ANNUAL SUMMARY PROGRESS REPORT ON YEAR 2 OF THE E-ARK PROJECT

Each year, EU Research Projects must submit to the European Commission a detailed report of their activities during the preceding 12 months.

These reports contain a short summary section which is published to inform the public about the work undertaken and the project's achievements.

This report summarises the work of the E-ARK Project during its second year and covers the period from 1 February 2015 to 31 January 2016.

ANNUAL SUMMARY PROGRESS REPORT ON YEAR 2 OF THE E-ARK PROJECT

Overview and objectives

The E-ARK project goal is to pilot archival services to keep digital records authentic and usable based on current best practices which address the three main endeavours of an archive: acquiring, preserving and enabling re-use of information. The potential benefits for public agencies, citizens and business will be demonstrated by providing easy and efficient access to the archived records. Archival processes at a pan-European level will be harmonized, 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. The E-ARK approach will be public-facing, open source, robust, replicable and scalable, and will address a wide range of organisations, taking full account of legal constraints.

The following specific objectives for this year have been achieved. We have built upon the year one best practices studies and the draft specifications of two of the Information Packages (IPs) in order to develop pilot versions of the Submission IP (SIP) and the Archival IP (AIP), and further develop the draft specification of the Dissemination IP (DIP). This work has been set in the legal context provided by a comprehensive legal study of the European landscape as it pertains to archives. The General Model, with its overarching workflows, methodology and use cases has been updated in order to provide a current framework for the project. The first versions of the newly integrated / developed open source tools / platforms necessary for supporting the archival processes are ready, as is the Integrated Platform Reference Implementation. We have carried on our “Big Data” work: the faceted search facility has been completed, and work continued on the data mining showcase. The Knowledge Centre has been developed in prototype, and the Maturity Model evaluated and assessed. Our dissemination strategy for the year has focussed on reporting on progress and recruiting early users. We have produced detailed and thorough preparatory material (pilot cards) for running our seven pilots in year three. All deliverables have been submitted on time, and all milestones met.

Co-Ordination (Work Package 1 - WP1) University of Brighton, UK

Project Management

E-ARK has been successfully managed throughout the year in accordance with the principles of the Prince 2009, MSP (Managing Successful Programmes) and M_o_R (Management of Risk) methodologies. The Project Management Team have maintained regular contact with all work packages through fortnightly teleconferences and occasional face-to-race meetings when the opportunity arose through other events having brought project partners together. Once each month, these meetings are extended to include representatives of all Consortium Members, thereby enabling any issue affecting any part of the project to be identified and addressed without delay. Six monthly Project Board meetings have been held as scheduled. A variety of media have been used for intra-project management during the year: Cisco Webex, Sharepoint, Redmine, Github and Google Drive. In January 2016, (the last month of the reporting period) project co-ordination moved to the University of Brighton, the transfer being formally ratified by the EC; the former co-ordinator, UPHEC; and the project partners. The project co-ordination processes / facilities outlined above continued unchanged over the transfer period, thus facilitating continuity of experience for the project partners.

Technical Coordination, National Archives Estonia

The daily technical management and coordination of E-ARK continued as established during the first year of the project. Virtual technical meetings have been organised on a monthly basis for all technical staff. In addition, three cross work package face-to-face meetings were held:

- February (Portsmouth, UPHEC);
- May (Vienna, AIT);

- December (Lisbon, IST).

Several targeted cross work package groups have also been set up. Most notably the work on Information Package specifications has been synchronised, resulting in the decision to create an overarching “Common Specification for Information Packages” document which provides the backbone for all more detailed specifications (SIP, AIP, DIP and the E-ARK content types’ specifications). Also notable is the collaboration in regard to E-ARK workflow specifications where specific details agreed upon in pre-ingest (WP3), ingest and preservation (WP4) and access (WP5) are always synchronised with the E-ARK General Model (maintained by WP2).

During the reporting period the focus of the project turned from delivering specifications towards delivering software tools. Some delays have occurred in tool delivery, mainly due to the late availability of detailed process and information package specifications. However, most E-ARK tools were available or close to completion by the end of M24 which allows the project to start the pilot instances on time.

A summary overview of the Year Two progress of the technical work packages now follows.

Work Package 2 (WP2) - Use Cases and Pilots, National Archives Hungary

WP2 submitted the D2.2 deliverable as a fundamental document of the legal environment of digital archiving. Major progress was achieved in tool developers support and pilot coordination. These two tasks have a lot of connecting activities, for which an appropriate methodological framework has been laid down. As a result, WP2 has been prepared for the management of the full scale/additional pilots and other validation activities of the final and final year.

Work Package 3 (WP3) - Transfer of Records to Archives, National Archives Estonia

For WP3 the main effort was to work further with the SIP specification by delivering profiles for relational databases and Electronic Records Management Systems (ERMSs). The specification work concluded with an updated version of the general SIP specification, the SIARD 2.0 standard (in close co-operation with the Swiss Federal Archives) and SMURF (Semantically Marked Up Records Format) specification for ERMSs and Simple File System Based (SFSB) records as the deliverable D3.3 E-ARK SIP pilot specification in M24. The current pilot SIP specification will be updated based on the experiences and feedback received from pilots in M36. WP3 was also active in the pre-ingest and ingest workflow definition process by supporting the work on workflow diagrams led by WP2 Use Cases. The development of pre-ingest and ingest tools is in progress and is guided by use cases with requirements created by E-ARK partners.

Work Package 4 (WP4) – Archival Records Preservation, Austrian Institute of Technology

For WP4, the second year started with a review of the use cases and requirements related to the AIP format. Based on the initial AIP specification draft presented in of E-ARK deliverable D4.2 “E-ARK AIP draft specification” [1] the work in Task 4.1 focused on preparing the next version of the AIP format specification (deliverable D4.3 “E-ARK AIP pilot specification”). In a cross-WP collaboration, the Common Specification was developed. From the WP4 perspective, this document laid the ground to develop the AIP format in close coordination with the other E-ARK IP formats developed in WP3 (E-ARK SIP) and WP5 (E-ARK DIP). In addition, WP4 started implementing the SIP to AIP conversion tool based on a Python/Django/Celery technology stack. Regarding the archiving of databases, E-ARK was looking into methods to add de-normalized and OLAP representations to the AIP which allows dimensional analysis via OLAP. Additionally, work in WP4 was related to standardize the way that content from Electronic Records Management Systems (ERMSs) can be archived. The E-ARK approach is based on the MoReq specification[2] and gives guidance on how to reuse it for SIPs and AIPs.

Work Package 5 (WP5) – Archival Records Access Services, Danish National Archives

We have delivered in M15 a draft specification of the Dissemination Information Package (deliverable D5.2 “E-ARK DIP draft specification”). It presents the first version of the E-ARK DIP format. It first illustrates and describes the workflows and use cases of archival access services and ultimately uses these

[1] http://www.eark-project.com/resources/project-deliverables/18-d42-e-ark-aip-draft-specification

[2] http://moreq.info

to present the set of requirements to be followed when designing the final DIP format. As access to archival records is largely dependent on the tools and environments used, the secondary aim of the deliverable was to also go beyond the DIP format and look closely at the tools needed for preparing and using the DIP. Therefore, this deliverable is also being used in E-ARK as the basis for the tool development and to assure that the tools and DIP requirements are well aligned. The identification of tool requirements was also completed during this year, and the subsequent development of Access tools begun. At the end of the year, developers were well advanced in certain areas and could present prototypes of the search functionality; of the AIP-DIP conversion tool; of the IP viewer; and of the geodata presentation tools.

Work Package 6 (WP6) – Archival Storage, Services and Integration, Austrian Institute Technology

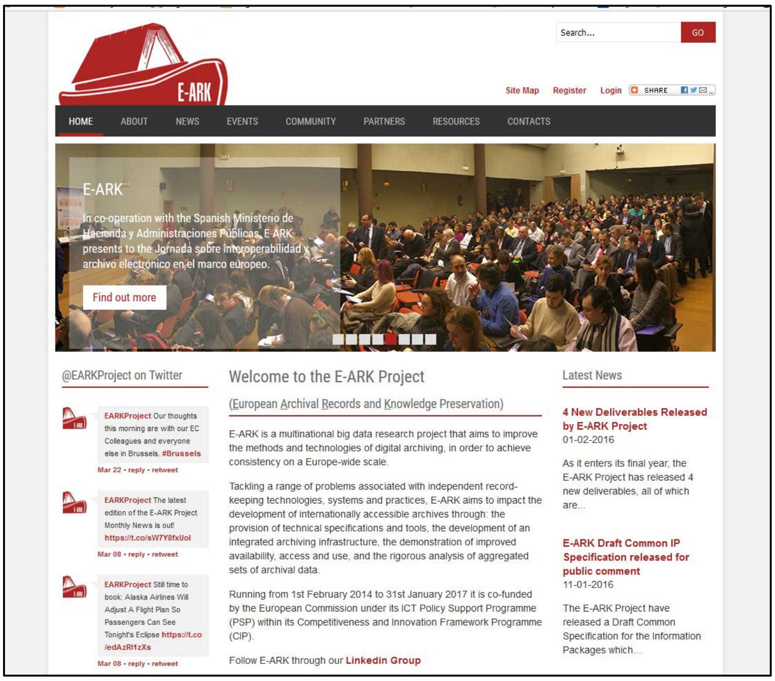
WP6 has produced two deliverables describing an architecture and prototype implementation for the storage, search, and access of E-ARK IPs in a scalable environment. The resulting E-ARK Integrated Platform Reference Implementation Prototype integrates (a) an IP creation and management system, (b) a repository for content search and access, and (c) a scalable storage and execution environment based on Apache Hadoop. The developed components have been harmonized and integrated with the ESSArch platform with respect to core functionality, interfaces (task definition, HDFS upload), used programming language (Python and Python libraries), and underlying execution systems (Celery, Linux). The resulting prototype enables users to control the creation of Information Packages and to upload the packages to the Hadoop-based infrastructure. The HDFS upload service has been extended to support the calculation of file checksums using MD5, SHA-1, and SHA-256 algorithms. A message broker has been employed to automate the ingestion of uploaded IPs into the Lily content repository. A search interface has been developed that supports searching - also complex objects like office and PDF documents - based on full-text and selected metadata fields within and across IPs. The displayed results are directly linked with the Lily content repository allowing users to access data items contained in IPs at the file level.

Work Package 7 – Evaluation and Assessment Instituto Tecnico Lisbon, Portugal

WP7 assessed the project pilots using a questionnaire based on the criteria defined in the first version of the Information Maturity Model (submitted under Deliverable 7.1). The questionnaire intended to assess the capabilities that will be developed / improved in the project prior to the deployment of the E-ARK solutions. The assessment process and results were described in Deliverable 7.2. The assessment process and the Maturity Model are currently being revised in order to create the second and final version of the information Maturity Model (to be submitted in M36). A functional prototype of the Knowledge Centre was developed and disseminated to the public (http://kc.dlmforum.eu/home). The prototype includes functional releases of the (1) EVOC service, (2) REQs service, (3) MoReq Schemas Validator, and (4) Maturity Assessment. The services goals and functionalities are described in Deliverable 7.3 along with the plans for future development of the Knowledge Centre**.**

Work Package 8 (WP8) – Project Dissemination, University of Brighton, UK

In our second year, our project has moved from its awareness-raising phase to that of reporting on progress and recruitment of early users. We have continued to disseminate information about our work and our results via a wide variety of channels: online, a dedicated mailing list, publication of papers and by presentations at scientific conferences. We have organised our first event dedicated to the work of the Project and achieved our quality targets. At the end of the project’s first year we extended our dissemination performance targets to ensure that they were challenging. We have both achieved and exceeded these, and will therefore extend these further again in our final Year.



E-ARK

In co-operation with the Spanish Ministerio de Hacienda y Administraciones Públicas, E-ARK presents to the Jornada sobre interoperabilidad y archivo electrónico en el marco europeo.

@EARKProject on Twitter

EARKProject Our thoughts this morning are with our EC Colleagues and everyone else in Brussels. #Brussels
Mar 22 • reply • retweet

EARKProject The latest edition of the E-ARK Project Monthly News is out! https://t.co/sW7Y8fxUol
Mar 08 • reply • retweet

EARKProject Still time to book: Alaska Airlines Will Adjust A Flight Plan So Passengers Can See Tonight's Eclipse https://t.co/edAzRl1zXs
Mar 08 • reply • retweet

Welcome to the E-ARK Project

(European Archival Records and Knowledge Preservation)

E-ARK is a multinational big data research project that aims to improve the methods and technologies of digital archiving, in order to achieve consistency on a Europe-wide scale.

Tackling a range of problems associated with independent record-keeping technologies, systems and practices, E-ARK aims to impact the development of internationally accessible archives through: the provision of technical specifications and tools, the development of an integrated archiving infrastructure, the demonstration of improved availability, access and use, and the rigorous analysis of aggregated sets of archival data.

Running from 1st February 2014 to 31st January 2017 it is co-funded by the European Commission under its ICT Policy Support Programme (PSP) within its Competitiveness and Innovation Framework Programme (CIP).

Follow E-ARK through our Linkedin Group

Latest News

4 New Deliverables Released by E-ARK Project
01-02-2016

As it enters its final year, the E-ARK Project has released 4 new deliverables, all of which are...

E-ARK Draft Common IP Specification released for public comment
11-01-2016

The E-ARK Project have released a Draft Common Specification for the Information Packages which...

At the start of the project E-ARK established three stakeholder Advisory Boards as an integral component of project governance, as well as to enhance project communication and dissemination activities.

Advisory Boards

The Advisory Boards' main contribution is to assess contributions to and from the project and to adjudicate on conflicting community views if these arise. In general terms, the Advisory Boards also:

- represent a range of stakeholder interests broader than those represented by project partners
- ensure E-ARK project outputs are compatible with relevant national and international standards and legislation
- keep the project work connected to best practice in digital preservation approaches, tools and services
- help disseminate information about and outputs of the E-ARK project to their stakeholder communities.

The Boards are open to all interested parties and represent a wide range of stakeholder interests. The current membership numbers for the three E-ARK Advisory Boards are:

• the Commercial / Technical Advisory Board	10 organisations (+1)
• the Archival Advisory Board	27 members representing 18 (+4) organisations
• the Data Provider Advisory Board	4 institutions, 1 individual (no change)

All completed project deliverables are circulated to the Boards following their submission to the EC. In year 2, Advisory Board members were sent the following deliverables:

- D5.2 (June 2015),
- D2.2, D6.1 (August 2015)
- D7.2 (October 2015)

- D3.3, D4.3, D6.2, D 7.3 (February 2016)

In addition, the internal deliverable, *Introduction to the Common Specification for Information Packages*, was circulated to Advisory Board members in January 2016. It is pleasing to be able to report that the amount of feedback received has increased during the second year of the project. All feedback received is acknowledged, and individual responses prepared by Work Package leads and the Technical Coordinator are sent to respondents.

At least one face to face meeting of the Advisory Boards is held every year. The purpose of the meetings is to inform Board members of progress of various work packages, to discuss issues and concerns, and to receive additional feedback on project deliverables. In year two of the project, two face to face meetings were held. The first of these was held in conjunction with the International Council of Archives (ICA) Annual Conference in Reykjavik, Iceland on 27 September. The second face to face meeting of the Boards was held in conjunction with the DLM Forum Member Meeting in Luxembourg on 15 October. Although 9 members attended the very successful DLM Advisory Boards meeting, attendance at the ICA Advisory Boards meeting was low, with only 3 members attending. It may not be viable to hold future AB meetings outside mainland Europe. Advisory Board members are sent an overview report approximately every quarter that details Work Package progress in the previous three months and foreshadows major project activities in the upcoming three months. In year two, reports were circulated in February, May, September and December 2015.

Expected Results / Impact

E-ARK will provide national, regional and local archives with an open source digital archiving framework, complete with accompanying metadata and other appropriate standards. The resulting harmonisation will help communication across archives, but still respect different national legislative backgrounds. Organisations supplying data to archives will be provided with streamlined pre-ingest standards / tools which reduce their costs. A wide range of users (including business, researchers and citizens) of archives will benefit from improved access to digital archives, also covering databases. There will be new “Big Data” methods, such as faceted searches, to enhance discovery. We are on track to produce these results / impact.

Contact info@earkproject.eu E-ARK Grant Agreement no. 620998