The application of E-ARK tools for archival interoperability to support a long-term sustainable Digital Single Market

Karin Bredenberg
National Archives of Sweden
P.O. Box 12541, SE-102 29 Stockholm
Sweden
Karin.Bredenberg@riksarkivet.se

Kuldar Aas
National Archives of Estonia
Nooruse 3, Tartu 50411
Estonia
Kuldar.Aas@ra.ee

Prof David Anderson
University of Brighton
Watts Building, Lewes Road,
Brighton, East Sussex, BN2 4GJ
United Kingdom
c.d.p.anderson@brighton.ac.uk

Dr Jaime Kaminski
University of Brighton
Watts Building, Lewes Road,
Brighton, East Sussex, BN2 4GJ
United Kingdom
J.Kaminski@brighton.ac.uk

ABSTRACT

Since the early 2000s, the European Union has promoted and supported the development of cross-border and digital means for conducting business and running governments. This strategic focus carried out through the initial Lisbon Strategy[1] and the follow-up Europe 2020 strategy[2] has contributed to a widespread shift toward e-governments and e-commerce. Ultimately, this has transformed the relationships between governments, citizens and business. From a technical perspective, a wide range of new information systems and communication methods have become available that streamline and automate transactions, enable integration of information and service delivery, and enhance collaboration between participants. It has also changed the ways of conducting public and private business, documenting activities, and making information available for reuse, aiding decision-making and supporting accountability. Much of this information has to remain available for long periods, often beyond the life expectancy of individual information systems or technological components. This poses a significant challenge for both the original data creators and digital archives that have to be capable of both dealing with the most current information technology, and with "saving" information from a variety of legacy systems. In order to deal with these challenges, there is a need to bridge the gap between the methods used in digital preservation and content creation and introduce additional practical interoperability into the solutions used for a Digital Single Market and digital preservation infrastructures. The European Commission recognised the need in 2013 and funded the E-ARK project[3] (2014 – 2017) which successfully delivered a set of principles, specifications and tools that provided an interoperable and international solution for the sustainable long-term availability of digital information. This paper describes the solutions first developed during the E-ARK project, focusing on the SIP format, as well as the steps the European Commission is taking towards building a thriving digital preservation community around the E-ARK outcomes.

CONFERENCE THEMES ADDRESSED

- Collaboration and capacity-building
- Technological infrastructure

KEYWORDS

Digital archiving, access, digital preservation, information management, information lifecycle, standards, specifications, Digital Single Market, Europe 2020

[1] http://www.europarl.europa.eu/summits/lis1_en.htm
[2] https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/european-semester/framework/europe-2020-strategy_en
[3] http://www.eark-project.eu

1 INTRODUCTION

Over the last few decades, governments and business have become increasingly digitized. For example, in the late 1990s, the Estonian government was largely paper-based. By the early 2000s, the digitization of processes began, resulting in the implementation of electronic records management systems and the creation of electronic records in typical office file formats. Following this rather simple digitization, the focus turned towards providing web-based services that relied on some interconnected databases documenting created and exchanged information in a highly normalized and structured way. Now the government is taking the next step and has started carrying out its activities pro-actively, offering zero-bureaucracy services, not at the request of citizens or businesses, but automatically once needed[4]. While nations and businesses might differ in their specific aspects, some global trends are evident. For example, methods used in information governance are often retired within a decade of their implementation, and replaced with new approaches that are both technologically and conceptually different.

The issue for any long-term repository dealing with e-government or e-commerce is that its managers need to be aware of these different approaches, systems and formats, and it must develop and implement appropriate specifications, tools, and workflows capable of effectively processing the information in all these various formats. As e-governments and e-business continue to thrive, we can expect a growing number of archival institutions to come under significant pressure because of lack of resources and knowledge.

One obvious solution to reduce the pressure is to reuse best practices and tools developed by other institutions. However, there is no single, sufficiently interoperable, widely understood and accepted approach for how to transfer valuable digital information to digital archives, or to preserve and access that data over the long-term. Most crucially, practical interoperability is lacking around OAIS Information Packages. While some commonly accepted specifications exist within the digital preservation community (METS[5], BagIt[6], Premis[7], etc), these are not sufficiently detailed or standardized to allow for the development of interoperable package creation, validation and manipulation software.

Work defining information packages began in Europe in the early 2000s. A period which saw the development of the OAIS reference model; the Producer-Archive Interface Methodology Abstract Standard (PAIMAS) ISO 20652:2006 describing the administrative part of a transmission; and especially the Producer-Archive Interface Specification (PAIS) ISO 20104:2015 based on the XML Formatted Data Unit (XFDU) Structure and Construction Rules ISO 13527:2010 (the PAIS schemas can be found at https://sanaregistry.org/r/daixml). PAIS was developed during the E-ARK project.

There were, however, standards that were available that were further ahead in their development. Moreover, these standards already had an established user base that had already created profiles for transfers and was willing to share their experiences. Using pre-existing standards was much quicker compared to creating profiles from scratch for a completely new schema. The E-ARK project, therefore, built upon existing tools that used standards like METS and PREMIS because moving to other formats would have required a complete rewrite of the tools which was a cost that could not be justified in this project.

The critical limitation with these standards is that they require users to create a profile for their specific needs. While this is a useful feature, it complicates interoperability. Consequently, commonly developed profiles for use by the whole community needed to be developed. This was the task that the E-ARK project took on and continues to this day.

The creation of the "Common Specification for Information Packages" and the profiles for the different information packages (SIP, AIP and DIP) which are used in the tools provided allows for specialization of the use in all different use cases.

The "content information type specifications" create a common means of describing different content types. The advantage for users is that they can use the same specifications for the same content when transferring them either between systems or to an archive, while also allowing for the use of another specification if needed. In the collaboration with content information type specification owners, as well as the creation of new content type information specifications it will be essential to increase the number of content information type specifications.

At the same time, available e-government interoperability standards do not consider digital preservation natively and are mostly to be seen handling "individual data items", not systems, and a complete export of information from the system.

2 E-ARK BACKGROUND

The European Commission (EC) acknowledged the need for more collaboration and interoperability in the area of long-term preservation in 2013 and funded the E-ARK project (2014 – 2017) to address the problem in an open and international collaboration between academia, national archival services and commercial systems providers.

An early phase of E-ARK involved a global study of the current state of digital archiving procedures, in order to establish the state-of-art in long-term digital preservation, as well as to find gaps in current knowledge and implementations. Initial desk research identified a number of reports that clarified and compared how

[4] https://www.mkm.ee/en/zero-bureaucracy
[5] http://www.loc.gov/standards/mets/
[6] https://tools.ietf.org/html/draft-kunze-bagit-14
[7] https://www.loc.gov/standards/premis/

institutions have approached the issues of digital archiving and digital preservation. A number of expert interviews and a survey to augment the desk research followed this up. Full details of the study are published on the E-ARK website [1][2][3].

The general conclusion the project drew from the reports consulted, discussions and surveys is that a number of archival approaches exist in institutional or national settings across Europe. However, available approaches are rather fragmented and practical reuse of these in other technical or legal contexts is extremely difficult and requires extensive remodeling and customization. From a more technical and pragmatic perspective, the study highlighted that solutions meant for archival purposes were usually custom developments only suitable for use in specific institutional or national contexts.

Based on these results it was concluded that there was critical need to develop an overarching and international methodology that addressed business and operational issues in digital archiving; and practical interoperability specifications and tools for ingest, preservation and re-use. As such, E-ARK worked towards ascertaining the practical needs of the community and establishing a practical approach by refining and further specifying already available (but insufficient) individual standards.

Ultimately, the E-ARK project decided to provide a single approach that was capable of meeting the needs of diverse organizations, whether public or private, large or small, and was able to support various complex data types. E-ARK aimed to demonstrate the potential benefits for public administrations, public agencies, public services, citizens and business by providing simple, efficient access to standards, tools and workflows for the main activities of an archive, including export from source business systems, transfer of digital content to long-term repositories, preservation and enabling access and re-use. Acknowledging the huge mass of information created by public and private sector information systems, scalability and robustness of the standards, tools and workflows were some of the core design constraints.

The E-ARK project also went beyond theory and implemented the resulting standards, tools and workflows in seven open pilots in various national contexts in 2016 – 2017. The pilots assessed the interoperability and ease-of-use aspects of the project outcomes by combining previously existing local solutions with services developed by other project partners. As far as possible, the pilots were executed in production systems using real data.

The E-ARK project concluded on 31 January 2017 with excellent marks from the EC and project reviewers. Immediately after the end of the project, the DLM Forum[8] took over the maintenance, dissemination, implementation and further development of all outcomes. As the first specific initiative, the standardization body "Digital Information LifeCycle Interoperability Standards Board" (DILCIS Board) was set up on 1 February 2017, taking over the ownership for the core Information Package specifications. However, in 2018 – 2019 the DLM Forum intends to establish a full-service offering on long-term digital archiving, including training, consultation, certification, and research aspects.

3 SPECIFICATIONS FOR ARCHIVAL INTEROPERABILITY

The definition of archival interoperability in the context of this paper is twofold. First, it is about the interoperability between source information systems (where data creation, gathering and management occurs throughout its active use) and long-term repositories. The reasoning for interoperability in this scenario is rather simple and partly already given in the introduction. Given the availability of a widely accepted and implementation-oriented standard, it is rather straightforward to:

- develop native archiving functionality into any source system, for the transfer of data to any long-term repository of choice (and vice versa for an archival reuse scenario), and
- reuse software components developed initially for other national or institutional contexts, therefore significantly decreasing the resources needed to implement a full-scale digital repository for any single archival institution.

The secondary aim of archival interoperability is to ensure that data already stored in long-term repositories can be exported in a simple, scalable and platform independent way. This ensures that an archival institution itself does not end up in a "legacy software hostage" situation and that repositories can exchange their software tools for better when they are needed and available.

The aspects of interoperability relevant in these scenarios are mainly semantic and technical. The most crucial requirements for an archival scenario are as follows:

- Data and metadata are in standardized formats, so their subsequent use is not inhibited by system differences;
- The data and metadata, and any other information required to use the data are combined in a single conceptual package;
- The package contains enough information to allow validation and identification of all its components in a standardized way at any point in time;
- The package is constructed in such a way that its information content can be understood in the long term without reference to external systems or standards by any systems or software available at any point in time.

The E-ARK project based its approach on the widely recognized OAIS Reference Model [4]. Consequently, the project followed the

[8] http://www.dlmforum.eu

definitions of the main archival processes and associated conceptual information package definitions articulated in OAIS:

- Ingest – the Submission Information Package (SIP);
- Archiving – the Archival Information Package (AIP);
- Dissemination – the Dissemination Information Package (DIP).

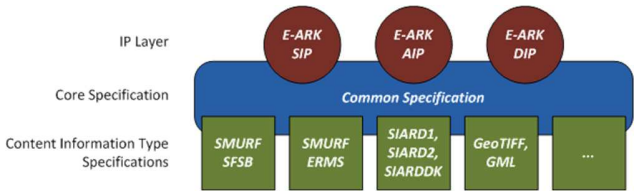

Figure 1: E-ARK specification ecosystem

The work on an interoperable environment of Information Packages resulted in the creation of a layered model[9] (Figure 1), which includes:

- The Common Specification for Information Packages: This specification outlines the essential requirements for any packages, and provides an overall basis for developing tools capable of working with any information packages, regardless of their type, size or content;
- The SIP, AIP and DIP Specifications: All of the specifications follow the requirements set in the E-ARK Common Specification but extend it with specifics of the relevant processes (pre-ingest and ingest, preservation, access). In this paper, we concentrate on the SIP format that sets out requirements for the packaging of information for transfer between producer systems and archives. The importance of the SIP specification is that the interoperability encouraged by it allows archives to replace repository systems as needed while remaining compatible with established ingest workflows. Secondly, a well-established SIP specification ensures that vendors will be able to adapt their electronic records management systems to be compatible with the specification, allowing the creation of integrated workflows between producers and archives.
- Content Information Type Specifications: To guarantee that the integrity and authenticity of transferred information are not compromised, we need to go beyond the actual data and consider system-specific aspects. For example, a typical real-world records management system contains records arranged into aggregations, metadata relating to records and their relationships to other entities, a business classification scheme, a set of retention and disposal schedules, user access controls and definitions, information to support the retrieval of a record by a search engine and so on. All these components, which make up a specific and complete information package, must be transferred together with the data in a way that ensures the integrity, authenticity and understandability of the whole package. The E-ARK model addresses this need with the concept of Content Information Type Specifications, which allow for the definition of relevant system-specific elements which need to be archived along with the data, and which ultimately are used to extend the scope of the common specification itself.

3.1 The Common Specification for IPs

The Common Specification for Information Packages [5] provides the backbone of archival interoperability. The OAIS compliant specification builds on the requirements presented above and provides a unified set of rules for packaging any data and metadata into a single conceptual package, seamlessly transferred between systems, preserved and reused in the long term. The core of the common specification is a definition of an Information Package structure (Figure 2).

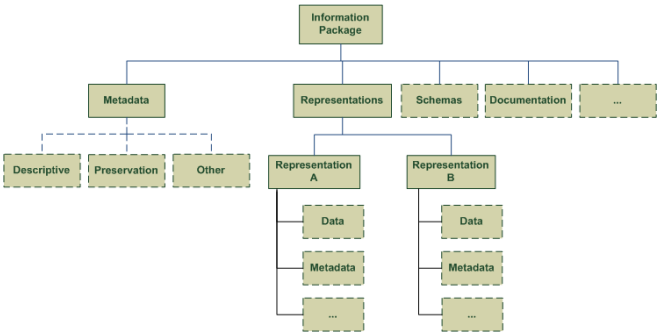

Figure 2: E-ARK Information Package structure

The structure allows for the separated inclusion of any metadata, data, relevant schemas and documentation. Further, the metadata in the package can be divided into descriptive (metadata needed to find and understand the data), preservation (metadata needed to ensure the integrity and authenticity of data, metadata and the whole package) and other (any other metadata which is deemed relevant to be included into the package by the source system or the archives).

A specific feature of the data component is the specific division into explicit representations. As such, the representations view allows for the incremental growth of the package throughout preservation actions, while keeping each representation separated, ultimately supporting the authenticity of the package and the information object. The representation component can also include metadata or other components specific to this exact version of the information

[9] All specifications are available at http://www.dilcis.eu/

object. This is especially useful in the case of large objects where it is not possible and reasonable to keep all representations within a single information package. The common Specification structure defines a representation to be effectively a "package within a package" it is indeed possible to split representations into separate physical packages, which are also understandable and usable on their own.

Lastly, to ensure that the whole package can be understood and reused in the long-term users can make the package self-sustainable by including any relevant schemas and documentation which might not be available externally in the future. Documentation may take many forms and constitutes what the OAIS calls Representation Information, i.e. information that cannot be easily classified as semantic or structural (e.g. software, algorithms, encryption, emulation environments, written instructions and many other things may be needed to understand the object). Again, it is possible to add such additional information either to the package as a whole or next to individual representations.

In addition to the structure, the Common Specification details the use of core structural and packaging metadata. Each package includes a core XML metadata file, which follows the widely recognized METS standard. The core METS metadata serves the main purpose of:

- identifying the package and its components in a persistent and unique way;
- providing a standardized overview of all components of the package;
- connecting relevant pieces of data, metadata and other components to each other.

As such, E-ARK uses METS metadata as the manifest or table of content for the whole package and ensures that the validation of the components of the information package follows commonly accepted rules. In order to achieve these goals, the E-ARK implementation of METS poses multiple additional limitations to the original specification. Most crucially, E-ARK clearly specifies the:

- use of identifiers in any information package;
- linking mechanism to be used for referencing any data or metadata included in the package;
- explicit use of the METS file section and structural map;
- definition of the type of the information package and content information type(s) included in the package.

It is worth noting that the E-ARK pilots demonstrated that archival software components from various developers could be easily joined into a single workflow, therefore validating the practical benefits of the structure and metadata requirements described above.

3.2 Submission Information Packages (SIPs)

The first stage of the digital archiving workflow is extracting information from the producer's business system and packaging it for transfer to the archive's system. The OAIS Reference Model conceptualizes information submitted to an Archive as one or more discrete transmissions of Submission Information Packages (SIPs). The E-ARK SIP specification [6] provides a detailed description of the structure and main metadata elements that should be part of an E-ARK SIP and functions as the initial input for the technical implementations of pre-ingest and ingest tools that automate the creation and transformation of SIPs.

In its simplest form, an E-ARK SIP is a packaged set of files and folders inside a ZIP or TAR container (Figure 3). A SIP can contain one or more representations of a single intellectual entity (e.g. Rep-001 and Rep-002 under the "representations" folder in the diagram). The SIP can hold metadata that is related to the intellectual entity as a whole; at the same time, each representation may also contain its own specific metadata, although separation of metadata in this way is purely optional.

Moreover, as provided for in the common specification, the information package folder must include a mandatory core metadata file named "METS.xml", which includes the information needed to identify and describe the structure of the package itself and the rest of its constituent components. One vital requirement for the E-ARK SIP specification is that it can be extended to support any content type a digital repository needs to ingest. Accordingly, the specification allows for the development of additional separate content type descriptions for different types of information being submitted to the archives.

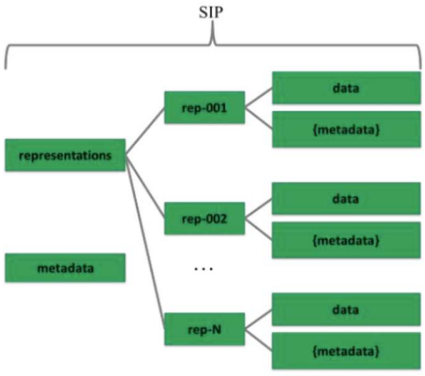

Figure 3: E-ARK SIP structure

The main differences between the SIP specification and the common specification are:

- the SIP specification proposes an optional model for recording a transfer agreement in a common machine-readable way;
- the SIP specification requires the description of additional actors relevant to the SIP transfer process within METS metadata.

3.3 Content Information Type Specifications

An information package does not just contain "data", but information structured according to a specific content type and described using specific metadata. The types of data files and their structural relationships, and metadata elements vary for different content types[10]. Metadata produced by a specific business system will variously be intended to support descriptive, structural, administrative, technical, preservation, provenance (relating to authenticity) and rights (relating to IP, retention and access) functions. For example, the ways to describe the context and content of a relational database can vary significantly from the in-depth description of a piece of interactive digital art or GIS data. Many archival organizations have to deal with a significant number of these content types, resulting in the need for individual specifications, principles and software available for each. As such, the success of the archive depends largely on its capability to add new components quickly and efficiently, which support specific content types along to the core elements of the repository supporting the content-agnostic manipulation and storage of information packages.

The METS standard used in the common specification does not offer one single structure in which content type specific data and metadata could be stored as a whole. In order to efficiently use metadata to support archival functions, the common specification defines separate METS sections as containers for the various metadata functions, such as the METS header for package management, the <dmdSec> for Encoded Archival Descriptions (EAD) [11] and other descriptive metadata standards, and the <amdSec> for preservation (PREMIS), technical and other functions.

The role of the Content Information Type Specifications is to:

- define a content type-specific internal structure and format for the data portion of the information package;
- describe how submitted metadata, documentation and other components have to be positioned and referenced from relevant sections in METS metadata (<amdSec>, <dmdSec>).

According to the E-ARK ideology, the ecosystem of Content Information Type Specifications is open, and anyone interested in content level interoperability is welcome to add new specifications as needed. However, the E-ARK project itself developed four content information type SIP profiles:

- *SMURF SFSB Profile* (Semantically Marked-Up Record Format for Simple File Systems Based records) [7]. This specification provides for the common and interoperable description and packaging of data and metadata originating from simple file-systems (i.e. computer files being stored on a "shared drive") and the requirements for describing these using EAD;
- *GIS Profile* [7]. A sub-component of the SMURF SFSB profile is the specification for the inclusion of GIS-derived information and supporting documentation in GeoTIFF and GML formats;
- *SMURF ERMS Profile* (Semantically Marked-Up Record Format for Simple File Systems Based records) [7]. The specification specifies further how to archive the necessary elements of an ERMS system, including the classification scheme, aggregations and classes, disposal schedules, and user access controls along with the records and their metadata;
- *Relational Database Profile SIARD 2.0* (Software Independent Archiving of Relational Databases) [8]. SIARD is an open format originally developed by the Swiss Federal Archives in 2004. The format is designed for archiving relational databases in a vendor-neutral form. The format proposes a common standard for describing core elements of the live DBMS: data; structure; stored procedures; triggers; views and queries. The new version of SIARD (i.e. SIARD 2.0) has been developed in collaboration between the Swiss Federal Archives and E-ARK[12].

4 TOOLS AND PILOTS

Developing specifications is at the core of any interoperability standardization activity. However, there are far too many examples of standardization, which, appear solid on paper, but fail to reach widespread implementation.

In order to provide pragmatic and practical specifications the E-ARK project committed to delivering real-life implementations regarding software tools. The project collaborated therefore with three software providers (ES Solutions, Sweden [13] ; KEEP Solutions, Portugal[14]; Magenta Aps, Denmark[15]) who updated their already available information management or digital preservation tools according to the specifications developed by the E-ARK project.

[10] The term "content information types" has been coined within the E-ARK project. In other sources, other terms might be used for a similar concept. In general, it is meant that information or content comes in different formats, like databases, 3D images, PDF records, etc. All of these have different characteristics, which are relevant to be described and kept with the data in order to establish appropriate future understanding and reuse capacity.

[11] https://www.loc.gov/ead/

[12] Early 2017 some errors and ambiguities were discovered within the SIARD 2.0 specification. In collaboration between E-ARK partners and the Swiss Federal Archives an update was prepared (SIARD 2.1). As of April 2018 SIARD 2.1 is waiting for approval and publishing by the Swiss national standardisation agency.

[13] http://www.essolutions.se/

[14] https://www.keep.pt/en

[15] http://magenta.dk/

In total, the project produced or updated more than 20 individual open source software components [16], which all follow the specifications. Most notably, the work resulted in overlaps in the E-ARK SIP specification support, meaning that there are multiple tools from different providers supporting the creation and transfer of an E-ARK SIP package. The software components cover all specifications and archival workflows, meaning that it is possible to build a holistic preservation workflow solely using these components starting from records management or relational database content extraction, through transfer packaging, ingest and preservation at the long-term digital repository, and provide access to archived content.

While developers tested all these tools individually within the project, the ultimate aim was to establish practical interoperability between information systems and software components developed by various companies. In order to validate such aspects the project conducted seven pilots in national contexts at the National Archives of Denmark, Norway, Estonia, Slovenia, Portugal and Hungary, and the Estonian Business Archives. All of these pilots aimed to integrate some of the E-ARK tools into already established digital preservation workflows and establish that the use of international specifications did indeed simplify the exchange of data and metadata between source systems and repository, and provide for easier reuse of preservation tools developed in other contexts. The project managed to conclude all seven pilots successfully and gain practical feedback for the further development of the specifications. Detailed reports on all the pilots are available at the project website [9][10][11].

5 FROM PROJECT TO PROGRAMME

While the E-ARK project delivered a solid set of foundational work during its lifetime, it was understood from the outset that principles, specifications and tools evolve in line with the development of practices in IT systems development, and information governance both at public agencies and private institutions. The specifications produced by E-ARK reflect the practical experiences of the archival and information management experts from nine European countries, and may reasonably be considered to represent best practice within that community. The intention was always to make the outputs of E-ARK as widely applicable as possible, and this requires input from other countries and other professional domains. It was therefore apparent from the beginning of the E-ARK project that, in order to achieve practical implementation-level interoperability, the project would need to go beyond technical work and:

- make e-government, e-business and digital archiving communities aware of the specifications and the overall aim of archival interoperability;
- provide relevant training and support opportunities for interested parties;
- carry out pilots in additional national and professional contexts (e.g. eHealth, construction, aviation);
- actively manage communication and feedback in order to develop the principles, specifications and tools further.

In short, the aim for the future development of specifications is "development through practice", where practical implementation experiences are well balanced with theoretical discussions.

The E-ARK consortium, recognizing these issues, started drafting a long-term sustainability plan for the E-ARK outcomes before the project's end in January 2017. As the first step, the ownership and management responsibility of the project's outcomes was transferred to the DLM Forum, an EC-supported international organization that brings together national institutions and individuals interested in the lifecycle issues of digital information from Europe and beyond. More specifically, the DLM Forum took immediate action and established an expert board on 1 February 2017 (the Digital Information LifeCycle Interoperability Standards Board, or the DILCIS Board[17]) tasked with taking care of the most crucial outcomes of the project – the interoperability specification.

In parallel, a core team of former E-ARK partners and DLM Forum representatives negotiated with the European Commission to use the project outcomes as the basis of an eArchiving Building Block within the Connecting Europe Facility (CEF) program[18]. The CEF program is one of the core policy programs for implementing the Digital Single Market within the European Union, with the explicit aim of reducing the impact of national and institutional borders in the use of technology. The work within CEF is di4vided into generic Building Blocks (e.g. eSignature, eIdentification, eDelivery) and sector-specific Digital Service Infrastructures (e.g. eHealth, eProcurement). As such, it was logical to add also digital preservation, or eArchiving, into the list of fundamental Building Blocks to facilitate a harmonized approach to preserving crucial digital information created while carrying out business or public activities. The original proposal was submitted to the CEF program in October 2016 and was unanimously accepted by EU Member States in March 2017. Ultimately, in December 2017 the European Commission officially announced the inclusion of the eArchiving Building Block into the CEF program, quickly followed by the publication of the first grants for specific activities.

As such, in 2018 the activities around archival interoperability continue with the aim of setting up a series of support services by summer 2019. These are a:

- support infrastructure for the further maintenance of the E-ARK specifications (i.e. continuation of the DILCIS Board), open to all interested bodies;
- clearly defined Sample Software Portfolio that consists of open source software that uses the latest versions of

[16] All tools are freely available from the E-ARK website: http://www.eark-project.com/resources/eark-tools

[17] http://www.dilcis.eu

[18] https://ec.europa.eu/inea/en/connecting-europe-facility

the specifications which will serve as a reference implementation;
- training offering that covers all the aspects of the specifications;
- service desk for helping interested bodies to understand the specifications and aiding in their implementation.

Of course, any individual or organization interested in the topic of archival interoperability, either within or outside Europe, is welcome to join! As mentioned above, we are especially interested in practical pilots of the specification and your participation in the DILCIS Board.

6 CONCLUSIONS

Ongoing access by citizens to information created by governments, and efficient reuse and availability of crucial business information is a *sine qua non* of the modern world. But access and re-use of government information of long-term value depend, crucially, on ensuring the reliable and error-free movement of records between government business systems and the long-term repositories charged with the responsibility of providing ongoing access to those records.

Additionally, we need to note that the movement of records between systems may occur many times during their lifespan and requires robust interoperability between those systems that, as demonstrated by the E-ARK project operational pilots, is much more cost-effective using a set of common and international specification and principles.

An average archival organization might have to deal with dozens of different specifications, standards and software components, all of which need to be actively maintained and further developed. Archival interoperability, especially the use of common technical specifications for information packages, can significantly reduce the resources needed for such activities for any single institution.

However, practical interoperability is not only about the availability of technical specifications but also about promoting expert discussions and global sharing of best practices as well. The European Commission has recognized this need and is now supporting the eArchiving Building Block based on the outcomes of the initial E-ARK project (2014 – 2017). The authors of this paper invite everybody interested in this topic to join the effort, either from Europe or beyond!

ACKNOWLEDGMENTS

The initial E-ARK project was in part supported by the European Commission within the Competitiveness and Innovation Programme 2007-2013, Grant Agreement no. 620998 under the Policy Support Programme.

REFERENCES

[1] *E-ARK Project. D3.1 E-ARK Report on Available Best Practices* (2014). Available from http://eark-project.com/resources/project- deliverables/6-d31-e-ark-report-on-available-best-practices

[2] *E-ARK Project. D4.1 E-ARK Report on Available Formats and Restrictions* (2014). Available from http://www.eark-project.com/resources/project-deliverables/7-e-ark-d41-report-on-available-formats-and-restrictions

[3] *E-ARK Project. D5.1 E-ARK GAP Report Between Requirements for Access and Current Access Solutions* (2014). Available from http://www.eark-project.com/resources/project-deliverables/3-d51-e-ark-gap-report

[4] *CCSDS. Open Archival Information System Reference Model (OAIS)* (2012). Available from http://public.ccsds.org/publications/archive/650x0m2.pdf

[5] *DILCIS Board. Common Specification for Information Packages v1.0* (2017). Available from http://dilcis.eu/specifications/common-specification

[6] *DILCIS Board. E-ARK Submission Information Package v0.14* (2017). Available from http://www.dilcis.eu/specifications/sip

[7] *DILCIS Board. Semantically Marked-Up Record Format v1.7* (2017). Available from http://dilcis.eu/content-types/smurf

[8] *eCH-0165 SIARD-Formatspezifikation* (2016). Available from https://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0165&documentVersion=2.0

[9] *E-ARK Project. D2.5 Recommended Practices and Final Public Report on Pilots* (2017). Available from http://www.eark-project.com/resources/project-deliverables/97-d25-1

[10] *E-ARK Project. D2.4 Pilot Documentation (Part 1)* (2017). Available from http://www.eark-project.com/resources/project-deliverables/87-d24docs-p1-1

[11] *E-ARK Project. D2.4 Pilot Documentation (Part 2)* (2017). Available from http://www.eark-project.com/resources/project-deliverables/88-d24docs-p2-1