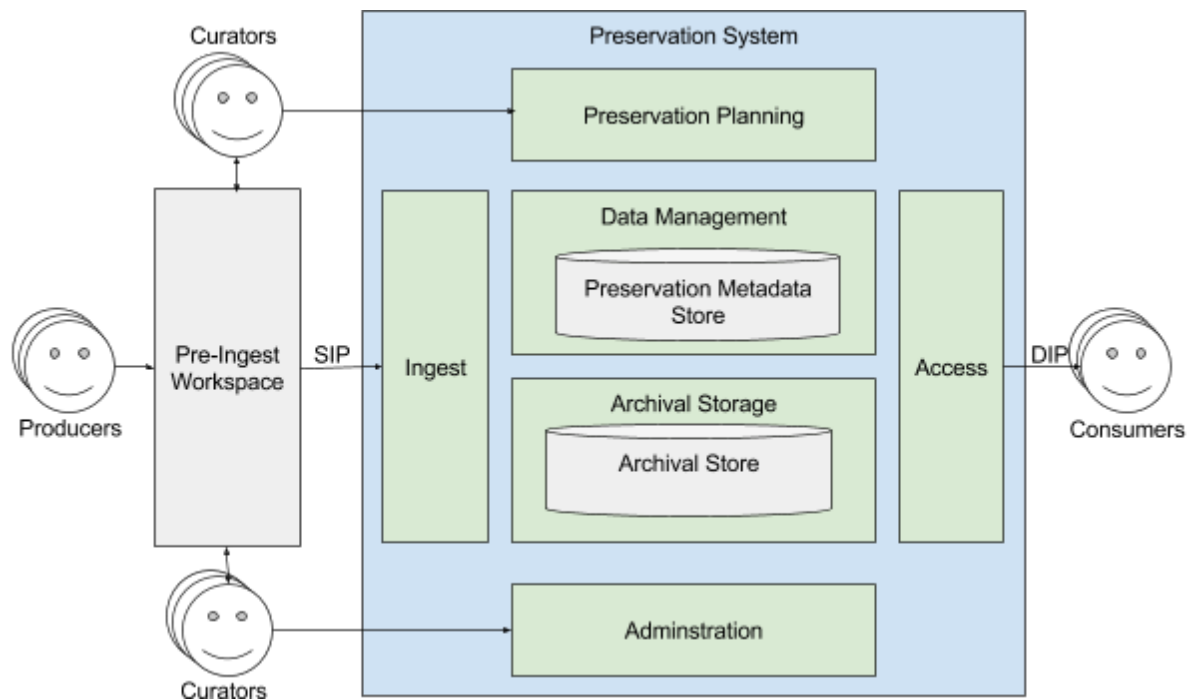


Logical Architecture

Overview

The aim here is not to produce a detailed design. A logical architecture is intended to provide a common understanding and consistent terminology.



Key

- The blue box marks the boundary of the OAIS repository system.
- The green boxes represent the OAIS functional entities of the same name.
- The grey boxes represent storage components, e.g. RDBMS, triple store, network file systems, or cloud storage, etc. Two storage components are coupled with the corresponding OAIS functional elements, the Archival Store and the Preservation Metadata store.
- Smiley faces are external agents, these aren't necessarily human so include external systems also and are analogous to OAIS roles.

OAIS Concepts

The logical architecture is based upon concepts drawn from the [Open Archival Information System \(OAIS\)](#) reference model.

Agents / Roles

The OAIS model describes three roles, producer, management and consumer. The model doesn't define a role for staff who operate the repository on a day to day basis. The management role is a higher level, strategic one. Operational staff are subsumed into the OAIS functional administration and preservation planning entities. Since this a logical system model curatorial staff have been represented as an agent / role.

Producers

Members of staff or external systems that create and submit data for long term preservation. Some external systems might provide data automatically, some might require manual intervention. Examples of producers are:

- UNHCR staff
- Records and Archives staff (a subset of UNHCR staff with responsibility for preservation)
- eSafe content management system
- Orange Logic

Curators

Members of staff or external systems that:

- process data to prepare it for ingest, e.g. augment with data from other systems or providing QA services;
- curate previously ingested material, e.g. making corrections, deleting material; or
- carry out digital curation processes such as preservation planning or format migration.

This is likely to be the responsibility of the Records and Archives staff

Consumers

Any person or system that interacts with the system to find and access preserved information.

- UNHCR staff
- Records and Archives staff
- Researchers
- General public
- AdLib based search and access system

Information Packages

OAIS names a collection of content and associated metadata an information package. It then defines three specific types of information package that model the life-cycle and use of digital content and metadata.

Submission Information Package (SIP)

A prepared information package that is delivered by a Producer for ingest into the preservation system, where it is transformed into AIPs.

Archival Information Package (AIP)

The content and metadata package that is held by the system for long term preservation.

Dissemination Information Package (DIP)

An information package that is derived from AIPs for delivery to a Consumer in response to some kind of access request.

Logical Components

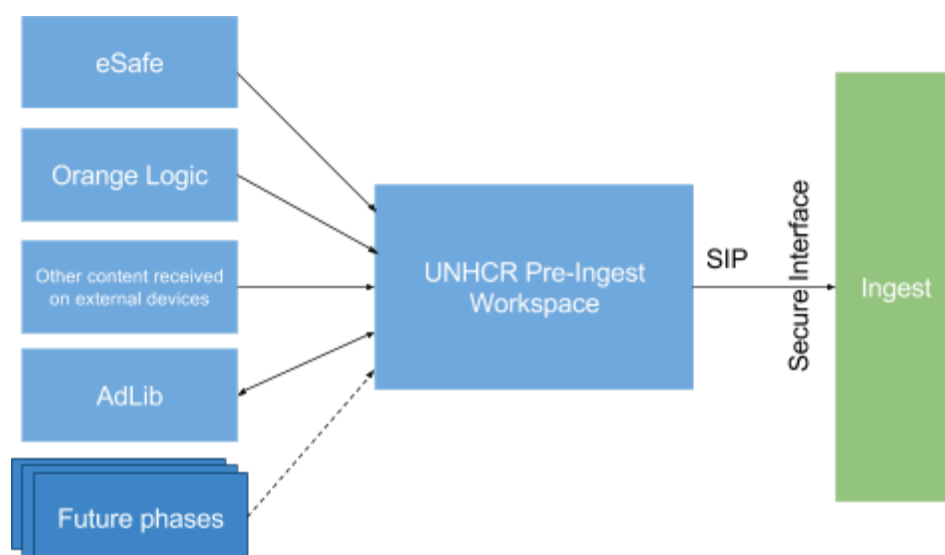
Pre-Ingest Workspace

A secure, local workspace for accessioning and preparing digital content for ingest into the preservation system. It is quite possible, but not certain, that the workspace is purchased alongside the preservation system. In any case the pre-ingest workspace is owned, hosted, supported and administered by UNHCR.

The workspace provides an operational curatorial area where UNHCR curatorial staff can:

- package content and metadata as Submission Information Packages for ingest;
- curate and augment package metadata, e.g.;
 - recording relationships between objects in SIPs;
 - collecting metadata from other UNHCR systems for addition to SIP; and
 - transformation of metadata into SIP forms.

The diagram below illustrates this particular role of the workspace and shows some of the content sources and metadata stores under consideration for initial ingest.



Preservation System

The Preservation System, the subject of the main tender, has been divided into six functional sub-components. The OAIS model classifies them all as functional entities. In addition we have named two infrastructure entities that are key components of the system to be procured. These are identified separately as there are specific non-functional requirements that concern them.

Infrastructure Entities

These have been classified separately as they're fundamentally sub-systems provided by technical solutions. Both are data stores that should ideally be hosted and maintained by a third party, ideally as a cloud solution.

Preservation Metadata Store

This component provides a queryable data store of preservation metadata. This can be:

- Descriptive metadata used for searching for preserved content.
- Full text indexes of content, again for use in finding content of interest.
- Technical metadata, e.g. file format ids, the results of format validation tasks, for use in content analysis and preservation planning.
- Relationships between content items, e.g. versioning, hierarchical file system structures, the grouping of content into logical collections.

The metadata store is almost certainly not provided as a single technology but as a combination of technologies like Relational Database Systems, triple-stores and full text indexing products.

Archival Store

The archival store is a long term preservation store for Archival Information Packages. The store is resilient, providing protection against data loss:

- The store should preserve multiple copies of AIPs in geographically remote locations.
- The store should provide protection against bit-rot, that is the potential corruption of data over time caused by hardware or software issues.

Functional Entities

The OAIS functional entities aren't simply technological solutions, but a combination of technology and institutional processes. That is they are a combination of staff and technology the provide functions of the preservation system.

Ingest

The ingest entity takes Submission Information Packages provided by Producers and transforms them into Archival Information Packages for long term storage. This process includes:

- Validating the contents and metadata provided in the SIP.
- Running preservation tools that perform format identification and validation of the submitted content.
- Adding metadata to the systems Metadata Store to support search, retrieval and preservation planning.
- Transforming the SIP into an AIP, where the SIP contents are augmented with metadata produced during ingest validation.

Archival Storage

This function manages the long-term storage and maintenance of digital materials entrusted to the preservation system. These functions include:

- ensuring archived content resides on appropriate storage;
- ensuring that the bit streams remain complete and renderable over the long term; and
- the retrieval of items in response to access requests by Consumers.

Administration

The administration function has operational day to day management responsibility for an OAIS repository. These functions include:

- Negotiating submission agreements.
- Managing access control.
- Analysis and reporting of system capabilities, e.g. available storage and compute resources
- Auditing and reporting of system users and events.

Data Management

The Data Management function maintains data stores containing:

- descriptive metadata identifying and describing the archived information;
- technical metadata generated by the preservation system; and
- administrative data supporting internal system operations.

Data Management functions include:

- maintaining the databases for which it is responsible;
- performing queries on these databases and generating reports; and
- conducting updates to the databases as new information arrives, or existing information is modified or deleted.

The Data Management function supports search and retrieval of archived content.

Preservation Planning

This function ensures that the content and metadata held in the repository remain

understandable and accessible in the long term. This entails:

- Analysis and reporting of system content.
- Development of preservation plans for at risk content.
- Automated execution of preservation plans at adequate scale.
- Automated QA of the results of preservation plans.

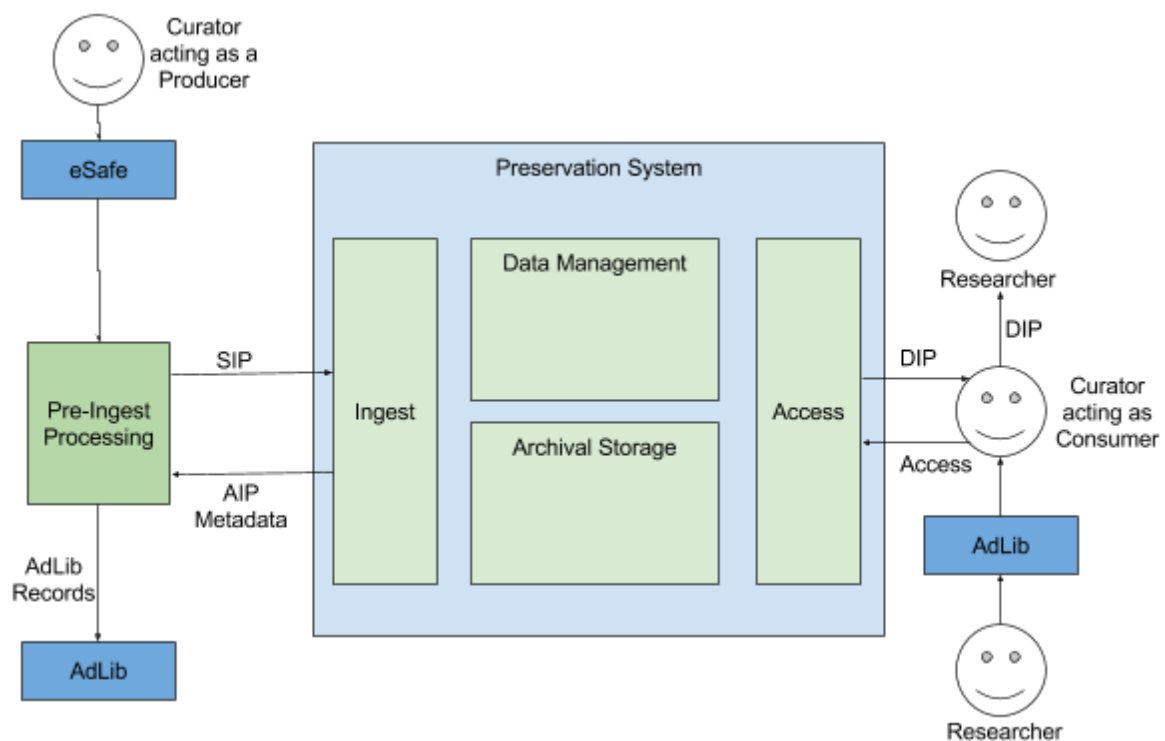
Access

The access function provides the services and functions that support discovery of and access to archived content. This function addresses the interfaces that allow authorized agents to search for and retrieve metadata and content from the repository. The information is returned as a Dissemination Information Package that is often generated on the fly in line with access policy. These responsibilities include:

- supporting queries of repository holdings to determine the existence, description, location and availability of content;
- processing requests for access to content and delivering the resulting DIP to the Consumer; and
- enforcing security and controlling access to archived content.

Proposed Workflow

Physical Architecture

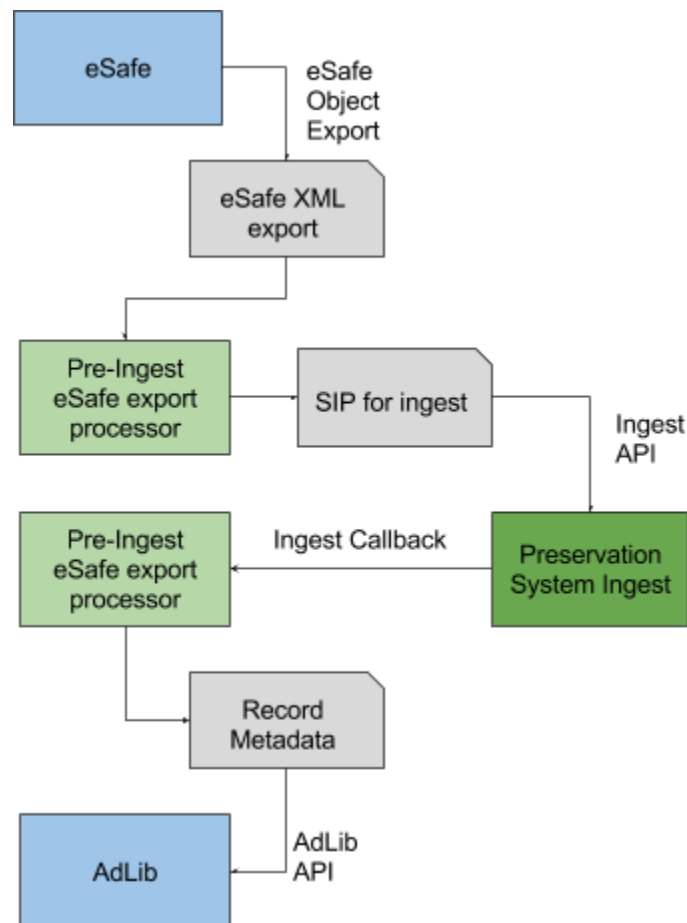


Workflow Overview

Here we give an overview of the proposed preservation workflow for records from the eSafe

system. The aim is to ingest the record content from eSafe for long term preservation while:

- Adding a “stub” to eSafe to inform users accessing the content of the new location.
- Transforming the eSafe record structure, metadata and content into Submission Information Packages for ingest into the preservation system.
- Extracting selected metadata to enable discovery and access of the preserved content via AdLib.



The proposed ingest workflow for eSafe is shown in the diagram above. Here we briefly describe each step in the process.

1. **eSafe export**

Staff select the eSafe content for preservation and use the eSafe object exporter to create an XML file containing the record content and metadata.

2. **Pre-Ingest eSafe export processing**

A purpose developed pre-ingest processor parses the eSafe export file and creates Submission Information Packages for ingest into the preservation system as well as extracting selected metadata for the AdLib record.

3. **Ingest via API and callback**

The pre-ingest component submits the created SIPs for ingest and collects the returned ingest information including the generated IDs for the ingested content.

4. **Adding access metadata to AdLib**

The extracted metadata and generated IDs are used to create AdLib records for the ingested content using the AdLib API. These allow users to search for material using the selected metadata fields and retrieve a link to the content in the preservation system.