Introduction to the E-ARK Specifications

Transcript v.02

[1] Standalone training from the eArchiving Initiative

Welcome to the E-ARK Specifications training course. This unit is part of the standalone training series developed by the eArchiving Initiative. In this session, we will guide you through the concepts, structures, and practical applications of the E-ARK specifications. By the end, you'll have a clear understanding of how these standards support long-term digital preservation and ensure interoperability across different systems.

[2] Lesson Structure

Lesson 1 offers a general introduction to the E-ARK specifications. In this video, you will become familiar with the overall package structure and gain a brief overview of both the information package specifications and the content type specifications.

[3] The Information Package Concept

In today's world, most of the information we create and rely on exists only in digital form. Emails, databases, photographs, official documents – they all make up our collective memory. But digital files are fragile. Without careful preservation, they can be lost to time, changing technology, or simple neglect.
At its heart, digital archiving is about making sure that digital materials remain accessible, understandable, and trustworthy – not just for a few years, but for decades, or even centuries.

To achieve this, archivists use the concept of an **information package**. You can think of it as a carefully prepared container that brings together the digital files themselves, plus the essential information that explains what they are, where they come from, and how they should be read in the future.

By following this structured approach, archives ensure that digital records are not only safely stored, but also remain meaningful and usable in the long run. In short, information packages are the foundation of trustworthy digital preservation, protecting our cultural and institutional memory for the future.

[4] OAIS Information Packages

The Open Archival Information System, or OAIS, is the internationally recognised framework for long-term digital preservation. It doesn't describe a single technology, but rather a reference model – a kind of blueprint that defines the functions and responsibilities of a trustworthy digital archive.

[5] OAIS Information Packages

At the centre of the OAIS model are the information packages. These come in three main forms: the **Submission Information Package**, or SIP, which is created and delivered to the archive by the producer; the **Archival Information Package**, or AIP, which is the internal version preserved and managed within the archive; and the **Dissemination Information Package**, or DIP, which is what the archive provides to users, the consumers. Together with OAIS processes, these packages ensure that digital materials are not only stored, but remain understandable and usable over time, no matter how technologies evolve.

[6] E-ARK Information Packages

Although the OAIS model is widely used, it's important to remember that it does not prescribe the exact internal structure of information packages. In other words, OAIS tells us that there should be SIPs, AIPs, and DIPs, and what they need to achieve, but it leaves open how they are actually organised, documented, or exchanged in practice. This flexibility is useful, but it also means that without further guidance, archives and institutions might each implement their packages in different ways, making interoperability difficult.

To address this, more detailed and structured approaches have been developed. One of the most significant is the **E-ARK specifications**, produced under the European Commission's **eArchiving Initiative**. These specifications build on the OAIS model and define common, standardised formats for information packages. By following them, organisations can ensure that their archival packages are consistent, interoperable, and easier to share across systems and countries. In this way, E-ARK provides the practical detail that complements the OAIS reference model.

[7] E-ARK Information Packages

An E-ARK information package provides a flexible structure for storing the content files together with all the relevant metadata. At first glance this structure may seem complex, but we'll explore its different parts step by step.

In the following example, we'll imagine a digital archiving scenario where a set of office documents needs to be packaged for submission to an archive.

[8] E-ARK Information Packages – Step-by-Step Guide

Let's begin with some files. They can be in any format, but in this scenario we'll start with a few Office documents. The actual files – the content to be archived – are always stored in a folder called *data*.

[9] E-ARK Information Packages – Step-by-Step Guide

Alongside the files, we also include metadata that describes them. Metadata is simply information about the files – it describes their content, origin, format, and context, so that they can be properly understood and managed in the future. The most widely used metadata standards are EAD, Dublin Core, and ISAD(G). With E-ARK packages you can work with any XML-based metadata format, but in practice we recommend using either EAD or Dublin Core.

[10] E-ARK Information Packages – Step-by-Step Guide

There are different types of metadata, each describing a particular aspect of the archived content. Metadata that records the origin, format, and context of the files is known as *descriptive metadata*. Another key type is *preservation metadata*, which documents the actions and events that take place during the preservation of the content. The de facto standard for preservation metadata is the PREMIS format. In the E-ARK package structure, these different types of metadata are kept in separate metadata folders. The structure is flexible, so you can extend it if your package includes other kinds of metadata.

[11] E-ARK Information Packages – Step-by-Step Guide

As you can see, we've only just begun and already a clear structure is taking shape. To formally define this structure, E-ARK relies on the METS format. METS, which stands for Metadata Encoding and Transmission Standard, is an XML-based format designed to describe the structure of complex digital objects. In an E-ARK package, it works like a manifest or map, linking together the content files and the various types of metadata.

[12] E-ARK Information Packages – Step-by-Step Guide

Now, let's imagine we have two different file formats of the same documents: the original Office files and their PDF versions. We want to preserve both, because while the Office files are the originals, the PDFs are often easier to share and use across different systems. A similar situation arises with images, where both high-resolution and low-resolution versions might be archived together.

[13] E-ARK Information Packages – Step-by-Step Guide

E-ARK handles this neatly by using the concept of *representations*, which allow multiple versions of the same content to be stored side by side within one package.

[14] E-ARK Information Packages – Step-by-Step Guide

Each representation can also have its own metadata, describing only the files within that particular version.

[15] E-ARK Information Packages – Step-by-Step Guide

To keep them separate, each representation is placed in its own uniquely identified folder structure.

[16] E-ARK Information Packages – Step-by-Step Guide

And each of these representation folders can also contain its own METS file, describing the internal structure of that folder.

[17] E-ARK Information Packages – Step-by-Step Guide

As we build up this structure, it's important to distinguish between mandatory and optional elements. To make this clear, we use a simple graphical notation: boxes with solid lines represent mandatory elements, dashed lines indicate recommended elements, and dotted lines show optional elements. This makes the diagrams easier to read and ensures consistency in how we describe packages.

[18] E-ARK Information Packages – Step-by-Step Guide

The Schemas folder of an E-ARK package contains the XML schemas that define and validate the metadata formats used in the package. Including these schemas ensures that the package can always be checked for consistency and interpreted correctly, even many years into the future. It is therefore important to place all the XML schemas used in the package into this folder, so that no part of the metadata becomes unreadable or ambiguous over time.

[19] E-ARK Information Packages – Step-by-Step Guide

And finally, the *Documentation* folder of an E-ARK package is used to store any additional explanatory materials, such as reports, transfer notes, or contextual information that help future users understand the package. These documents are not part of the formal metadata, but they provide valuable background and clarity for archivists and researchers.

[20] E-ARK Information Packages – Step-by-Step Guide

As we have seen, alongside metadata and the METS file, each representation may also include its own documentation folder, containing supporting materials specific to that representation.

[21] E-ARK Information Packages – Step-by-Step Guide

In addition to the standard folders, you can also include optional folders in the package, giving you the flexibility to capture special requirements or project-specific needs.

[22] E-ARK Information Packages – Step-by-Step Guide

And of course, these additional folders can also be included at the representation level, ensuring that each representation has the flexibility to carry its own supporting material.

[23] E-ARK Information Packages – Step-by-Step Guide

So, this is what the complete structure of an E-ARK information package looks like, bringing together content, metadata, documentation, and structural descriptions into a single, well-organised package.

[24] E-ARK Information Packages – Step-by-Step Guide

In most cases, however, the structure is much simpler, so you don't always need to build the full structure to create a valid E-ARK package. Keep in mind that only these elements are mandatory.

[25] E-ARK Information Packages – Step-by-Step Guide

However, it is generally recommended to include additional elements, as they improve the clarity, usability, and long-term preservation of the package.

[26] E-ARK Information Packages – Step-by-Step Guide

And if needed, you can also make use of the optional parts, which provide further flexibility for specific requirements or contexts.

[27] E-ARK Common Specification

The E-ARK specifications cover all three types of information packages defined in the OAIS model: Submission Information Packages (SIPs), Archival Information Packages (AIPs), and Dissemination Information Packages (DIPs). Each of these serves a different purpose—whether it's transferring content into the archive, preserving it over the long term, or providing access to users.

However, much of the structure and content of these packages is the same. To avoid unnecessary repetition, E-ARK has created a Common Specification, which defines all the shared elements. The individual SIP, AIP, and DIP specifications then build on this common foundation, describing only the differences and package-specific requirements. This approach ensures consistency, reduces complexity, and makes the specifications easier to use in practice.

[28] E-ARK Content Type Specifications

In addition to the package specifications, E-ARK also provides Content Type Specifications (or CITS). These define how specific types of digital content—such as databases, geospatial data or electronic records—should be prepared and packaged for long-term preservation. Each content type has unique characteristics and requirements, so the specifications ensure that both the content and its metadata are represented in a consistent and preservation-friendly way.

The Content Type Specifications build on the E-ARK Common Specification, aligning the unique needs of different data types with the shared structure of E-ARK information packages. This ensures that archives can handle a wide variety of digital content while maintaining interoperability and clarity across systems. By following these specifications, organisations can be confident that their digital objects remain accessible and understandable in the future.

The E-ARK Content Type Specifications currently include:
- CS Archival Information (for descriptive metadata)
- CS Preservation Metadata
- SIARD and CITS SIARD (to archive databases)
- CITS ERMS (to archive records from an electronic records management system)
- CITS Geodata (to archive geospatial data)
- CITS eHealth1 and eHealth2 (to archive healthcare related data)
- and CITS 3D (to archive 3D product model data)

[29] DILCIS Board

The E-ARK specifications are maintained and developed by the DILCIS Board (Digital Information LifeCycle Interoperability Standards Board). The DILCIS Board is an international group of experts committed to maintain and

sustain a set of **interoperability specifications** which allow for the **transfer, long-term preservation, and reuse of digital information** regardless of the origin or type of the information. The Board is responsible for ensuring that the specifications remain up to date, practical, and aligned with the needs of the digital preservation community. By coordinating feedback from practitioners, archives, and system providers, the DILCIS Board supports the long-term sustainability of the specifications and promotes their adoption across Europe and beyond.

[30] DILCIS Board on GitHub

The DILCIS Board manages its development work openly on GitHub. This page hosts the official E-ARK specifications, related documentation, and tools. By using GitHub, the Board ensures transparency and enables collaboration, allowing the community to report issues, suggest improvements, and contribute directly to the ongoing refinement of the specifications.

[31] Thank you

Thank you for watching this overview! To explore the details further, please check out the in-depth videos in this E-ARK Specifications Course.