



# Creating E-ARK conformant Archivematica preservation workflows

Stephen Mackey, Penwern

*eArchiving Initiative Training Webinar*

# Agenda

- Background: Penwern
- Background: HaDEA OneClick and Engineering & Science Library projects
- Archivemata: what is it, strengths, limitations
- The E-ARK Archivemata Toolkit
- Demonstration: AIP and DIP creation
- Case studies, integration in systems
- Current status, the future

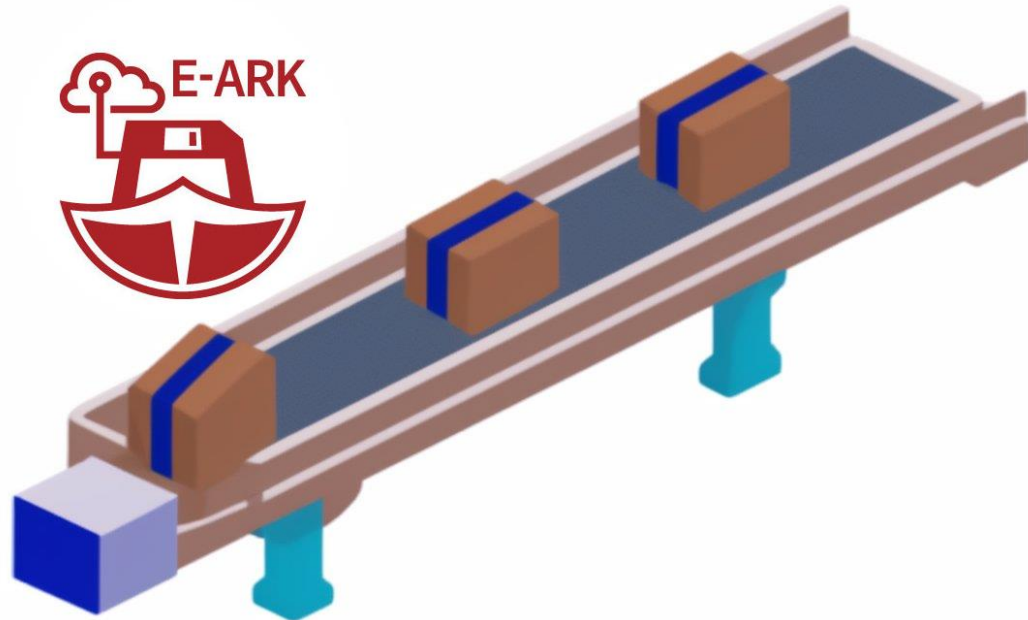
# Penwern

- Digital preservation and eArchiving consultants, project managers, product managers and software developers since 2017
- Members of E-ARK since 2020
  - eHealth1 (E-ARK3)
  - 3D Product Model and 3D Heritage CITS – new
  - Package validation testing
- Provide services to Artefactual Inc (Archivematica)
- Subcontractor to Piql AS on E-ARK3 and HaDEA projects



# Background: HaDEA OneClick and ESL

# OneClick



- August 2021 – January 2023
- XAMK (Finland), Disec (Finland), EKA (Estonia), Piql (Norway)
- Objectives:
  - Create E-ARK SIPs from ERMS exports (in one click)
  - Provide SIP, AIP and DIP support in Disec repository
  - Create Archivematica workflow to accept OneClick SIPs, produce AIPs and DIPs
- Hands-on tutorial  
[https://www.youtube.com/watch?v=aXgCv\\_Zlrps](https://www.youtube.com/watch?v=aXgCv_Zlrps)

# Engineering & Science Library

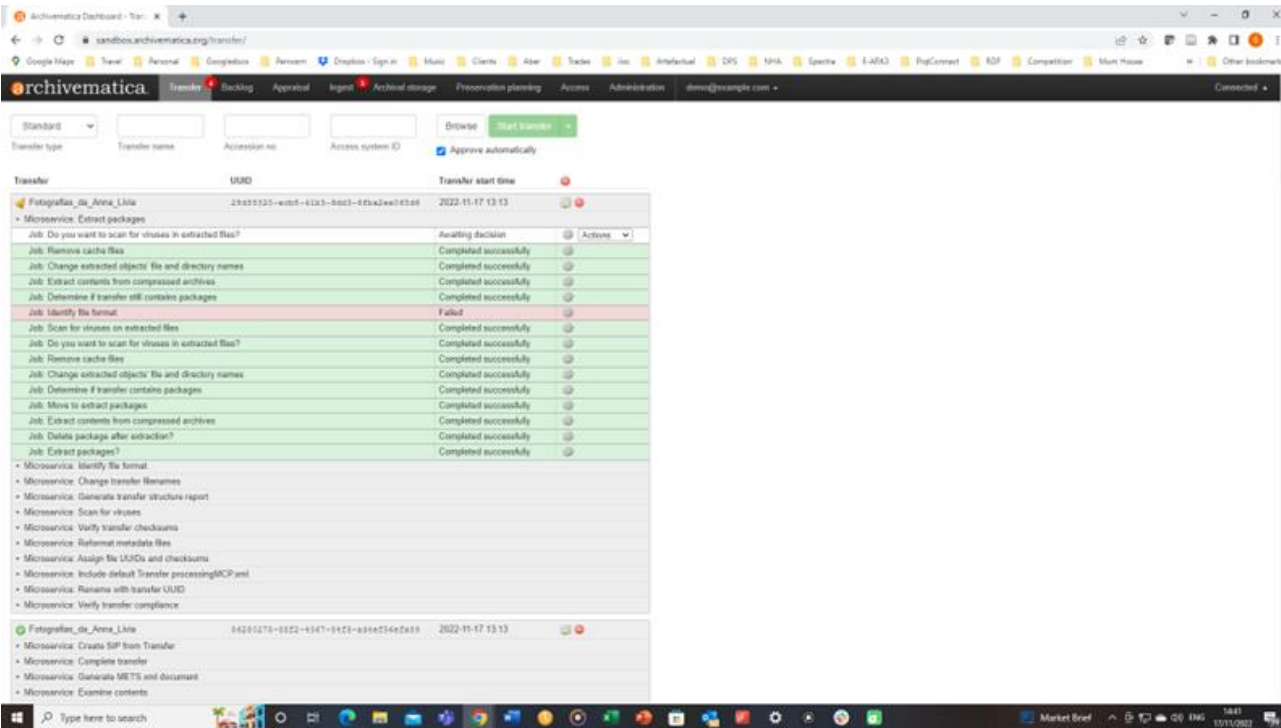
- August 2021 – June 2023
- Piql (Norway, Spain), Airbus Space and Defence (Spain), Central Library of Slovak Academy of Sciences (Slovakia)
- Objectives:
  - Development of on-site (Airbus) and hosted (UKSAV) E-ARK conformant archival systems with embedded Archivematica preservation workflows for science and engineering content
  - Archiving of 2D technical drawings and documents using ERMS CITS (Airbus)
  - Archiving of scientific manuscripts and documents (UKSAV)



# Archivematica

# What is Archivemata?

- An integrated suite of open-source software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model.
- Users monitor and control ingest and preservation micro-services via a web-based dashboard.
- Archivemata uses METS, PREMIS, Dublin Core, the Library of Congress BagIt specification and other recognised standards to generate trustworthy, authentic, reliable and system-independent Archival Information Packages (AIPs) for storage in a preferred repository.
- Archivemata can be found at: <https://www.archivemata.org/en/>





# Why use Archivematica?

- Performant and scalable
- Standards compliant
- Free, open-source solution
- Dashboard control over preservation plan, migration policies, format policy registry
- Will accept external metadata, PREMIS events, file manifests
- Can be integrated via its API and can be configured to run automatically as a service with set policies
- Can be run within a bespoke workflow using the Enduro workflow manager
- Large user base
- Active development and lead developer, Artefactual systems, Inc
- Support and hosting contracts available from vendors including Artefactual Systems

# Some limitations

- Unique Archivematica AIP format
- Unique Archivematica transfer (SIP) format
- DIP production only at Ingest (primarily for AtoM integration)
- Designed for user workflow interaction – detailed UI
- Dublin Core support only in METS
- Not a repository (limited search, view/render functionality)

# The Toolkit

# Who can benefit?

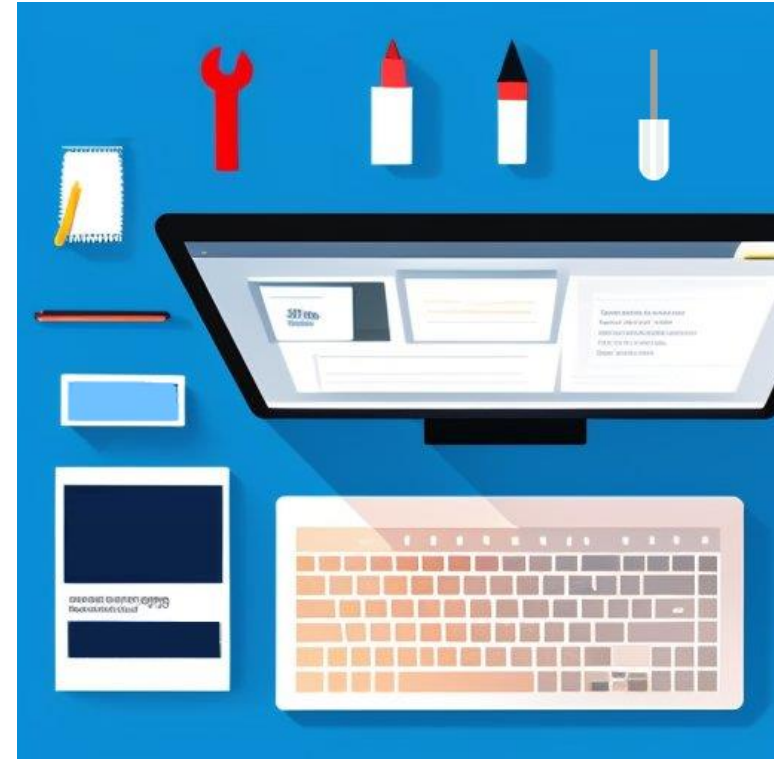
Any organisation that wishes to build a digital preservation system or archive that combines the advantages of:

- Conformance with E-ARK common and package specifications
- The Archivemata digital preservation platform.



# E-ARK Archivemática Workflow Toolkit

- Full E-ARK compliance – CSIP, SIP, AIP, DIP, CITS (may require modification)
- Archivemática is unchanged general distribution
- Free, open-source, can be modified
- Can be integrated into custom workflows using its API
- Uses Enduro, an open-source workflow manager from Artefactual
- Code can be found at: <https://github.com/Penwern/>



# Enduro

*"Enduro is a tool designed to automate the processing of transfers in Archivemata pipelines. It's part of a preservation solution that is being used by the National Health Archive (NHA, Norway) and the National Center for Truth and Reconciliation (NCTR, Canada)."*

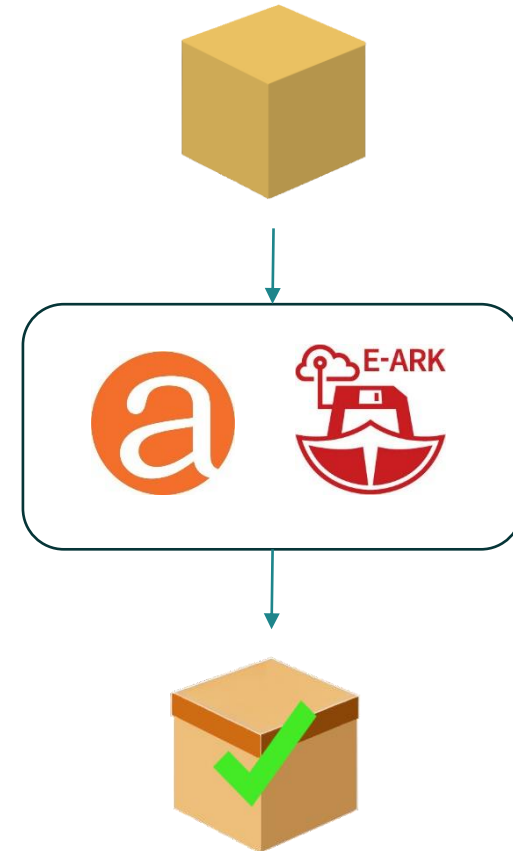
Enduro is a workflow management tool that is controlled by code. The AIP and DIP Creator provide that code such as to manage a workflow that embeds Archivemata into an EARK compliant system.

The version used by the Toolkit utilises Cadence (<https://cadenceworkflow.io/>) as an orchestration engine, a new version is in development that uses Temporal (<https://docs.temporal.io/>) but should be compatible.

Proof of concept project from Artefactual Systems, but in production at NHA and NCTR.

# What do the tools do?

- Accept and validate E-ARK submission packages (SIPs)
- Create Archivemata transfers (SIPs) and initiate their ingest into Archivemata
  - Including Dublin Core metadata into Archivemata METS, other metadata and sidecar files are preserved
  - Import PREMIS events (future)
- Create valid E-ARK AIPs with embedded Archivemata AIPs (preservation objects) as a separate representation
  - Metadata is stored in package /metadata folder
- Create valid E-ARK DIPs on demand from AIPs



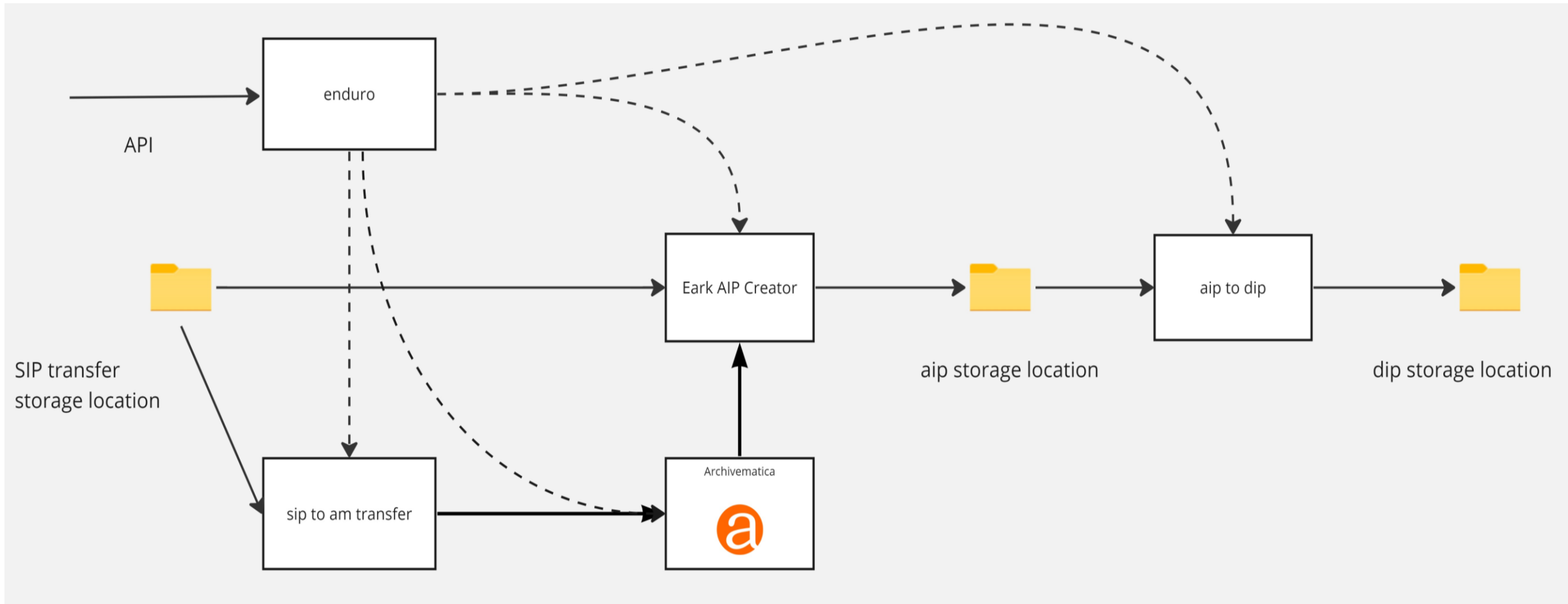
# What do the tools consist of?

- Archivemata, currently version 1.13.2, standard distribution
- Enduro, specific configuration with Go workflow code for the SIP to AIP and AIP to DIP processes
- SIP to Archivemata transfer, Python script
- SIP to EARK AIP, Python script
- EARK AIP to E-ARK DIP, Python script





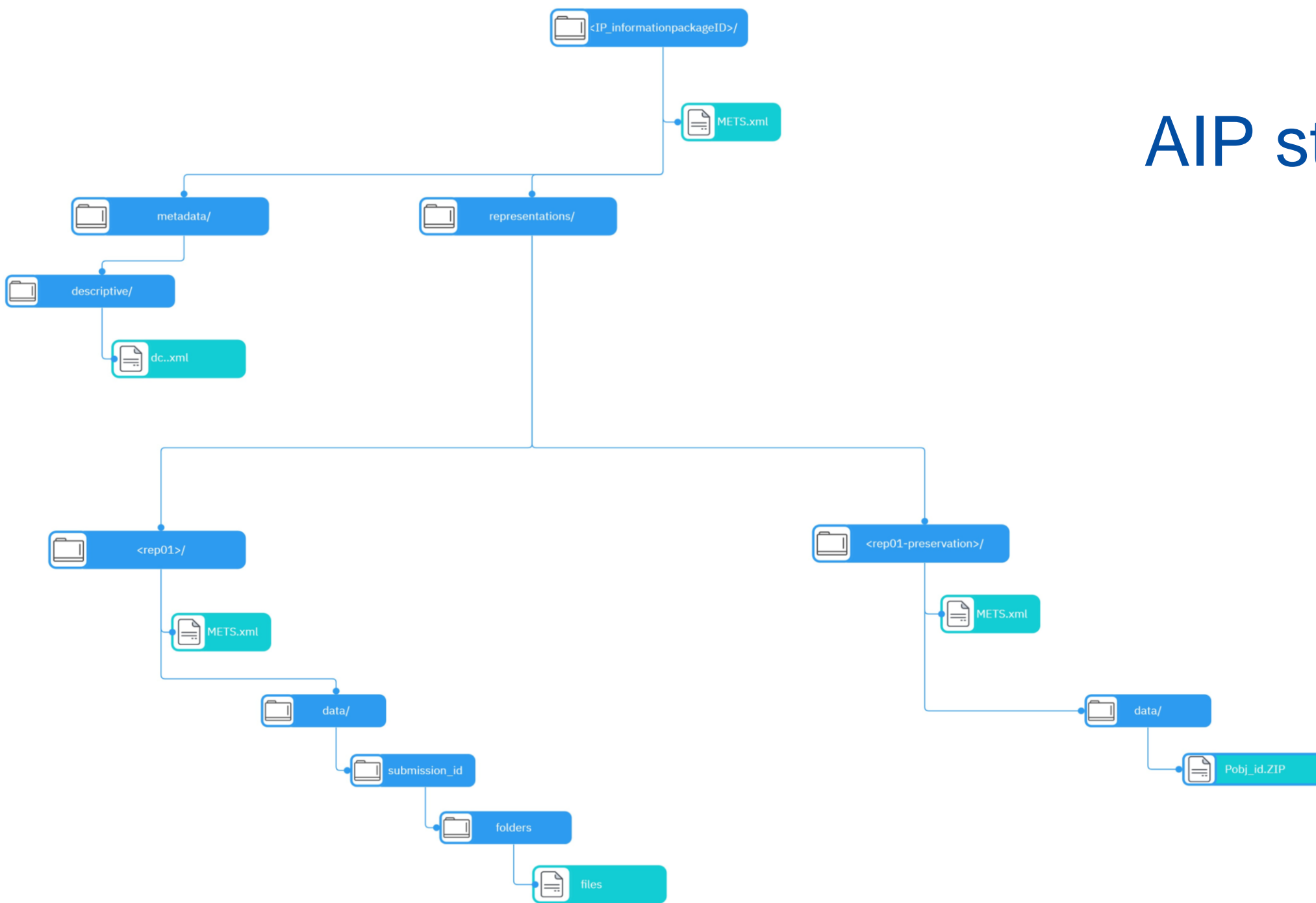
# Workflow



# Key features

- Dublin Core metadata included in the E-ARK SIP is included in the Archivemata transfer, Archivemata METS file and E-ARK AIP metadata folder
- Non-DC metadata is included in the E-ARK metadata folder
- Original files are included in the Archivemata AIP by default (optional)
- Original files are also included in a submission representation in the E-ARK AIP (optional)
- PREMIS events can be imported into the Archivemata METS if submitted in the E-ARK SIP (future)
- AIP folder structure is currently simple but can be expanded (documents, schemas, preservation metadata)
- Can support different CITS (ERMS is supported in Airbus implementation)

# AIP structure



# Demonstration

Video can be found at: [https://www.youtube.com/watch?v=aXgCv\\_Zlrps](https://www.youtube.com/watch?v=aXgCv_Zlrps)

# Case Studies, Integration

# Case study – OneClick

- The E-ARK Archivematica AIP and DIP Creator was used within a proof of concept to provide a simple means to extract data from an Electronic Record Management System (ERMS), create conformant E-ARK submission packages (SIPs) and convert to archival packages (AIPs) and dissemination packages (DIPs).
- The extracts were gigabytes in size
- The output packages were validated by the E-ARK validation team for conformance to the specifications.
- The project also produced a SIP Creator for producing E-ARK SIPS from ERMS exports

# Case Study – Engineering and Science Lab

## Airbus

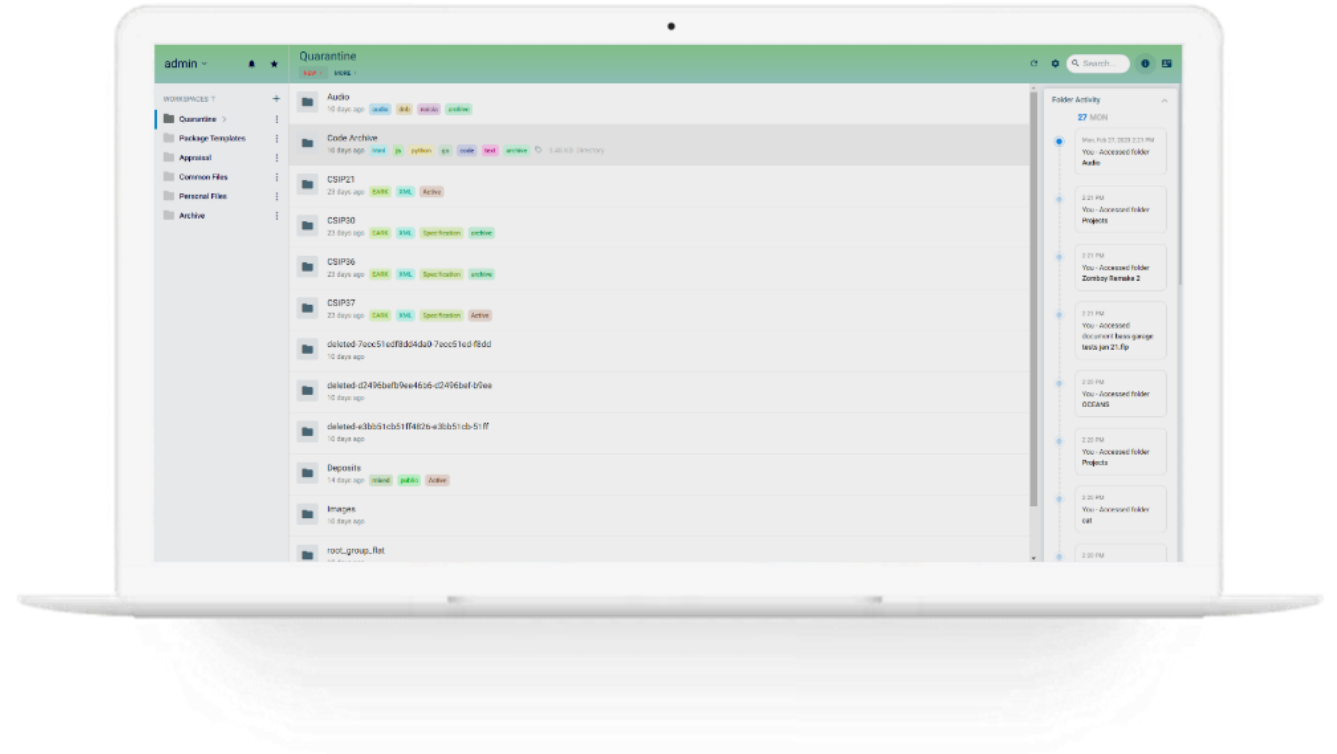
- Certification Quality Documents
- Engineering Technical Documents
- 2D Drawings
- 3D models (in the future)
- Extracts from ERMS using ERMS CITS and metadata schema
- AIP and DIP production with toolkit
- Used nHive application on-premise from Piql for repository and workflow (no Enduro)

## UKSAV

- Digitised books and manuscripts
- Research documents
- Ingest from digitisation workflow into nHive repository software from Piql AS
- Dublin Core metadata
- AIP and DIP production with toolkit
- Used hosted nHive application from Piql for repository and workflow (No Enduro)

# Case Study – Penwern Curate

- Simple, modern, performant platform providing:
  - Pre-ingest
  - Appraisal
  - Description
  - Preservation
  - Search, access
  - Public access

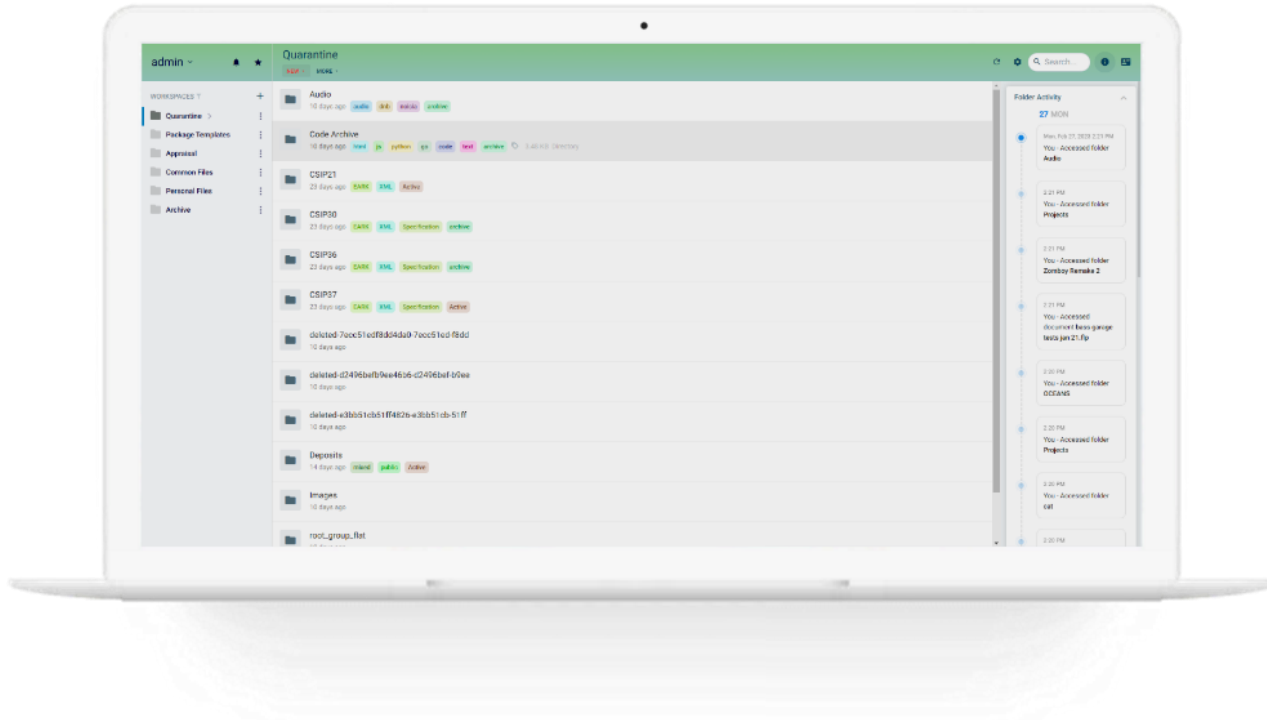


<https://www.penwern.co.uk>





# Penwern Curate



Curate can create E-ARK AIPs from Archivemata AIPs and reassociate package and file metadata in its search system

Curate has inbuilt functionality via Archivemata a3m to normalise objects or collections of objects and create Archivemata AIPs

a3m is a lightweight version of Archivemata being developed by Artefactual Systems, focused on AIP creation. It has neither external dependencies, integration with access systems, search capabilities nor a graphical interface.

# Demonstration

# Toolkit current status

- Code is available on github at: <https://github.com/penwern>
- There is no funding for maintenance or enhancement of the toolkit, but:
  - Penwern can answer simple questions ([info@curate.penwern.co.uk](mailto:info@curate.penwern.co.uk)) or [stephen@penwern.co.uk](mailto:stephen@penwern.co.uk)
  - Penwern can provide development and support services
  - Code can be altered/forked
- Testing has not been done with the Temporal workflow engine
- a3m would be advantageous but has not been tested in the workflow

# Q & A



**Thank you**

## Contact



<https://e-ark4all.eu/>



[info@e-ark-foundation.com](mailto:info@e-ark-foundation.com)



[@EU\\_eArchiving](https://twitter.com/EU_eArchiving)



<https://www.linkedin.com/company/eu-e-archiving-initiative>



<https://www.youtube.com/@e-ark>