql)
Note	Access data from E-ARK web / HDFS storage and from locals system. SOLR is used for search the full text index generated of the ducuments.

Pilot 3

Pilot Data

Information Packages (IP)	IP		Note
	E-ARK SIP	X	
	non E-ARK SIP	X	Preservica XIP
	E-ARK AIP	X	
	non E-ARK AIP	X	Preservica AIP
	E-ARK DIP	X	
	non E-ARK DIP	X	APEX, HIVE, Oracle Warehouse Builder

Pilot data description	
Data Set 1	**Hungarian Prosecution Office database**
Description	Old (not normalized) database in CSV exports of DBASE files.
Data type	CSV files
Metadata format	none
Quantity	more then 300.000 cases and 500.000 name. (1,6 GB)

Data Set 2	Scanned meeting minutes of the Central Coimmettee of the Hungarian Socialist Party
Description	Scanned documents in file systems in PDF file and corresponding metadata (EAD)
Data type	PDF files
Metadata format	EAD
Quantity	30-50.0000 files

OAIS Process				Pre-Ingest					
Main Process Stepps	**Content definition Technical feasibility Legal issues etc.**	**Create/Review transfer agreement**	**Select data**	**Data Extraction**	**Manual compilation of non ERMS content**	**Metadata mapping**	**Create SIP**	**Post-packaging quality control**	**Submit SIP**
Scenario 1	**SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format**								
Used E-ARK tool									
Used local tools	Define restrictions	Paper-based TA	Local ERMS		Manual	SIPKA Actor adds metadata by a GUI interface of a standalone software application	SIPKA		DVD, Personal delivery
Perfomer (actor)	Data provider and Archives	Data provider and Archives	Data provider	Data provider	Data provider	Data provider	Data provider		Data provider
Prelemineries and Start condition									
Input			Transfer Agreement	Selected Data	CSV files	CSV files in a predefined folder structure	Documentation (pdf, Excel files), primary data in CSM file, metadata in XML file		
Output		Transfer Agreement	Data selected	CSV files	CSV files in a predefined folder structure	XML file	OAIS compliant Information Package		
Scenario 2	**SIP Creation and Ingest of unstructured files**								
Used E-ARK tool						RODA-In	RODA-In		
Used local tools									Total Commander (local tool)
Perfomer (actor)	Archives	-	Archives	Archives	Archives	Archives	Archives		Archives
Prelemineries and Start condition									
Input			set of files			Metadata known by the archivists	set of files, metadata		
Output			subset of a set of files			XML within a SIP under construction	E-ARK SIP		

Ingest							
Receive SIP	**Security check**	**Integrity check**	**Sintax check**	**Format check**	**Review and Approve SIP**	**Modify SIP**	**Generate and Store AIP**
SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format							
							RODA-In, SIP2AIP
SIPKA	SIPKA	SIPKA	SIPKA	SIPKA	SIPKA	SIPKA	BÜRKE, SIARD Suite, Oracle, ELEV SIP Creator
Archives	Archives	Archives	Archives	Archives	Archives	Archives	Archives
Specific SIP	Specific SIP	Specific SIP	Specific SIP	Specific SIP	Specific SIP	Approved SIP	Modofied SIP
	check report	check report	check report	check report	Approved SIP	Modofied SIP	E-ARK SIP, SDB XIP
SIP Creation and Ingest of unstructured files							
RODA-In,earkweb SIP creator	RODA-In,earkweb SIP creator	RODA-In,earkweb SIP creator	RODA-In,earkweb SIP creator	RODA-In,earkweb SIP creator	RODA-In,earkweb SIP creator	RODA-In,earkweb SIP creator	RODA-In,earkweb SIP creator
Archives	Archives	Archives	Archives	Archives	Archives	Archives	Archives
E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	E-ARK SIP	Approved SIP	Modofied SIP
	check report	check report	check report	check report	Approved SIP	Modofied SIP	E-ARK AIP

OAIS Process			Access				
Main Process Stepps	**Search & Select Records**	**Manage Order**	**Prepare DIP**	**Prepare DIP delivery**	**Deliver DIP**	**Access DIP**	**DIP Management**
Scenario 3	**Extract SIARD Package from Preservica/E-ARK AIP**						
Used E-ARK tool	-	-	DPTK (SIARD2RDBMS)				
Used local tools	SDB/Preservica (local tool)	ScopeArchive (local)	Oracle RDBMS, APEX, Oracle BI		Oracle RDBMS, APEX	APEX, Oracle BI	Administration is ScopeArchive.
Perfomer (actor)	User, Archivist	Archivist	Archivist		Archivist	User, Archivist	Archivist
Prelemineries and Start condition							
Input		user request	SDB DIP (OAIS Compliant)	RDBMS data	Oracle schema	RDBMS schema	Oracle RDBMS schema for access
Output		new/modified record	RDBMS shema (a subset and anonymized version of RDMS data in a different shema)		Oracle schema	RDBMS schema available through APEX application or/and Oracle BI	DIP created from Oracle RDBMS by DBPTK and RODA-In and stored in HDFS-Storage. Modified record in ScopeArchive
Scenario 4	**Search and present SIARD based information with E-ARK access tools**						
Used E-ARK tool	E-ARK Web (Search), Lily Ingest, HDFS Storage		DPTK (SIARD2RDBMS)				
Used local tools		ScopeArchive	Oracle RDBMS, APEX, Oracle BI		Oracle RDBMS, APEX	APEX, Oracle BI	Administration is ScopeArchive.
Perfomer (actor)	User, Archivist	Archivist	Archivist		Archivist	User, Archivist	Archivist
Prelemineries and Start condition							
Input	Search terms	user request	DIP exported by E-ARK Web from HDFS storage	RDBMS data	Oracle schema	RDBMS schema	Oracle RDBMS schema for access
Output	Result list	new/modified record	RDBMS shema (a subset and anonymized version of RDMS data in a different shema)		Oracle schema	RDBMS schema available through APEX application or/and Oracle BI	DIP created from Oracle RDBMS by DBPTK and RODA-In and stored in HDFS-Storage. Modified record in ScopeArchive

Scenario 5	**Access information from unstructured files**						
Used E-ARK tool	E-ARK Web (Search), Lily, HDFS-Storage		AIP2DIP			IP Viewer	
Used local tools		ScopeArchive	image manipulation software (if needed)		Internet, LAN, File system		Administration is ScopeArchive.
Perfomer (actor)	User, Archivist	Archivist	Archivist		Archivist	User, Archivist	Archivist
Prelemineries and Start condition							
Input	Search terms	user request	DIP exported by E-ARK Web from HDFS storage		DIP	DIP	DIP
Output	Result list	new/modified record	DIP (with a subset of the exported DIP from the HDFS)		DIP	DIP presented by IP Viewer	DIP stored in HDFS-Storage. Modified record in ScopeArchive

5.8 Additional project pilots

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 directly not members of the E-ARK consortium.

List of the candidates for non-full-scale project pilots:

- Access scenario to one of the ingested data sets of the pilot (NAN)
- Additional access scenario (NAE)
- Ingest and access of a raster type Geodata (in GeoTIFF format + GML boundary) (NAS)
- Ingest and access of a set of raster geodata (multiple files, but with one reference GML file) (NAS)
- Ingest and access of a new time series to an existing dataset (the administrative units until 1994 will be available, and the changes until 2005 will be added) (NAS)
- Access scenario using RODA and some local tools for Searching Geodata in the Access phase. (NAS)
- Automatic ingest of records from Porto municipality (KEEPS)
- Data mining access scenario using one of the ingested data sets of the pilot (NAH)
- Pre-Ingest scenario using EssArch ETP and ETA tools in a Linux environment

Please note that this list can be changed according to agreements with data providers and possible new testing opportunities.

The additional project pilots will be defined in a separate document.

5.9 External validation activities

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.

The external validation activities will be defined in a separate document.

List of the candidates for external validation activities:

- Data Mining Showcase (a joint cooperation of the University of Brighton, University of Portsmouth, NAS and NAH)
- Evaluation of SIARD2 compatible tools and use of SIARD in E-ARK (Schweizerisches Bundesarchiv, Bern)

More evaluation activities are planned when the tools are stable and tested within the project.

6. Schedule

No #	Requirement	MoSCoW	Comment
6.1	Each pilot site should execute the pilot according to the following schedule	M	Small differences may occur.

Pilot Definition

Schedule

Pilot	Scenario	Deployment		Execution						Documentation		
		March, 2016	April, 2016	May, 2016	June, 2016	July, 2016	August, 2016	September, 2016	October, 2016	November, 2016	December, 2016	January, 2017
Pilot 1 (DNA)	1. Extracting records from database											
	2. Extracting records from database											
	3. Extracting records from database											
	4. Extracting records from database											
Pilot 2 (NAN)	1. SIP Creation and Ingest of unstructured records											
	2. SIP Creation and Ingest of unstructured records											
	3. SIP Creation and Ingest of structured records											
Pilot 3 (NAE)	3.1 Extract records from EDRM and ingest into Preservica											
	3.2 Extract records from EDRM and ingest into Preservica											
	3.3 Provide access to records from governmental institution through CMIS interface											
	3.4 Provide access to records from governmental institution through CMIS interface											
Pilot 4 (EBA)	4.1 SIP Creation and Ingest of business records from bespoke business system											
Pilot 5 (SNA)	5.1 SIP Creation and Ingest of records with Geodata											
	5.2 SIP Creation and Ingest of records with Geodata											
	5.3 Search and Access information using Geadota											
	5.4 Search and Access information using Geadota											
Pilot 6 (KEEPS)	6.1 Automatic ingest of records from a semi-active archival management system											
	6.2 Automatic ingest of records from a semi-active archival management system											
Pilot 7 (NAH)	7.1 SIP Creation and Ingest of old (not normalized) database in SIARD 2.0 format											
	7.2 SIP Creation and Ingest of unstructured files											
	7.3 Extract SIARD Package from Preservica / E-ARK AIP (APEX/OWB access)											
	7.4 Search and present SIARD based information with E-ARK access tools											
	7.5 Access information from unstructured files											

7. Success Criteria

The project has defined success criteria at different levels of the full-scale pilot:

- Pilot-level success criteria
  The whole full-scale pilot can be considered unsuccessful if these criteria are not met.
- Scenario-level success criteria
  The scenario cannot be considered successful unless these criteria are met.
- Archive-level success criteria
  Success criteria about the maturity model improvement of the archive executing the pilot.

Some success criteria have a yes/no success value (e.g. “RODA-In should be tested in this scenario”), the others, the measurable success indicators, have a success rate % (e.g. “Create E-ARK SIP from more than 2000 DB records. Success rate 95%”). The criteria will be evaluated with the help of the evaluation tables at the end of this chapter.

No #	Requirement	MoSCoW	Comment
7.1	The E-ARK full-scale pilot, as a whole, and the individual pilots and scenarios must be evaluated against the success criteria defined in this chapter	M	

Pilot-level success criteria

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	M	
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	M	
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	M	

Scenario-level success criteria

No #	Requirement	MoSCoW	Comment
7.5	An individual pilot scenario is successful if the following measurable criteria are met	M	

Pilot 1	SIP Creation on relational databases
Success Criteria	
General	The following E-ARK tools will be tested in a pilot environment: Database Preservation Toolkit (Yes/No)
Scenario 1	Extract records from Ms SQL database more then 100 GB. (95% success rate)
Scenario 2	Extract records fromlarge database containing documents. (To be finalized after data provider approval)
Scenario 3	Extract records from Ms SQL database containing 50-60 tables and about 90.000 records. (95% success rate)
Scenario 4	Extract records from MySQL database about 5 million records.(To be finalized after data provider approval)

Pilot 2	SIP creation and ingest of records
Success Criteria	
General	The following E-ARK tools will be tested in a pilot environment: ESSArch Tools Producer (ETP), ESSArch Tools Archive (ETA), ESSArch Preservation Platform (EPP). 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 (potentially) 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 dissminated during piloting. Another success criteria is if we succeed in specifying a conversion of EDRMS output produced in accordance with the national NOARK standard, into Moreq 2010. This is not directly linked to the ESSArch Tools which can handle content regardless of whether it is NOARK og Moreq 2010.
Scenario 1	SIP Creation and Ingest of unstructured records. (To be finalized after data provider approval)
Scenario 2	SIP Creation and Ingest of unstructured records. (To be finalized after data provider approval)
Scenario 3	SIP Creation and Ingest of unstructured records. (To be finalized after data provider approval)

Pilot 3	Ingest from government agencies
Success Criteria	
General	The following E-ARK tools will be tested in a pilot environment: Alfresco Export Module, along with components of the Integrated Protoype (E-ARK Web): ERMS viewer (Alfresco), CMIS Portal viewer. (Yes/No)
Scenario 1	Extract records from EDRM, create and ingest SIP of more then 650 records of "Decrees General" and "Board Meeting Minutes" documents. (95% success rate)
Scenario 2	Access information from more then 650 records "Decrees General" and "Board Meeting Minutes" documents. (95% success rate)
Scenario 3	Extract records from EDRM, create and ingest SIP of more then 3400 cases of the Estonian Unemployment Insurance Fund (Töötukassa). (95% success rate)
Scenario 4	Access information of more then 3400 cases of the Estonian Unemployment Insurance Fund (Töötukassa). (95% success rate)

Pilot 4	Business Archives
Success Criteria	
General	The following E-ARK tools will be tested in a pilot environment: Database Preservation Toolkit, RODA-In (Yes/No)
Scenario 1	Migration and Ingest of more than 200 000 business records from bespoke business system (success rate 85% due complicated database architecture)

Pilot 5	Preservation and access to records with geodata
Success Criteria	
General	The following E-ARK tools will be tested in a pilot environment: ESSArch Tools Producer (ETP), ESSArch Tools Archive (ETA), ESSArch Preservation Platform (EPP), OMT - Search and Dsiplay GUI, OMT - Order Management Tool , IP Viewer, along with components of the Integrated Protoype (E-ARK Web): Order Submission Service, Lily-Ingest, Geoserver, Peripleo, with the integration of an Archival EAD based Catalogue and QGIS (Yes/No)
Scenario 1	SIP creation, verification and ingest of more then 1000 records with a vector geodata layer. (90% success rate)
Scenario 2	Finding , accessing, modifying and exporting a DIP containing a vector geodata layer of more then 1000 records. (90% success rate)

Pilot 6	Integration between a live document management system and digital archiving and preservation service
Success Criteria	
General	Test the E-ARK compatible RODA Repository in a pilot environment. (Yes/No)
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)
Scenario 2	Ingest of no less than 500 accounting records in E-ARK SIP format automatically generated by a specially developed integration tool (90% success rate)

Pilot 7	Access to Databases
Success Criteria	
General	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 Protoype (E-ARK Web): SIP2AIP, HDFS-Storage, Lily-Igest, Search, AIP2DIP, with the integration of Oracle (OLAP Viwer). (Yes/No)
Scenario 1	Create SIP and Ingest more then 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)
Scenario 2	Create SIP and Ingest more then 30.000 pages of scanned pdf images of meeting minutes of the former Hungarian Socialist Party. (95% success rate)
Scenario 3	Provide access for more then 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)
Scenario 4	Provide access for more then 300.000 cases of old (not normalized) database of the Hungarian Prosecution Office. (90% success rate)
Scenario 5	Provide access for more then 30.000 pages of scanned pdf images of meeting minutes of the former Hungarian Socialist Party. (95% success rate)

Maturity model based success criteria

No #	Requirement	MoSCoW	Comment

7.6	An individual pilot is successful if, at the end of the pilot, the maturity level of the executing archive is equal or higher than the maturity level assessed before the start of the pilot	M	This criterion is not used for Pilot 6, since the executing organisation is not an archive.

7.1 Pilot evaluation tables

The following are samples of the pilot evaluation tables to record success criteria fulfillment.

Pilot-level success criteria

E-ARK Full scale pilots - Success Criteria				
General Success Criteria				
All E-ARK uses cases have been piloted				
	Use Case	**Pilot - Scenario**	**Successfull?**	**Comment**
Pre-Ingest	Extract and Ingest relational database based on SIARD 2.0	Pilot 1 - Scenario 1 -4 Pilot 7 - Scenario 1	Yes	
	Extract and Ingest ERMS records based on MoReq2010			
	Extract and Ingest computer files from simple file-system - GML			
	Extract and Ingest computer files from simple file-system - Other (please specify)			
Ingest	Ingest E-ARK SIP (Generate E-ARK AIP)			
Access	Access databases via Sofia (sql)			
	Access databases via SOLR (no-sql)			
	Access single ERMS records via alfresco cms			
	Access geodata via qgis			
	Access data with OLAP via oracle			
All E-ARK tools have been piloted				
	Tool		**Successfull?**	**Comment**
Pre-Ingest	Database Preservation Toolkit	Pilot 1 - Scenario 1 -4 Pilot 7 - Scenario 1	Yes	
	...			
	...			
	...			
Ingest	SIP to AIP coversion tool			
Access	Oreder Management			
	...			
	...			
	...			

Scenario-level Success Criteria

Pilot-level Success Criteria					
Pilot	**Scenario - Success Criteira**	**Should exceed (%)**	**Factual (%)**	**Successfull?**	**Comment**
Pilot1	Scenario 1 - Exporting not less then ... records of ...	95	98	Yes	Errors reported succesfully
	Scenario 2	90	100	Yes	
Pilot2	Scenario 1	95	99	Yes	
	Scenario 2	95	98	Yes	
	Scenario 3	90	76	No	Errors were not reported succesfully
	Scenario 4	95	96	Yes	
...					
...					

8. Pilot preparation check list

In order to follow pilot preparation activities and status, we define a preparation data sheet. In this table the preparation activities and statuses of the planned tools, data sets and other infrastructure elements can be recorded. The preparation data sheet provides an overview of the pilot preparatuion.

No #	Requirement	MoSCoW	Comment
8.1	The E-ARK full-scale pilot preparation should be recorded in the following Pilot Preparation status follow-up table.	M	

Pilot preparation status follow-up

Pilot #	Pilot Preparation

				Preparation status					
Software component Preparation tasks related to the software components	**Tool / Version number** from Software Component Matrix (for E-.ARK tools)	**Scenario** from Scenarios sheet	**Process** from Processes sheets	**Tool selected** Yes /No / (issue)	**Tool available for Pilot** Yes / (planned date of availability)	**Tool/Version installation** Installed / (issues)	**Tool configuration** No needed / Configured / (issues)	**Knowledge overtaken** Yes / (issues)	**Tool ready for Pilot** Ready / (issues)

			Prepration status			
Pilot dataset Preparation tasks related to pilot data	**Dataset #** from Pilot Data sheet	**Scenario** from Scenarios sheet	**Data selected** Yes / (issues)	**Legal issues** None / (issue)	**Data available** Yes / (planned date) / (issue)	**Dataset ready for Pilot** Ready / (issue)

			Prepration status		
Infrastructure Preparation tasks related to pilot infrastructure	**Scenario** from Scenarios sheet	**Process** from Processes sheets	**Element selected** Yes / (issues)	**Issues** None / (issue)	**Element ready for Pilot** Ready / (issue)

Sample Pilot Preparation list

Pilot 7	**Pilot Preparation**								
				Preparation status					
Software component	**Tool / Version number**	**Scenario**	**Process**	**Tool selected**	**Tool available for Pilot**	**Tool/Version installation**	**Tool configuration**	**Knowledge overtaken**	**Tool ready for Pilot**
Preparation tasks related to the software components	from Software Component Matrix (for E-.ARK tools)	from Scenarios sheet	from Processes sheets	Yes /No / (issue)	Yes / (planned date of availability)	Installed / (issues)	No needed / Configured / (issues)	Yes / (issues)	Ready / (issues)
SIARD Suite (local?)		Scenario 3	Access	yes	yes	Probable the install kit, published on 24.02.2016, contains a fixed versiom	A previouse version was already configured.	yes	Probably
DBPTK (RDBMS to SIARDX) (Database Preservation Toolkit)		Scenario 1, scenario 3, scenario 4	Pre-ingest	yes	Yes	A previouse version is already installed and tested	A previouse version was already configured.	yes	Ready
RODA-In (SIP creation tool)		Scenario 1, Scenario 2	Pre-ingest, Ingest					http://rodain.roda-community.org	
Integrated prototype (SIP creation tool)		Scenario 1, Scenario 2	Pre-ingest, Ingest??	yes	No date available, under development. The Integrated prototype will be installed by 15. March according our plan	No, it is planned by 15.03,2016	No configuration	Further knowladge transfer needed: training and user manual	Not ready, it must be installed
DBPTK (SIARDX to RDBMS)		Scenario 3, Scenario 4	Access	yes	yes				
SIP2AIP (Integrated Prototype)		Scenario 1, scenario 3, scenario 4	Ingest-Storage	yes	No date available, under development. The Integrated prototype will be installed by 15. March according our plan				
Lily - Ingest (Integrated Prototype)		Scenario 3, Scenario 4	Access	yes	No date available, under development. The Integrated prototype will be installed by 15. March according our plan				
HDFS Storage (Integrated Prototype)		Scenario 3, Scenario 4	Storage-Access	yes	No date available, under development. The Integrated prototype will be installed by 15. March according our plan				
E-ARK WEB (Search) (Integrated Prototype)		Scenario 3, Scenario 4	Access	yes	No date available, under development. The Integrated prototype will be installed by 15. March according our plan				
AIP2DIP (Integrated prototype)		Scenario 1, Scenario2, Scenario 5	Access	yes	No date available, under development. The Integrated prototype will be installed by 15. March according our plan	No, it is planned by 15.03,2016	No configuration	Further knowladge transfer needed: training and user manual	Not ready, it must be installed

SDB/Preservica (local)		Scenario 1, scenario 2, scenario 4, Scenario 5	Storage, access	yes	yes	installed	yes	yes	Ready
ELEV SIP Creator (local)		Scenario 1	Ingest	yes	yes	installed	yes	yes	Ready
SIPKA (local)		Scenario 1	Pre ingest	yes	yes	installed	yes	yes	Ready
Bürke (local)		Scenario 1	Ingest	yes	yes	installed	yes, but it must be reconfigured to the new Oracle database installed on the Hadoop cluster	yes	Ready
Oracle RDBMS (local)		Scenario 1, scenario 3, scenario 4	Pre ingest, ingest, access	yes	yes	installed	yes	yes	Ready
Oracle APEX (local)		Scenario 3	Access	yes	yes, but the appropriate further development is part of the pilot	yes, but...	yes, but further configuration have to be done based on the developments	yes	Ready
Oracle BI (local)		Scenario 3	Access	yes	yes, but the pilot needs a lot of developments within the software	yes, but...	yes, but further configuration have to be done based on the developments	yes	Ready
Oracle AWM (local)		Scenario 3	Ingest, Access	yes	yes, but the pilot needs a lot of developments within the software	yes, but...	yes, but further configuration have to be done based on the developments	yes	Ready
SIARD-OLAP		Scenario 3	Access	yes	No date available	No, it is being planned	No configuration	Further knowladge transfer needed	Not yet developed

			Prepration status			
Pilot dataset Preparation tasks related to pilot data	**Dataset #** from Pilot Data sheet	**Scenario** from Scenarios sheet	**Data selected** Yes / (issues)	**Legal issues** None / (issue)	**Data available** Yes / (planned date) / (issue)	**Dataset ready for Pilot** Ready / (issue)
BÜR (Registry of the Hungarian Prosecution Offices)	Data set 1	Scenario 1, Scenario 3, Scenario 4	yes	Non	yes	yes
Meeting minutes of the AGITPROP Departement of the Hungarian Communist Party	Data set 2	Scenario 2, Scenario 5	yes	Non	yes	yes

			Prepration status		
Infrastructure Preparation tasks related to pilot infrastructure	**Scenario** from Scenarios sheet	**Process** from Processes sheets	**Element selected** Yes / (issues)	**Issues** None / (issue)	**Element ready for Pilot** Ready / (issue)
Application developed in Oracle APEX to present normalized and denormalized database	Scenario 3, Scenario 4	Access	yes		Ready, but a lot of development is needed which is planned to be the part of the pilot
Denormalization of normalized relational databaase by Oracle BI	Scenario 1	Pre-ingest, Ingest???	yes		Ready, but a lot of development is needed which is planned to be the part of the pilot
Installation of Hadoop cluster	Scenario 4, Scenario 5	Storage, Access	yes		By 29.02.2016
Generating OLAP Cubes	Scenario 3	Pre-ingest, Ingest???, Access	yes		Ready, but a lot of development is needed which is planned to be the part of the pilot
Generatin an appropriate archival version of OLAP Cubes	Scenario 3	Pre-ingest, Ingest???, Storage, Access	Not selected yet		Not ready, it is still under design

9. Pilot feedback requirements

No #	Requirement	MoSCoW	Comment
9.1	Pilots are required to give regular feedback to tool developers about issues, requirements, recommendations or best-practices.	M	
9.2	Pilots should use the below Feedback forms to upload feedbacks.	S	
9.3	Pilots must inform the contact person of the corresponding tool developer as soon as possible in case of - critical errors, - issues preventing progress with pilot activities, - issues meaning significant risk for the successful execution of the scenario.	M	
9.4	Pilots must inform WP2 as soon as possible in case of - critical errors, - issues preventing progress with pilot activities, - issues meaning significant risk for the successful execution of the scenario.	M	
9.5	Pilots should inform the contact person of the corresponding tool developer in a monthly tool report about - non-critical errors, - performance issues, - inconveniencies when using the tool, etc.	S	

Feed-back reports

- Installation report
  Issues, experiences, recommendations and best-practices during the installation of the tools.
- Error report
  Software errors, software related issues during the running of the tools.
- Monthly tool report
  Tool usage, used data (type, volume), overall opinion about the tool (user friendly, easy-to-use, performance) in an easy-to-complete format.

10. Support requirements

This section defines the requirements set by the pilots towards the tool developers about providing support during the pilot.

No #	Requirement	MoSCoW	Comment
10.1	Tool developers are required to provide support about their products during - testing, - installation and - the execution of the pilots.	M	
10.2	Tool developers should name a direct contact person along with availability during the pilot set up and execution.	M	
10.3	Tool developers should react within 24 hours to support demands.	M	

11. Pilot documentation requirements

Pilot documentation will be part of the final pilot report. Our objective was to provide a detailed, step-by-step documentation about the pilot activities with minimal documentation effort by the pilot responsible and staff. Along this objective we have defined three types of reports:

- Pilot progress report
  Easy-to-use status report about the weekly/monthly progress of the pilot.
- Scenario execution report
  Detailed step by step report on: input / output, additional activities, experiences, issues, data usage, etc.
- Recommendations and best practices

12. References

This document references the following E-ARK documents:

- D2.1 – E-ARK General Model (deliverable of month 6)
- E-ARK General Model v2.0 (revised edition)
- Pilot Definition excel tables for each pilots