

E-ARK PROJECT – BEST PRACTICE SURVEY RESULTS ON ARCHIVING OF DIGITAL MATERIAL

Clive S G Billenness
University of Portsmouth
School of Creative Technologies
Portsmouth
+44 208 123 2782
clive.billenness@port.ac.uk

Kathrine H E Johansen
The Danish National Archives
Rigsdagsgaarden 9
DK-1218 Copenhagen K
+45 41 71 72 20
khej@sa.dk

David Anderson
University of Portsmouth
School of Creative Technologies
Portsmouth
+44 23 9284 5525
cdpa@btinternet.com

ABSTRACT

This paper describes the poster presented at iPres2014 by the EC-funded E-ARK Project, detailing the results of an international user survey recently conducted into Best Practice in the full life cycle of the long term preservation of digital records by archival organizations in the state and private sectors.

General Terms

Communities, Preservation Strategies, Preservation Workflows, Best Practice.

Keywords

Digital Archives, User Survey, E-ARK, EC, ICT-PSP, SIP, AIP, Pilot, e-infrastructure, digital archives, data mining, OAIS, Big Data, born-digital records, ingest, access

1. INTRODUCTION TO E-ARK

European Archival Records and Knowledge preservation (E-ARK) was launched in February 2014 and is a new, 3-year pilot project within the European Commission's ICT Policy Support Programme. With 16 partners in 11 EC countries comprising end users, research institutions and systems suppliers, its objective is to provide a single, scalable, robust approach capable of meeting the needs of diverse organisations, public and private, large and small, and able to support complex data types. E-ARK will demonstrate the potential benefits for public administrations, public agencies, public services, citizens and business by providing simple, efficient access to the workflows for the three main activities of an archive - acquiring, preserving and enabling re-use of information.

E-ARK will implement a number of pilot systems in different countries addressing challenges which differ in content and scale in order to create, by the end of the project, in 2017, a suite of openly-accessible end-to-end solutions capable of integration into third-party products and which will be sustained into the future.

Our work is worldwide the first attempt to bring together working elements of archival systems. As such it is an ambitious project which has several key features: creating standardized pre-ingest

iPres 2014 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

formats / specifications; expanding MoReq modules to be used as a key element of the infrastructure; using CMIS and Big Data techniques to promote new ways of access to digital archives, etc. It also addresses a wide range of users: public bodies, commercial institutions, individual citizens and researchers.

Our project will also provide a Digital Preservation Maturity Model which will enable organizations to not only assess their current performance, but also to measure improvement.

More information about the project is available from our website at www.eark-project.eu.

2. THE BEST PRACTICE SURVEY

2.1 Methodology

During the summer of 2014, the project conducted a survey into available best practices in Data Ingest, Available Requirements and Formats for AIP's and into the gaps between existing Access Solutions and users' needs.

We conducted an online survey to which responses were received from 184 respondents from 32 countries. Respondents covered a wide range of domains:

- Archives
- Private Companies
- Public Organisations
- Libraries
- Universities and
- Private Organisations

The questions within the survey were differentiated to be appropriate to each group.

We then followed up with person-to-person interviews with representatives from seven archives and four digital archiving solutions vendors.

The full results of our work are available for download from our website.

2.2 Available best practices in data ingest

We identified that two approaches must be differentiated: the ingest of whole systems and the ingest of individual records.

No widespread practices for records export can be identified. The most commonly used standards are ISO15489-1 and MoReq, but there seem to be lack of consensus on what is good practice for records export

Ingest workflows are generally considered to be in accordance with the OAIS model but the implementation varies greatly.

PIMAS steps can be used to describe/cover the most common steps in the ingest workflow.

The understanding of what constitutes a SIP varies greatly between organisations. Some consider simple computer folders as SIPs, others consider metadata standards as SIP and other again have defined a specific structure for SIPs.

Similarities between SIP formats can be found the use of metadata standards where METS, PREMIS and EAD emerge as significant.

Most SIP formats include the following four components:

- xml-file for describing the structure of the SIP,
- an xml-file for descriptive metadata,
- a unique identifier (UID), and
- a folder with content

2.3 Available requirements and formats for AIP's

Based on a set of generic criteria for an ideal conceptual AIP, the following existing AIP formats were identified as the best existing AIP concepts:

- US Patent 13/219,630 Method And System For Preparing Digital Information For Long-Term Preservation
- BSI TR-03125 Preservation of Evidence of Cryptographically Signed Documents, Annex TRESOR-F: Formats and Protocols
- The AIP from Archivemata
- The AIP from ESSArch
- The German developed DA-NRW AIP format
- The AIP format from Preservica

For AIP containers, the three most significant formats are ACE (.ace), RAR (.rar) and Tape Archive (.tar), each of which was evaluated for its suitability for further use within the E-ARK Project. At this stage, Tape Archive is recommended as the most suitable format to be used in the ongoing E-ARK work.

2.4 Gap analysis between existing access solutions and users' needs

Our study revealed that users are concerned with the following requirements for archival access services:

- Contemporary solutions that meet the standards of modern IT services
- Services that are easy to use i.e. that do not require specific technological or human skills
- Speed of access and usability and flexibility of services
- Integration or interoperability between different parts of an access service e.g. between finding aids and presentation tools
- Possibilities to search across Information Packages in data and metadata
- Functionalities that support their specific purpose

We concluded that:

- There is still limited experience with providing access to born-digital material
- Generally users' needs are not met very well by existing services

The most prominent gaps are between existing solutions and users' needs are:

- lack of flexible and modern solutions,
- lack of interoperability between components,
- lack of comprehensive metadata in finding aids,
- lack of functionalities to support use of data in presentation tools

3. NEXT STEPS IN E-ARK

Our project will now proceed to develop requirements and recommendations for the export of source records.

We will build an ingest workflow based on PIMAS.

We will include the four most common components of SIP's in the E-ARK SIP format.

We will continue to develop our products with a persistent core but with sufficient flexibility to accommodate the specific needs of individual organizations.

4. TO FIND OUT MORE ABOUT E-ARK

A PDF version of our poster, together with copies of all our reports, can be downloaded from our project website:

www.eark-project.eu

At the website, it is also possible to sign up for a mailing list to be kept informed about developments in the project.

You can also follow our Twitter feed at @EARKProject.