

European **A**rchival **R**ecords and **K**nowledge Preservation

#earkproject

www.eark-project.eu

@EARKProject

Coping with the data explosion

Interoperability solutions from the E-ARK project

Kuldar Aas, Technical Coordinator, E-ARK Project

Kuldar.Aas@ra.ee

ICA International Congress, Seoul, 5-10 September 2016



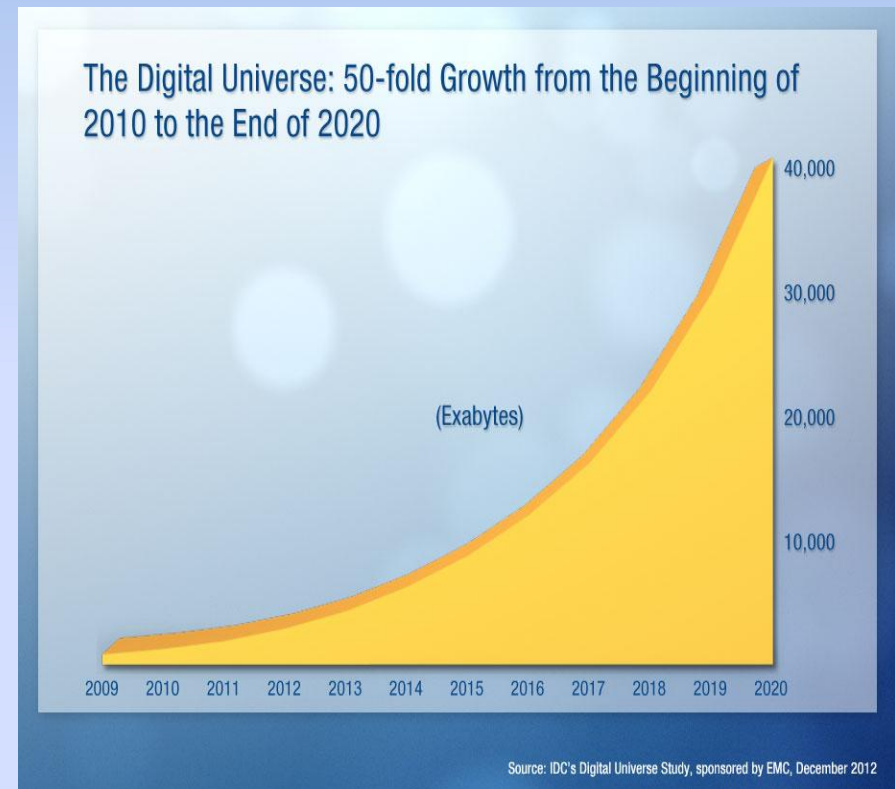
Topics

- The “Issue”
- What’s E-ARK
 - Specifications
 - Tools
- Life beyond E-ARK



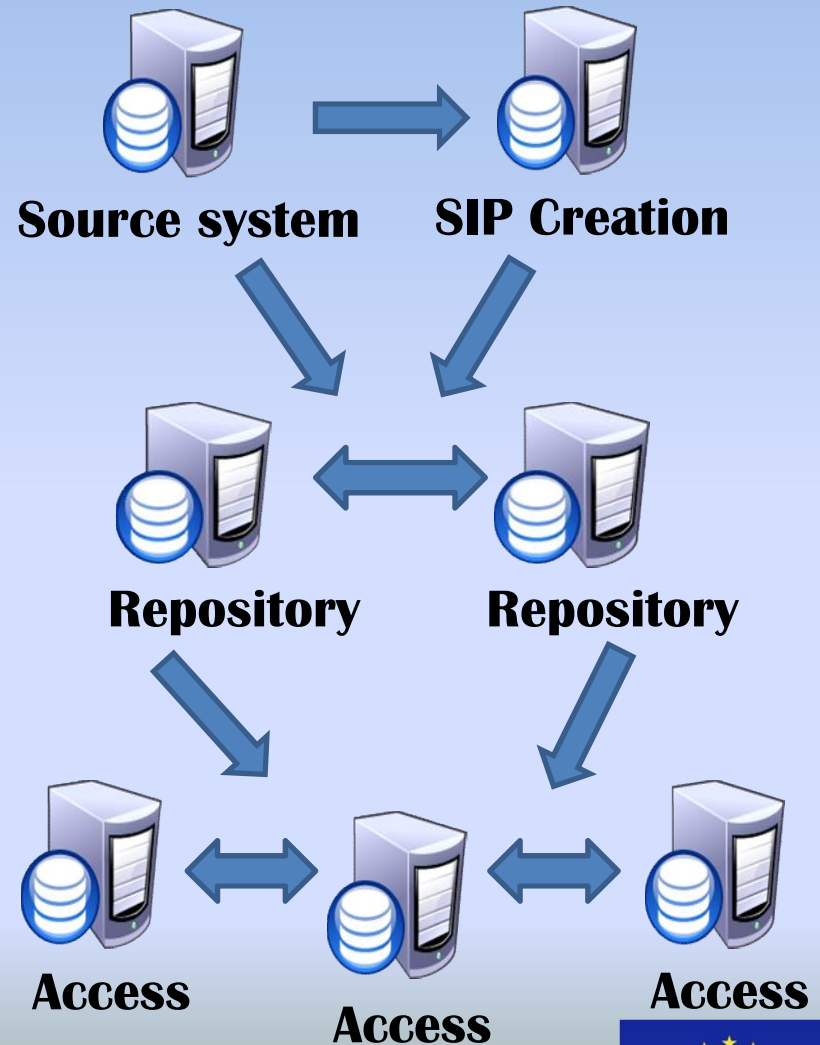
Status quo

- Continuous growth in (archival) records
- Data providers and system developers ignorant of archiving needs and specifications
- Bespoke local development needed far too often
- Lacking funding and skilled resources



Interoperability?

- Multitude of national and institutional standards
- International standards (e.g. METS, PREMIS) exist but leave room for interpretation
- Technical and semantic interoperability is lacking for practical purposes





STATENS ARKIVER

THE DANISH NATIONAL ARCHIVES



THE E-ARK PROJECT IS CO-FUNDED BY THE EUROPEAN COMMISSION UNDER THE ICT-PSP PROGRAMME

www.eark-project.eu



THE NATIONAL ARCHIVES OF NORWAY



RAHVUSARHIIV
THE NATIONAL ARCHIVES OF ESTONIA



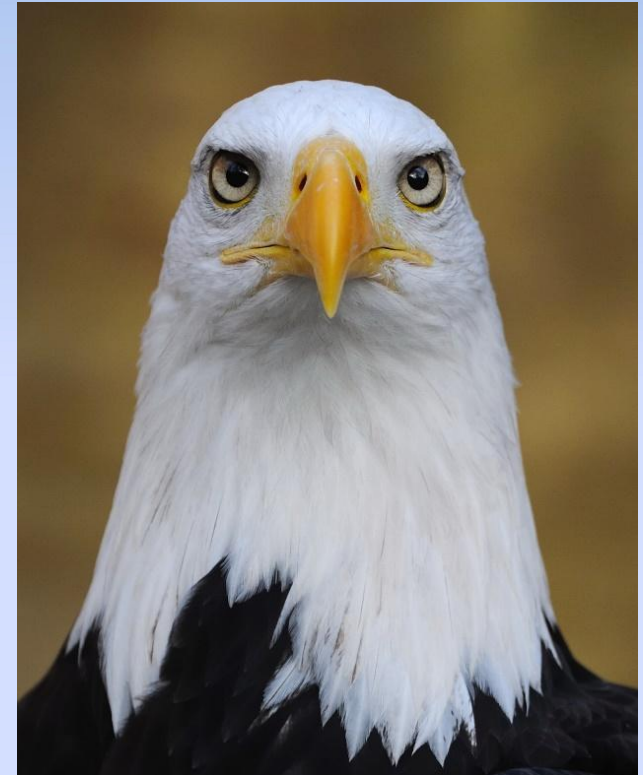
University of
Portsmouth



E-ARK

Vision: All digital preservation systems receive, store and provide access to information regardless of its size, type or format according to a set of agreed principles which allow systems to identify, verify and validate the information in a uniform way

Goal: Interoperability between data sources, archives and reuse environments is improved to a point where digital preservation tools can be reused across borders and institutions



E-ARK: core benefits

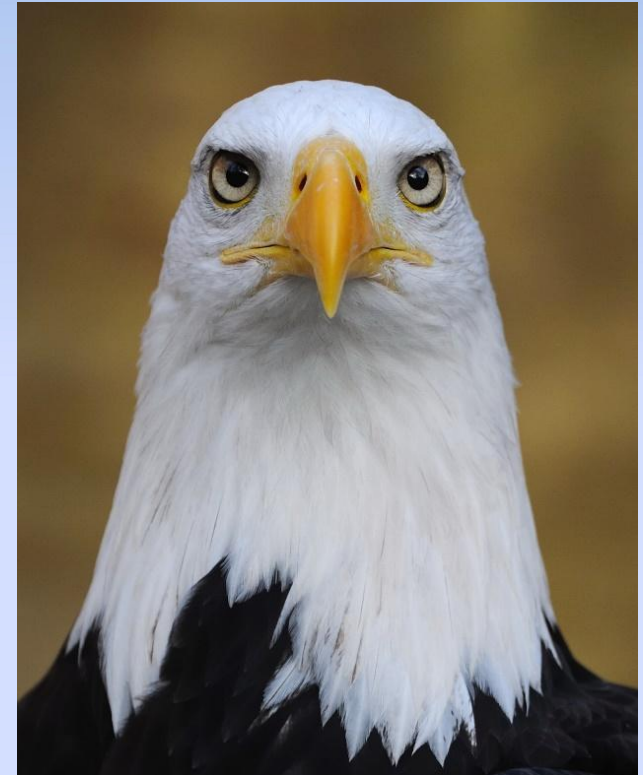
*Extended possibilities for
collaboration in tool development*



*Shared development = less costs
for each individual partner*



*Digital preservation is affordable
for everyone!*



Outcomes

Standardisation of available best-practices

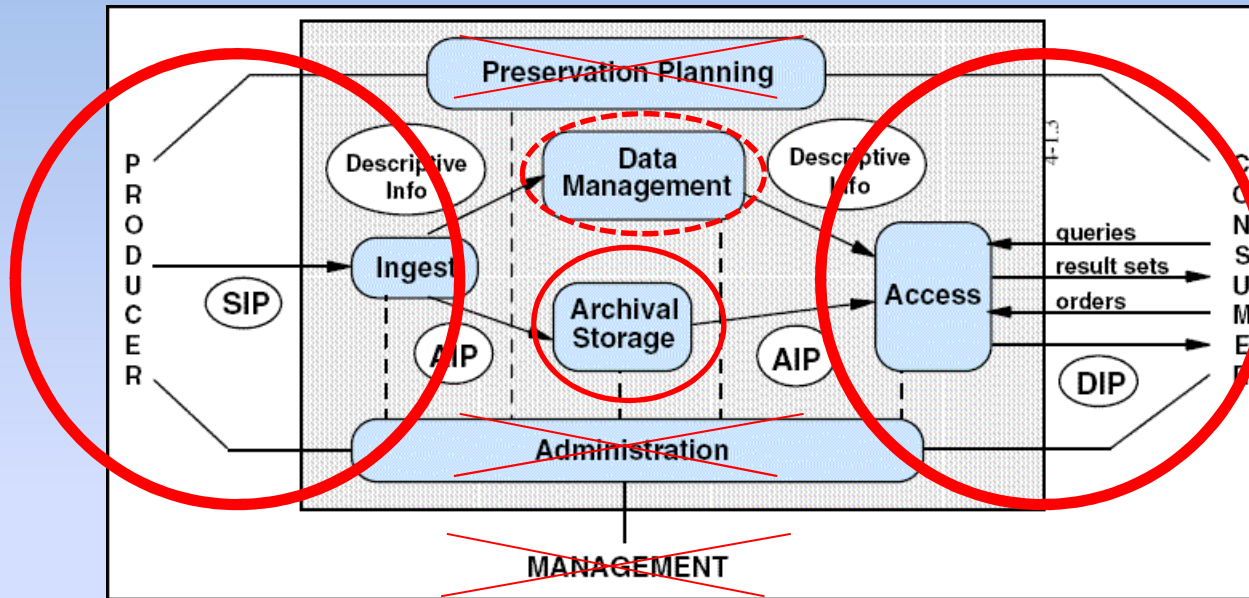
- Common SIP, AIP and DIP format specifications
- Pre-ingest, ingest and access workflows

Open source tools

- to implement the specifications
- to be scalable, modular, robust, computational, and adaptable
- to be implemented individually or as an integrated reference implementation



Scope



SIP

- Package prepared by Pre-Ingest WP3

AIP

- Package created for long-term archive WP4

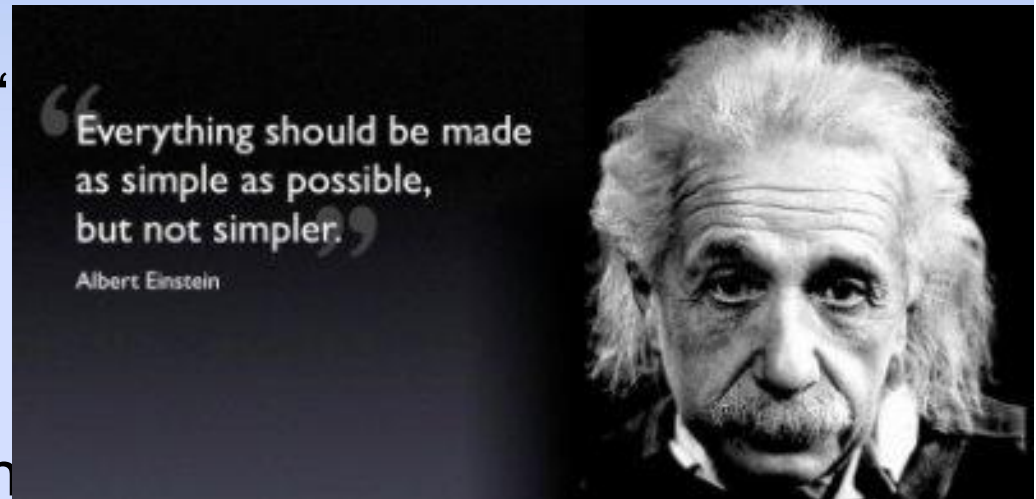
DIP

- Package created for access WP5



E-ARK Specifications

- Guiding principles in E-ARK Information Package standardisation
 - Focus on the needs of the „normal workflow“
 - Led by current implementations across the world
 - Careful balance between standardisation and simplicity
 - Robust and scalable



E-ARK Specifications

E-ARK Common Specification

E-ARK Information Package Specifications

E-ARK SIP

E-ARK AIP

E-ARK DIP

Content Information Type Specifications

SIARD2

SMURF

Geo-data



...



Common Specification

- **E-ARK Common Specification for Information Packages**
 - Provides core principles for any IP
 - Most crucial: a common structure
 - Influenced by the Swedish CS
 - Aims to support automation of high-level package identification and validation (integrity, technical validity, etc.)
 - Is built on widely used XML standards (METS, PREMIS)
 - Concentrates on structural metadata, includes elements of administrative and technical metadata
 - Is flexible enough to be used for any type of data
 - Allows for further specification and localisation if needed

Project 620998: European Archival Records and Knowledge Preservation - E-ARK



PROJECT DELIVERABLE

Project Acronym: E-ARK
Grant Agreement Number: 620998
Project Title: European Archival Records and Knowledge Preservation

DELIVERABLE DETAILS

DELIVERABLE REFERENCE NO.	N/A
DELIVERABLE TITLE	Introduction to the Common Specification for Information Packages in the E-ARK project
REVISION	0.14

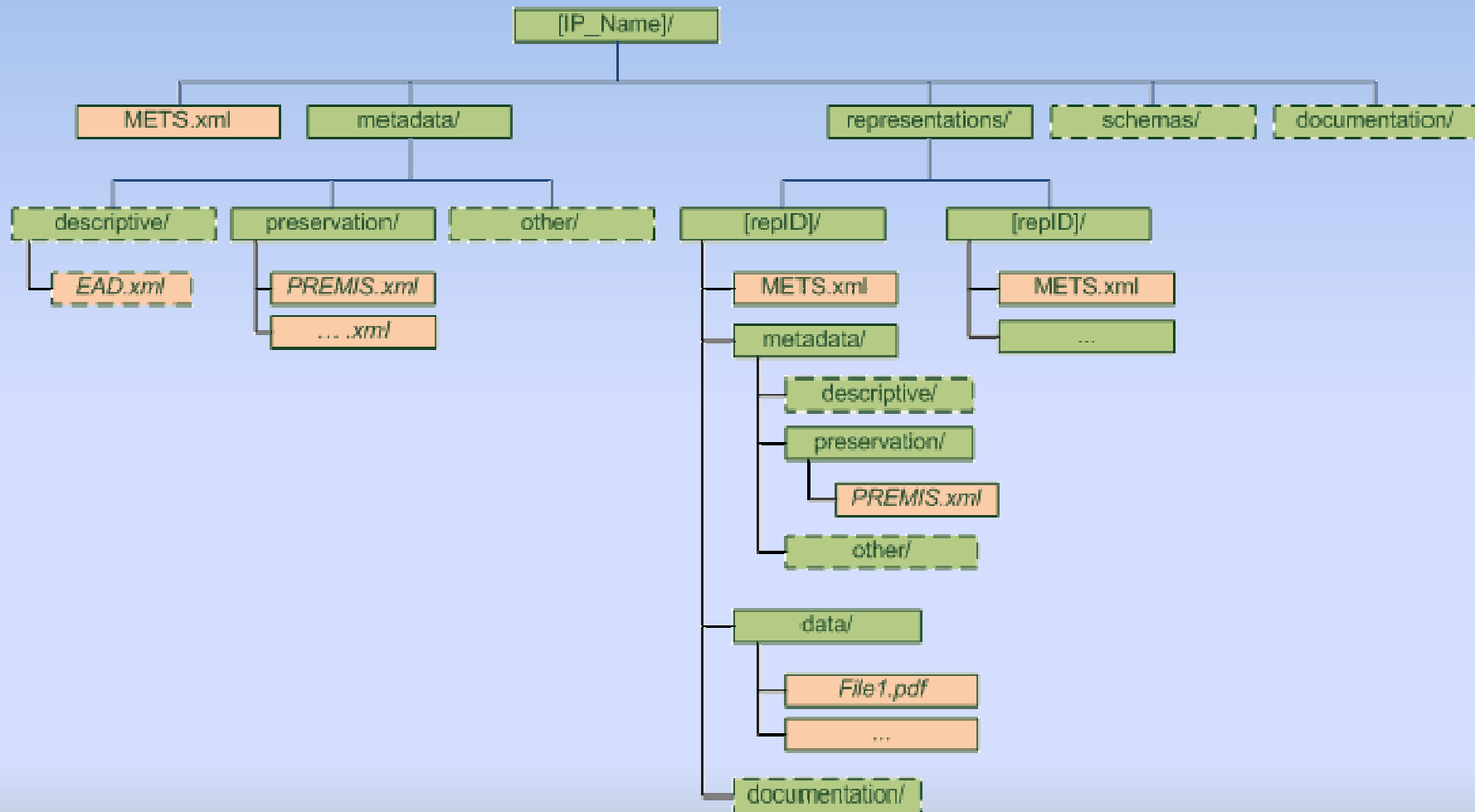
AUTHOR(S)

Name(s)	Organisation(s)
Tarvo Kärberg	National Archives of Estonia (NAE)
Karin Bredenberg	National Archives of Sweden / ES Solutions (ESS)
Björn Skog	ES Solutions (ESS)
Anders Bo Nielsen	Danish National Archives (DNA)
Kathrine Hougaard Edsen Johansen	Danish National Archives (DNA)
Alex Thirfays	Danish National Archives (DNA)
Sven Schlarb	Austrian Institute of Technology (AIT)
Andrew Wilson	University of Portsmouth / University of Brighton
Kuldar Aas	National Archives of Estonia (NAE)

Page 1 of 38
Internal Deliverable: Introduction to the Common Specification for information Packages in the E-ARK project

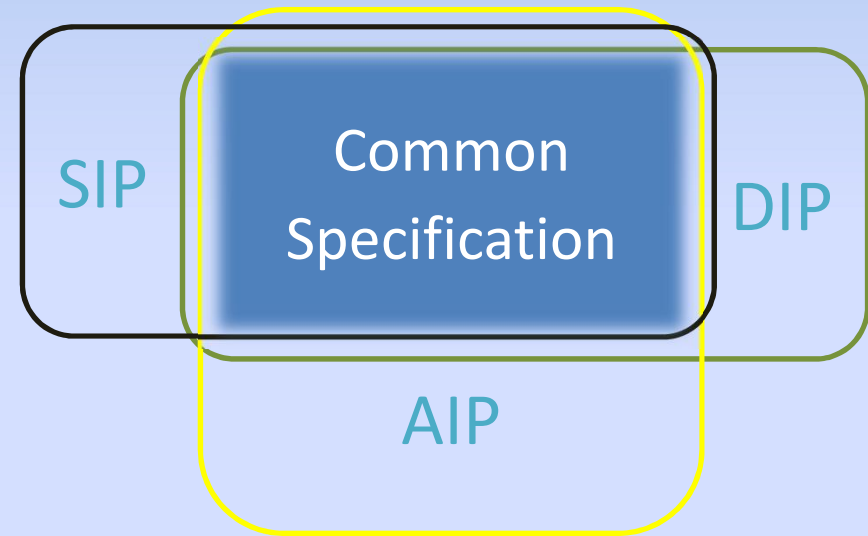


Common Specification: structure



E-ARK IPs

- E-ARK SIP
 - Administrative metadata for ingest and pre-ingest processes
 - Transfer agreement
- E-ARK AIP
 - Preservation metadata
 - Additional structural layer to allow keeping the original SIP untouched
- E-ARK DIP
 - Administrative metadata (log the creation of the DIP)
 - Description of complex access environments in PREMIS



E-ARK Content Information Types

- E-ARK Content Information Types allow description of the details of the package content
 - Descriptive metadata (e.g. sub-specification of EAD, MARC, CIDOC-CRM)
 - Technical metadata (e.g. specific requirements for digital image metadata)
 - Internal structure of the data folder (e.g. the physical structure of series and records in an ERMS-derived Information Package)
 - Specific representation information (e.g. define the composition of files which constitute an archived relational database and the appropriate viewer)
- Content Information Types can be
 - Data specific (for databases, websites etc.)
 - Metadata specific (for EAD, MARC, Dublin Core etc.)

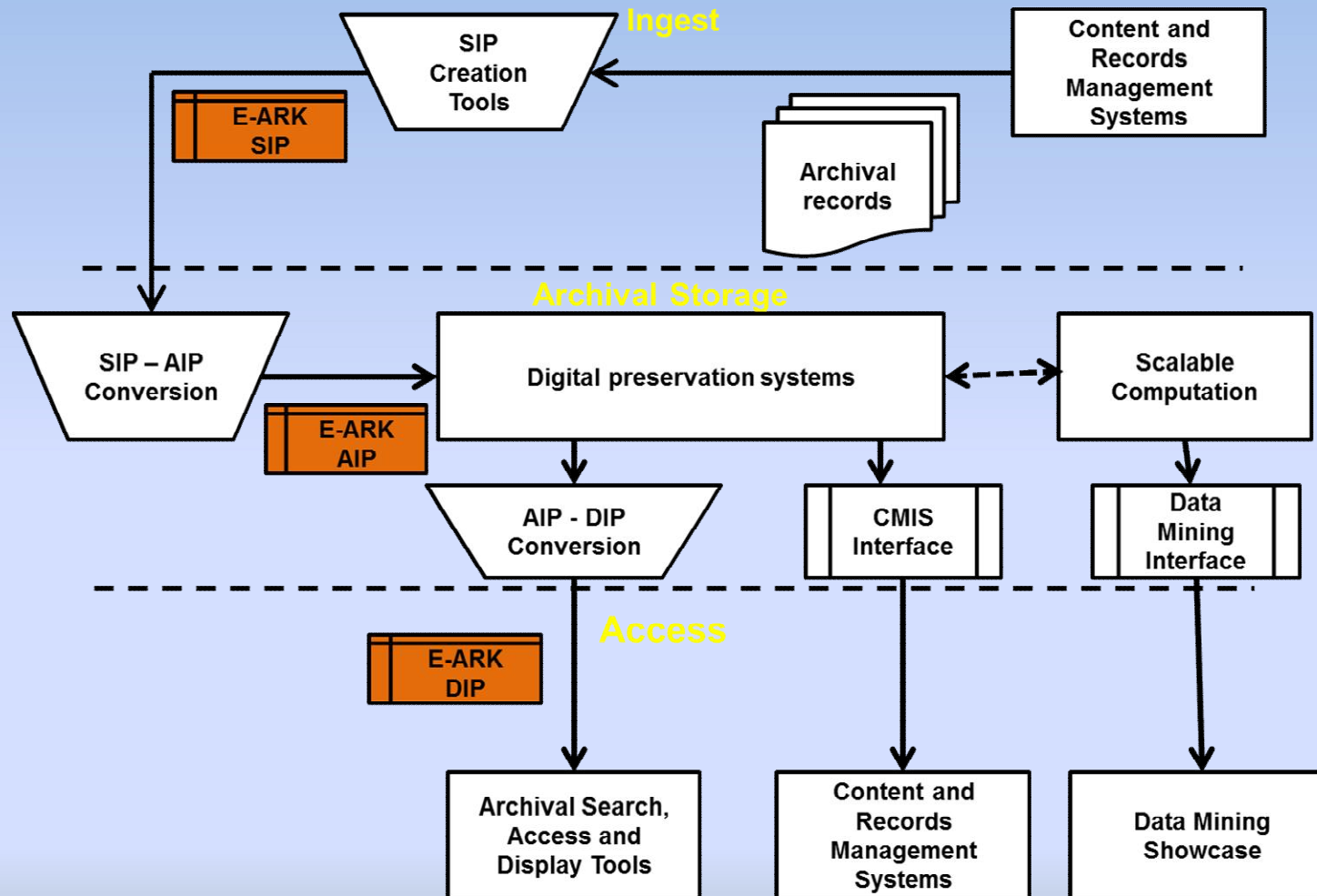


E-ARK Content Information Types

- E-ARK provides three data specific content type specifications
 - SIARD2
 - Specification for relational databases
 - Updates the original SIARD format in regard to improved scalability and standards support
 - In collaboration with the Swiss Federal Archives
 - SMURF (Semantically Marked-Up Records Format)
 - Specification for ERMS-exported and „loose“ records
 - Extends EAD with records management elements
 - Partly synchronised with MoReq2010
 - Geo-data
 - Specification to define a basic descriptive layer
 - Defines the structure and location of geo-info elements within the Information Package
- E-ARK provides a metadata specific content type for EAD
 - Possible to be used within any Information Package



E-ARK: tools



Sustainability

- E-ARK ends on 31st of January 2017 ..
- Maintenance of specifications ensured through the DLM Forum
 - Specific workgroup on standards currently being set up
 - Non-members of DLM welcome to join!
- Discussions to ensure the sustainability of tools in collaboration with the Open Preservation Foundation (OPF)



Questions?

Make sure to also check:

www.eark-project.com for all deliverables and specifications and
<https://github.com/eark-project> for all software

Meets us at:

iPRES 2016, October 3-6, Bern, Switzerland

E-ARK Demo Day, November 8, Brussels, Belgium

DLM Forum meeting and E-ARK Demo Day, November 15-17, Oslo,
Norway

E-ARK Final Conference, December 6-8, Budapest, Hungary

