how to preserve digital publications

The Deposit System for Electronic Publications

A Process Model

Titia van der Werf
Koninklijke Bibliotheek

nedlib Report Series 6

DSEP: A Process Model

The Deposit System for Electronic Publications:

A Process Model

Titia van der Werf
Koninklijke Bibliotheek

nedlib Report Series 6

Production Management Bureau D'Arts, Amsterdam
Design Herlof Schürmann, Amsterdam
Printed by Mart.Spruijt, Amsterdam
NEDLIB Report Series Editor Titia van der Werf
Available from Koninklijke Bibliotheek
p.o. Box 90407
2509 lk The Hague
e-mail info@kb.nl
Title The Deposit System for Electronic Publications: A Process Model
ISBN 906259150-7
Author Titia van der Werf/Koninklijke Bibliotheek
Date November 2000

This study is jointly funded by the European Commission's Telematics for Libraries Programme.

This is report 6 of the NEDLIB Report Series.

content

Foreword

This report is the result of a joint effort by all NEDLIB partners[1] to understand the requirements and to define the processes for handling electronic publications in a digital deposit environment. It builds on the concepts presented in the Reference Model for an Open Archival Information System (OAIS).[2]
Titia van der Werf from the Koninklijke Bibliotheek in the Netherlands has written the report.
Hilde Høgås, Fred Moerman and Knut Tore Breivik from the National Library of Norway (Mo-i-Rana) have contributed to the description of the access part of the DSEP system.
All NEDLIB participants, including Johan Steenbakkers, Lex Sijtsma, Juha Hakala, Hans Liegmann, Jörg Berkemeyer, Eric Burger, José Borbinha, Fernando Cardoso, Genevieve Clavel, Elisabeth Freyre, Catherine Lupovici and Giovanni Bergamin have contributed to this report and reviewed it at different drafting stages.

This report is jointly funded by the NEDLIB partner institutions and the European Commission's Telematics for Libraries Programme.

1 NEDLIB stands for 'Networked European Deposit Library'. For more details about the project aims, duration, partners and results please refer to:
URL: http://www.kb.nl/nedlib/

2 Referencing Model for an Open Archival Information System (OAIS), Don Sawyer / NASA and Lou Reich / CSC.
In the course of project NEDLIB the following subsequent versions of the OAIS Reference Model have appeared and been used as a basis for this NEDLIB report: White Book, Issue 4, September 1998; White Book, Issue 5, April 1999; Red Book, Issue 1, May 1999
URL: http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

Summary

National deposit libraries are responsible for the long-term preservation and availability of publications and other documentary heritage. As they extend their deposit tasks to include digital material, national libraries need to establish deposit systems for electronic publications (DSEP).
Project NEDLIB, which stands for Networked European Deposit Library, aims to develop a common architectural framework and basic tools for *building* such DSEP *systems*. The project looks into the functional and technical aspects of a DSEP.

In this report a *common workflow* for handling deposited electronic publications is defined on the basis of existing practices and future requirements reported by the NEDLIB libraries. All the basic processes for handling digital deposit collections, i.e. acquisition, cataloguing, storage, long-term preservation and access are mapped onto the broader framework of the Digital Library. The processes belonging to established library practices, such as acquisition, cataloguing and access, are supported by existing systems and are considered to be part of the broader Digital Library System (DLS). The processes relating to digital storage and long-term preservation are not catered to by existing DLS configurations, in any satisfactory manner. As a result NEDLIB has decided to scope the DSEP specifically to support these processes. In addition it has adopted the Open Archival Information System Reference Model (OAIS-RM), as basis for the 'high-level' design of the DSEP.

The *high-level design* of the DSEP specifies a logical model and related functionality for archiving electronic publications, regardless of origin, medium or format. In this sense the model is 'generic'. Similarly, the model supports generic access handling regardless of changing requirements for access rights management and end user service delivery. The DSEP framework is designed to address 'today's access' and to ensure 'future proofing'. It enables consistent implementation without imposing any implementation design constraints.
The DSEP model is designed to be general enough to address common requirements of national deposit libraries, yet specific enough to serve as a basis for local implementations by individual libraries.

As part of the modelling work a list of terms and definitions[3] and a guide to technical standards and solutions for implementing a digital deposit system[4] have been compiled.

3 NEDLIB List of Terms – NEDLIB REPORT 5 – by Genevieve Clavel-Merrin, Den Haag 2000.
URL: http://www.kb.nl/nedlib/results/NEDLIBterms.html

4 Standards for a DSEP: standards for the Implementation of a Deposit System for Electronic Publications (DSEP) – NEDLIB REPORT 4 – by Bendert Feenstra, Den Haag 2000. – URL: http://www.kb.nl/nedlib/results/dsepstandards.rtf

Objective

The main objective of this report is to present the results of the ongoing NEDLIB effort to arrive at a 'generic' high-level design of a Deposit System for Electronic Publications (DSEP) that can serve as a basis for local implementations by individual deposit libraries.

1 Deposit library requirements for handling electronic publications

This section outlines the workflow required for handling electronic publications belonging to deposit library collections. The workflow is based on existing practices and future requirements as reported by the NEDLIB libraries.[5,6,7] The workflow has subsequently been *fine-tuned* during a longer process of consensus building within the NEDLIB consortium.
The workflow consists of eleven steps. For an overview of all the steps please refer to Annex 1: Deposit library workflow.

1 selection for collection building: This process is mainly human-driven and involves the decision making process for including or excluding electronic material from the deposit collection. The decision-making process is based on national deposit policies, regulations and agreements made with publishers and other content providers. The selection process is therefore highly dependent ofon local conditions.

2 acquisition: this process involves all administrative transactions needed to receive new electronic publications in deposit – including receipt of new title information (metadata) from publisher's, ordering and sending recall.reminders. This process requires an exchange of bibliographic and administrative information between deposit library and publisher. In general, publications are acquired free of charge under deposit regulations or against agreed prices.

3 delivery/capture/harvest: the actual deposit process involves actions required to get a copy of an electronic publication from the publisher's production/distribution system to the library's deposit system, according to agreed procedures. The procedures may include delivery via electronic transfer or by conventional shipping procedures or capture by means of web harvesting techniques.

4 registration: this is the process by which the new incoming electronic publication is checked into the deposit system. It may involve actions such as registration, acceptance/rejection and notification of receipt.

5 verification: these are required control routines for checking the source of the data transfer, the physical integrity of the medium, the file formats and the logical integrity of the document. These procedures may involve installation and de-installation. They may result in returning the electronic publication to the sender with error-messages and a notification to acquisition.

6 description: this is the stage in the publication handling process where an entry is created in the library catalogue, to ensure the publication can be found in the library search systems. *Cataloguing* is done according to national cataloguing rules. This process may involve (automatic) re-use of primary metadata supplied by the publisher/author, but it may also involve manual value-adding by the librarian, such as performing authority control on author names, adding annotations and subject description. The bibliographic records are integrated in the library's online catalogue (OPAC) and published in the National Bibliography (NB). The National Bibliography contains the authoritative bibliographic descriptions of publications included in the deposit collection. *Content-indexing* is the (automatic) process that creates full-text indexes and other finding aids for the electronic publication. *Subject-indexing* is the intellectual process whereby subject specialists classify publications according to a subject scheme and qualify them with the help of controlled vocabularies. The description process may require installation of the electronic publication.

7 storage-handling: this is the process that takes care of the storage of the electronic publication in the deposit system. It involves medium migration of the electronic publication from its former carrier to

5 Functional Specification for DSEP – NEDLIB project document – by José Luis Borbinha, Fernando Cardoso, version 15-10-1998
URL: http://www.kb.nl/nedlib/results/func-spec/doc0000.html

6 High Level Design – NEDLIB project document – by José Luis Borbinha, Fernando Cardoso, version 25-09-1998
URL: http://www.kb.nl/coop/nedlib/high-level-design/doc0000.html

7 NEDLIB Local Situations – NEDLIB project document – by José Luis Borbinha, Fernando Cardoso, version of July 2000
URL: http://www.kb.nl/nedlib/results/local_situations_v2.htm

the physical storage of the deposit system, regular refreshing and duplication for backup purposes. It also involves actions such as regular integrity checks and quality control monitoring. Part of this process also consists of the storage location handling, ensuring the management of the physical location of all files in the storage system.

8 preservation: this process consists of all actions required for the long-term preservation of the deposit collection. This may include specification of document formats, emulation of document behaviour, regular medium refreshing and document re-formatting, other transformations to be performed on (parts of) the electronic publication, preservation metadata updates, integrity checks, authenticity assessments and quality assurance procedures.

9 delivery: this is the process that makes the deposit copy of an electronic publication available in such a way that it is fit for use by library visitors. The process allows for the retrieval of the electronic publication from the deposit store, by making a copy of the item and transferring it to the access stage. It may entail extracting parts of the electronic publication, or adding a full-text index to it, or re-formatting of (parts) of the publication for viewing, printing or downloading. It may involve providing for a viewing configuration, etc. Delivery may be embedded in services such as document delivery or print-on-demand.

10 access: access is a whole set of facilities that are part of the library end user environment and support access to the library collections. It includes availability of finding aids, user identification, authentication and authorisation, user-rights management, user-profiles, etc. Access conditions to the deposit collections vary with each library because of different deposit regimes and agreements with publishers. It is also anticipated that access conditions will change over time. The access environment supports the access process initiated by the library user.

11 monitoring: the whole workflow of handling electronic publications, as defined by steps 1 to 10, is monitored for quality control.

2 Positioning the deposit system in the digital library

Many of the workflow steps identified in the previous section are not specific to the handling of digital deposit collections but belong to the broader scope of the Digital Library System (DLS). The workflow needs to be mapped to the digital library environment as a whole. To this end we define broad categories of digital library processes and map the workflow required for handling electronic publications onto these categories. We identify which processes are within the scope of a Deposit System for Electronic Publications (DSEP), and which belong to the Digital Library System (DLS) as a whole. As a result of this mapping activity the scope of project NEDLIB is clarified as well.

The Digital Library processes can be defined by the following broad categories:

Production The library production process covers activities such as digital publishing and digitising library collections. Many current library resources are spent on large-scale digitisation programmes in order to broaden access to library collections and make them available via the web. NEDLIB does not address digitisation. The DSEP however may cater to digitised collections as well as to born-digital publications.

Collection building The library selects digital material according to collection profiles. The deposit collection is based on legal regulation and/or voluntary agreement with publishers. In addition, most libraries define selection criteria to be able to delimit their deposit collection and set priorities to collection building. Many current efforts of deposit libraries are geared towards extending their deposit to digital material and revisiting legal regulations. All electronic publications from publishers are deposited for last-resort access. In some cases, the publications may also be made directly accessible under agreed license conditions, during the commercial lifetime of the publications.
NEDLIB does not address these issues. Workflow step 1 '*selection for collection building*' is not included in the DSEP process model.

Acquisition Selected material may be acquired and owned by the library. However, not all library collections are acquired physically. Increasingly, digital publications remain under the custody of publishers or third party providers, and libraries facilitate availability and access through licensing agreements and access negotiation with content providers (virtual collections).
As a rule, deposit collections are physical collections that fall under a country's deposit regime and archiving the deposit collection for last-resort access is the responsibility of the deposit library.
Acquiring deposit material and agreeing on deposit procedures with individual publishers has taken a new dimension in the digital environment, as electronic publications can now be deposited electronically. Most development work at deposit libraries presently concentrates on this process. It requires much tailoring, as some publishers provide table of contents and others don't, some provide full-text versions for indexing and others don't – it often leads to re-visiting the deposit procedure with publishers and upgrading the quality of deposit submissions. For publishers, this interaction helps them to redesign their publishing process according to higher quality standards
Workflow step 2 *'Acquisition'* and step 3 *'Delivery/capture/harvest'* are part of this process.
Step 2 is the administrative part of acquisition. It is a general library function that applies to paper-based collections as well as to digital collections. It is not included in the DSEP process model.
Step 3 is the physical acquisition part, whereby an electronic publication is physically deposited. It is included in the DSEP process model.

Cataloguing Electronic publications belonging to the library collections are usually catalogued by the library and the resulting bibliographic descriptions are made available through the library's online searching systems such as the Online Public Access Catalogue (OPAC). Some virtual collections may not be catalogued, but as a rule all physical collections are. In particular, electronic publications that belong to the library's deposit collection are catalogued as part of the Universal Bibliographic Control (UBC) effort of national libraries world wide. The descriptions are published in the National Bibliography. Subject indexing and classification also take place at this level of activity. Much has been done to revise cataloguing standards and rules and to accommodate new requirements for describing electronic publications. NEDLIB does not address these issues. Workflow step 6 *'Description'* is not included in the DSEP process model.

Digital storage Electronic publications acquired by the library within the framework of physical collection building require a digital store and an IT-based processing environment for registration, management, preservation and retrieval.
Deposit libraries are increasingly confronted with the need to build digital stores for their growing digital collections. NEDLIB addresses this need in particular.
Workflow step 4 *'Registration'*, step 5 *'Verification'*, step 7 *'Storage-handling'* and step 8 *'Preservation'* are all part of the digital storage process and belong to the DSEP process model. The DSEP process model is mainly positioned in this process category.

Packaging library products The library collections are enhanced and packaged into end products and services (National Bibliography, OPAC, Internet subject guides, document delivery, printing-on-demand, etc) for library customers. An important part of this process category includes product labelling, marketing, providing user guides and help information, etc. New products and services are defined at this level, in particular the way in which the deposit collection is made available to the library user. This may lead to revisiting the role of the National Bibliography in the digital environment, for example. Workflow step 9 *'Delivery'* maps onto this process category. The processing needed to make electronic publications fit for use is included in the DSEP process model. The model however does not include marketing, business models and product development related to deposit collections.

Service Delivery All electronic information products and services provided by the library are made available through the library network, in particular via the library web-site. The library web-site is the shopping arcade of the Digital Library. It presents a window to the library products and services and it functions as the digital library counter for service delivery. This environment interfaces with the end user environment, at the library premises, at home or at work. Access control and user authorisation issues are being addressed at this level. NEDLIB does not address these issues.
Workflow step 10 *'Access'* maps onto this main process category and is not included in the DSEP process model.

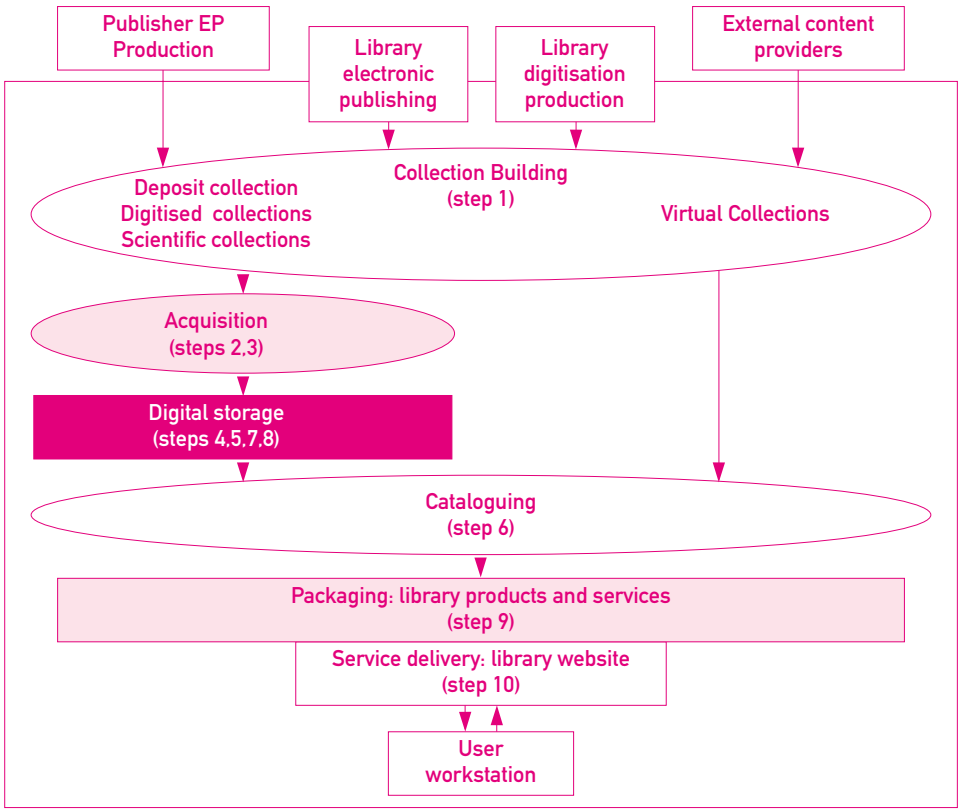


Figure 1 The Digital Library Model

The workflow steps for handling electronic publications of deposit collections, described in the previous section, are mapped onto the Digital Library model in Figure 1. The processes within scope of the DSEP belong to the 'Digital Storage' process category [Figure 1 shaded dark red box] and partly to 'Acquisition' and 'Packaging' [Figure 1 shaded light red boxes].

The last workflow step, step 11 *'Monitoring',* is a process affecting the Digital Library as a whole. However each process requires to be monitored individually as well. Monitoring and quality control is therefore also part of the DSEP process model.

3 Applying the Open Archival Information System Reference Model

This section discusses the Open Archival Information System Reference Model (OAIS-RM), a proposed ISO archiving standard.[8] It describes how NEDLIB has applied, scoped and extended this Reference Model to the needs of the Deposit System for Electronic Publications (DSEP).

Applicability of the OAIS Model

The OAIS Model is applicable to any archive. It is specifically applicable to organisations with a responsibility to make information available for the long term.
The NEDLIB consortium adopted this Reference Model as a basis for modelling the DSEP. The decision to adopt OAIS was taken at the NEDLIB Paris meeting in December 1998. The OAIS Reference Model (issue 4.0, dated September 1998) had developed into a mature conceptual framework, providing a coherent and consistent view of functions and data flows pertaining to digital archives. At that point in time, NEDLIB had started to map the workflow for handling electronic publications onto a structure of functional entities[9] and the project decided to extend this effort by mapping the workflow onto the OAIS-RM functional entities. Not surprisingly, it turned out that the functional structure of a deposit system could appropriately be represented by the OAIS model.

Advantages of applying the OAIS Model

From the start it was recognised that by applying the OAIS model, deposit libraries could benefit from the advantages of international standardisation. By using a common reference model, a common terminology and a common conceptual framework, it is much easier to share ideas and exchange experiences. Not only between deposit libraries, but also across institutional boundaries, for example, between libraries and archives.
In the NEDLIB project all work has been related to OAIS: process design was done on the basis of OAIS modelling, tools were described in terms of OAIS functional entities[10] and tested according to an OAIS based scorecard,[11] metadata were specified in the context of the OAIS data model.[12] This has facilitated the consensus building process considerably. This has proven to be true also during concertation meetings with non-NEDLIB deposit libraries and other, related initiatives and projects.
On the longer-term it is hoped that IT-vendors and system developers will adopt the OAIS framework as a basis for implementing deposit systems and for developing ready-to-market products. This would facilitate open systems development for the benefit of a much larger community then would be the case if archival institutions invested in tailor-made systems on an individual basis.
This is also mentioned as being part of the objectives of the OAIS Reference Model, namely:

- To expand consensus on the elements and processes for long-term digital information preservation and access
- To promote a larger market which vendors can support
- To guide the identification and production of OAIS related standards[13]

NEDLIB is keen to help progress the OAIS standardisation process and to provide feed back in order to ensure that generic deposit library requirements are catered to by the Reference Model.

8 op.cit. 2

9 op.cit. 6

10 Specification of Tools – NEDLIB project document – by Juha Hakala, version of October 2000

11 Scorecard for Functional Testing of Tools – NEDLIB project document – by Lex Sijtsma, December 1999.

12 Metadata for long-term preservation – NEDLIB REPORT 2 – by Catherine Lupovici et Julien Masanès, Den Haag 2000. URL: http://www.kb.nl/nedlib/results/preservationmetadata.pdf

13 op.cit. 2. See section 1.1 Purpose and scope

Description of the OAIS Model

The Reference Model addresses a full range of archival functions including Ingest, Archival Storage, Data Management, Access, and Administration (Figure 2).

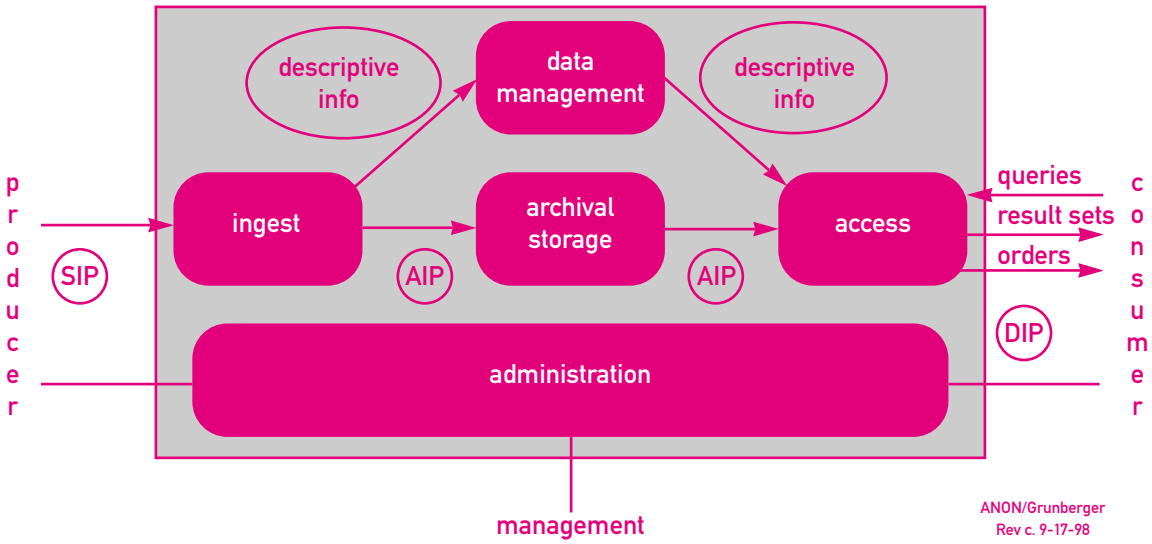


Figure 2 OAIS Functional Entities
[OAIS, Figure 4.1]

It also addresses the data models used to represent digital information in archives from a preservation perspective. The OAIS defines an Information Object as Data Object interpreted using its Representation Information. This is shown schematically in Figure 3.

Figure 3 Obtaining Information from Data
[OAIS, Figure 2.2]

In order for this Information Object to be successfully preserved, it is critical for an OAIS to clearly identify and understand the Data Object (the bits) and its associated Representation Information (implicitly hidden in the interpreting/rendering software).

Information transmissions internal and external to the OAIS archive functions occur by way of information packages. The information package contains the information object that needs to be preserved for future access. Three different types of Information Packages are defined: the Submission Information Package (SIP), the Archival Information Package (AIP) and the Dissemination Information Package (DIP). This is shown in Figure 4.

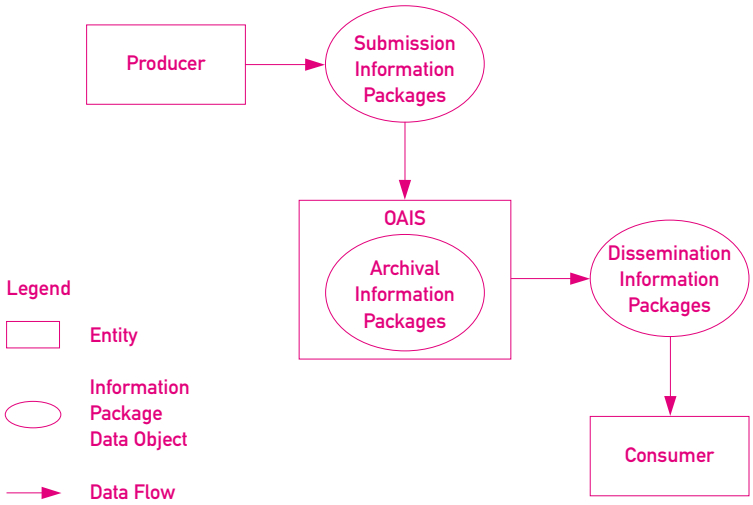


Figure 4 OAIS Archive External Data Flows
[Adapted from OAIS, Figure 2-4]

An Information Package (IP) is a conceptual container of two types of information called Content Information and Preservation Description Information (PDI). The Content Information and PDI are viewed as being encapsulated and identifiable by the Packaging Information. The resulting package is viewed as being discoverable by virtue of the Descriptive Information.
These Information Package relationships are shown schematically in Figure 5.

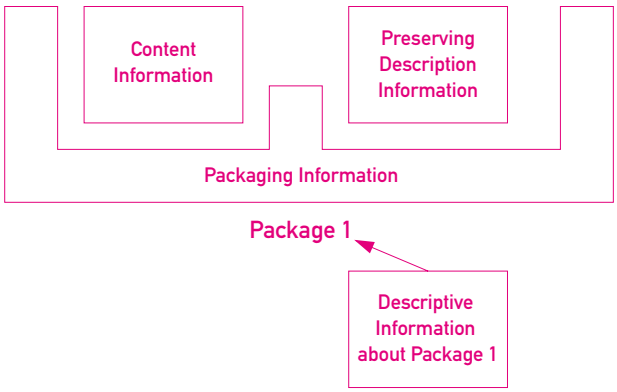


Figure 5 Information Package Concepts and Relationships
[OAIS, Figure 2-3]

Preservation perspectives

The concept of Information Object, with the explicit distinction between the Data Object (the actual bit-stream) and the Representation Information (information that enables decoding of the bit-stream into meaningful information) is central to the OAIS model. Preserving both the bit-stream and its Representation Information through time are crucial requirements.
The OAIS document provides some perspectives on the issues of information preservation using digital migration across media and across new formats or representations, but it is not clear which processes are needed and which functionality is required. It discusses medium migration (refreshing or copying a publication onto a new carrier) as a preservation procedure belonging to Archival Storage. As formats become obsolete and the viewers needed to interpret and render these formats become obsolete as well, measures to preserve the content of a publication and all related aspects such as look and feel, layout, structure and functionality, need to be taken. To this end, several strategies may be followed, such as

migration and emulation. The OAIS model does not discuss different preservation strategies and how they affect the model. It does incorporate data migration, i.e. 'transformation' of digital content. In all cases, transformation leads to a 'new version' of the original publication. It is not clear however where transformation takes place in OAIS. It does not belong to Archival Storage and this is understandable because Archival Storage does not have (and does not need to have) any knowledge of the content of a publication. The Administration entity has an 'Archival Information Update' function that provides a mechanism for updating the contents of an AIP stored in Archival Storage, by accessing it as a DIP, updating its content and resubmitting it as a SIP to Ingest. However the Reference Model does not clarify if and in what way this function belongs to a preservation process.

Scoping OAIS to the DSEP

The OAIS Reference Model defines the environment surrounding the archive and the interactions with Producers, Consumers and Management. Similarly, the DSEP Model defines the Digital Library environment surrounding the DSEP and the interactions between the DSEP and the Digital Library System (DLS).
NEDLIB has established that much of the OAIS functionality maps onto the broader digital library configuration. Many OAIS functions map to equivalent DLS functions, such as:

- negotiating submission agreements *(Acquisition)*,
- description (*Cataloguing*),
- creating finding aids (*Packaging*),
- user authorisation, managing access controls, access provision to customers (*Service Delivery*)

These are all general DLS functions.
In other words, OAIS functionality bundled in entities such as Ingest, Data Management and Access overlaps the general functionality of a digital library system (DLS). Consequently, OAIS functionality is situated partly outside and partly inside the actual limits of a DSEP.
In the model developed by NEDLIB, the Ingest, Data Management and Access functionality belonging to the DSEP system is much more limited than in the OAIS Reference Model. In fact, Ingest, Data-Management and Access have functionality only directly relating to storage & retrieval handling and preservation processing.

The NEDLIB model defines how the DSEP and the DLS interact. Two interfacing processes have been defined through which all input and output interactions with the DSEP take place. The interfacing processes have been *defined* as 'Delivery&Capture' at the input side and 'Packaging&Delivery' at the output side of the DSEP. These processes interact with the DSEP on the basis of well-defined input and output standards and conventions, such as a SIP and a DIP. They take care of the interactions with the publishers and other information providers at the input level, and with the library access systems at the output level. In this way the interfacing processes cater to the specific and changing requirements for interaction with the outside world whilst the inner DSEP processes can be considered to operate in a much more controlled and generic environment. These modelling considerations are expanded upon below.
It is acknowledged that some processes, namely creating finding aids, such as content indexing and cataloguing, need to some extent to interact with the DSEP to perform their tasks, but this interaction need not be direct and can occur through the input/output interfaces.

Figure 6 shows the result of this scoping exercise.

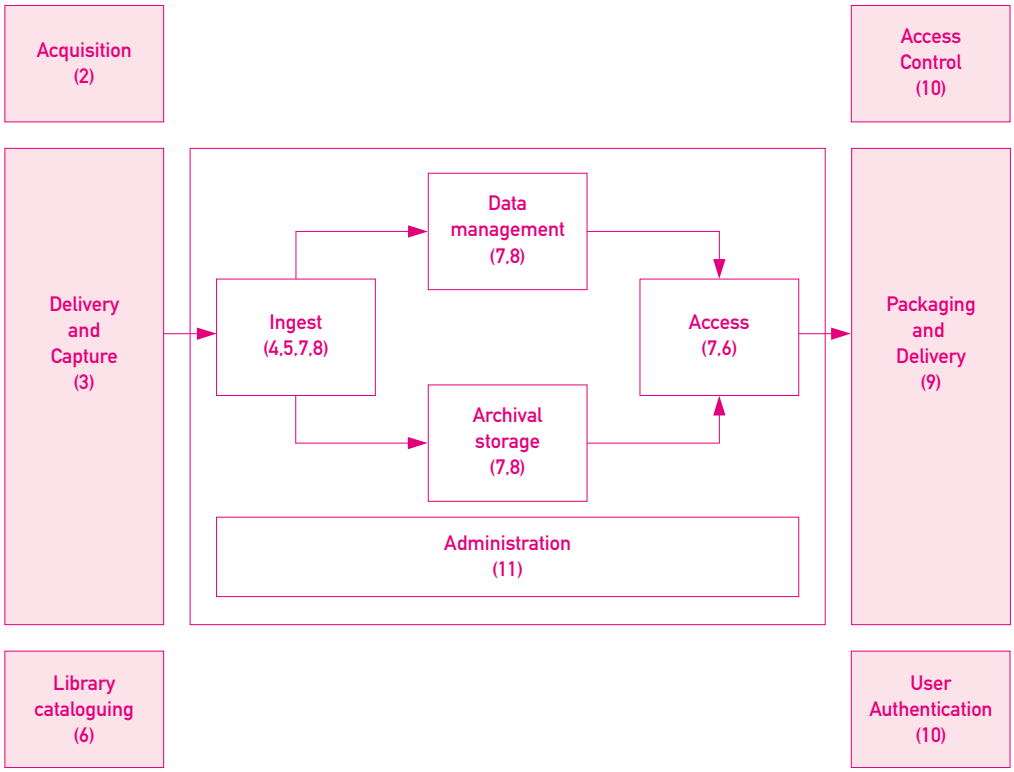


Figure 6 OAIS functional entities scoped to DSEP processes

Extending the OAIS model

NEDLIB has looked into preservation issues in some detail from different perspectives. It has commissioned Mark Bide to provide an overview of e-publishing standards and characteristics of electronic publications with a discussion of related preservation issues.[14] It has commissioned Jeff Rothenberg to initiate experiments to test the emulation of hardware approach with real digital deposit publications and to provide procedural models to support emulation-based preservation in the context of OAIS and DSEP.[15] It has looked in some detail at the metadata for preservation requirements and how they relate to the OAIS Information Model.[16]

What NEDLIB found missing in the OAIS Model was a conceptual entity symbolising the preservation processes required of an OAIS, whatever the preservation strategies followed. Therefore NEDLIB has added in its DSEP model a Preservation entity that manages the preservation processes required of a DSEP. Although it is recognised that the preservation function may affect all DSEP processes, NEDLIB has added this separate preservation entity to make this function more visible and more explicit in the model. Much in the same way as metadata processing affects all DSEP functions, still, OAIS has defined a separate Data-Management entity to visualise the metadata processing function.

Both transformation and emulation approaches are worked out in some detail in the DSEP model. The resulting output is either a new version of a formerly deposited publication, in which case it is ingested anew in the system, or it is a set of specifications for building emulators that can render a whole generation of publications on a future (unknown) reference platform. In both cases, new SIPs are ingested into the DSEP with generation of new preservation metadata.

14 Standards for Electronic Publishing: an overview – NEDLIB REPORT 3 – by Mark Bide, Den Haag 2000.
URL: http://www.kb.nl/nedlib/results/e-publishingstandards.pdf

15 An experiment in using emulation to preserve digital publications – NEDLIB REPORT 1 – by Jeff Rothenberg, Den Haag 2000
URL: http://www.kb.nl/nedlib/results/emulationpreservationreport.pdf

16 op.cit. 12

At the time of writing, the OAIS model is still in its ISO-review stage. As a result of an interim review in the first half of 2000, to which NEDLIB has contributed,[17] the current draft CCSDS recommendation for an OAIS-RM[18] does propose to add a 'Preservation Planning' entity.
This version of the DSEP model is based on the latest recommendation document.

Modelling considerations

This paragraph discusses some design considerations encountered during NEDLIB's work on OAIS and DSEP. It explains why certain less obvious choices have been made. The choices have been discussed with experts in the field, with other national libraries not participating in NEDLIB and with representatives from the OAIS standardisation effort.

Conceptual design vs implementation design

The Reference Model document assumes that implementers will use the OAIS model as a guide while developing a specific archival system implementation. The document stresses that it 'does not assume or endorse any specific computing platform, system environment, system design paradigm, system development methodology, database management system, database design paradigm, data definition language, command language, system interface, user interface, technology, or media required for implementation.'[19]

In his preservation test-bed report for NEDLIB Jeff Rothenberg rightfully stresses how important it is 'to keep in mind that the OAIS is intended as a reference model rather than a system design model. One implication of this (both for the OAIS and for the DSEP specification, which is derived from it) is that the functions or processes shown in these models do not necessarily correspond directly to the functional modules of a system that would implement that model. The functional decomposition of a system into appropriate modules is a design issue, and various implementations may well lend themselves to functional decompositions that are quite different from the "reference processes" of the OAIS.'[20]

In order to apply the OAIS model to the DSEP, the actual logical processes belonging to the DSEP have been identified and mapped to the OAIS functional entities. The DSEP model is a logical process design of the DSEP system. It refers to the OAIS functional entities without assuming that these correspond to an optimal functional decomposition of the desired system. In fact it introduces a new functional entity, the preservation entity, as a conceptual framework for the preservation process. This is not done to suggest that the preservation process is an isolated process that can be implemented as a separate preservation module, but rather to stress the importance of the preservation process as part and parcel of the DSEP. Even though the preservation process may pervade several other DSEP processes it is conceptualised in the DSEP model as a separate entity.

Similarly, NEDLIB has interpreted the OAIS concept of the Archival Information Package (AIP) as a conceptual package rather than an actual data structure. In this way it was possible to disassociate the different component parts of an Information Package (data and metadata) and to consider where various kinds of metadata are needed or generated by the different DSEP processes. The DSEP data flow model therefore indicates which metadata subsets logically 'belong to' which DSEP processes. The AIP, containing the information object, remains the unit to be preserved and retrieved for access in a DSEP, but actual implementations of an AIP are free to decide which metadata subsets are to be stored together with the information object into one data structure for preservation. It is suggested however that only

17 NEDLIB contribution to the ISO-review of the OAIS-RM, June 2000. – URL: http://www.kb.nl/nedlib/results/OAISreviewbyNEDLIB .html

18 CCSDS recommendation for an OAIS-RM, draft version circulated during the International CCSDS Workshop in Oxford, November 2000.

19 op.cit. 2. See section 1.4

20 op.cit. 15 . See section 2.1

those metadata that belong to the original publication and without which the publication is not complete, be preserved together with the information object – as a *bit stream* – in the AIP. An example of this kind of metadata is 'structural metadata' which specify the document structure, such as for example XML document type definition (DTD) or Ebind-DTD files. The metadata required by the different DSEP processes should not be preserved as part of the AIP, but be functionally accessible at all time. They need to be updated on a frequent basis and migrated to new data management software or data structures, as necessary, without having to unpack and modify the preserved publication itself, which remains unchanged in its AIP.

Modularity

Two interfacing processes have been defined through which all input and output interactions with the DSEP take place. This has been done to protect the DSEP from being exposed to changing external requirements that can be less easily controlled. From a design point of view it is best to ensure that the DSEP is a system operating autonomously, independently from external variables. Other digital library processing systems belonging to the DLS should similarly, be able to operate independently and without any knowledge of internal DSEP technical solutions and conventions. The DSEP should operate like a black box to the other systems. In this way changes in one system do not affect other systems. The DLS is conceived as a modular system, in which the cataloguing system, the acquisition system, the DSEP and other library systems operate individually, each with well-defined responsibilities. Because the internal workings of the DSEP are private to itself, the DSEP module can be developed independently of other existing DLS modules.

Generic vs local requirements

In order to minimise dependency, the number of interfaces for interaction with the DSEP have been reduced to a minimum of two. These are basically defined as input and output interfaces. The input interface (Delivery&Capture) serves to prepare a publication for ingest into the DSEP. The output interface (Packaging&Delivery) serves to unpack the retrieved publication from the DSEP and make it for use. The interfacing processes bundle all functionality that is necessary to address local, non-generic particularities of the digital library environment. The process 'Delivery&Capture' defines the non-generic part of 'Ingest': it is tailor-made to accommodate the different local deposit formats and procedures agreed between deposit library and publishers within each country. The process 'Packaging&Delivery' defines the non-generic part of 'Access': it is tailor-made to accommodate the different local access conditions agreed between the deposit library and publishers. In addition the deposit formats and procedures and the access conditions may alter significantly through time, making it necessary to adapt the interfacing processes regularly to changing external conditions. Separating the changing and local features of a DSEP from the more fixed and generic ones, makes it possible to focus on a generic system that is relevant for all deposit libraries and that can be implemented everywhere without the need for drastic localisations.

4 The deposit system process model and data flow description

This section describes the process model for a deposit system for electronic publications (DSEP). It details each sub-process pertaining to the DSEP and the processes interfacing with it.
For an overview of all the processes please refer to the DSEP Process Tree (Annex 2).

The top-level processes

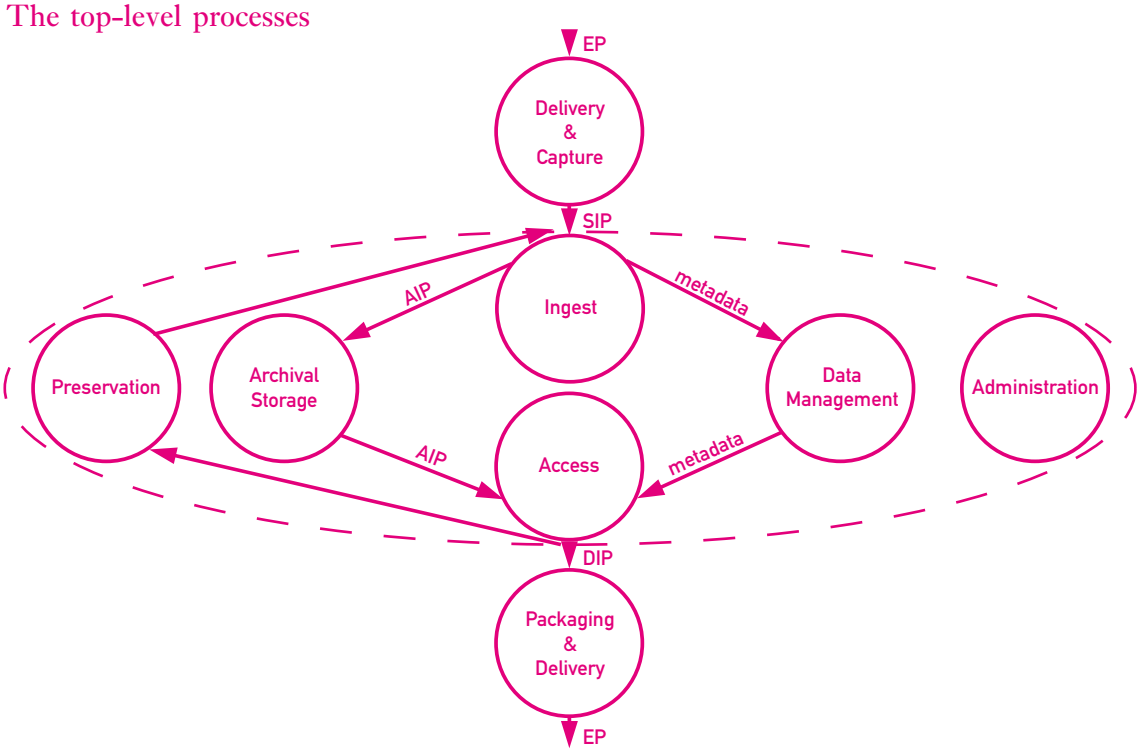


Figure 7 Top-level processes

The DSEP supports the following top-level processes:

1 Ingest
Ingest loads the publication prepared by 'Delivery&Capture' into the deposit storage and the associated metadata into Data Management.
2 Archival Storage
Archival Storage takes care of the storage and retrieval of the electronic publications and of the integrity of the bit-stream.
3 Data Management
Data Management takes care of the storage and retrieval of metadata associated with the publication and with systems administration.
4 Access
Access makes the archived publication and its associated metadata available through the interfacing process 'Packaging&Delivery'.
5 Administration
Administration is responsible for the operation of the system.
6 Preservation
Preservation is responsible for the long-term access and readability of electronic publications.

Interfacing processes for input and output:

The DSEP interacts with the following interfacing processes:

7 Delivery & Capture is the input interface of the DSEP
Delivery&Capture takes care of the pre-processing of electronic publications to be ingested into DSEP.

It receives or captures electronic publications and offers a working space for verification, redistribution of data for processing by external systems (Acquisition and Cataloguing) and repackaging according to the specifications of a SIP for ingest into the DSEP.

8 Packaging&Delivery is the output interface of the DSEP
Packaging&Delivery takes care of the post-processing of electronic publications retrieved from DSEP. It negotiates access requests, delivers and installs electronic publications, together with appropriate viewing or running software and metadata, for direct access on the library visitor's workstation, via the library access systems or via external systems.

Detailed process description

4.1 DELIVERY&CAPTURE

Delivery&Capture is conceptualised as the 'pre-processing' interface to the DSEP. It is needed because deposit libraries cannot dictate submission formats to publishers: in principle, they have to accept all formats published on the market.
Most development work at deposit libraries presently concentrates on this process. It requires much tailoring, as some publishers provide table of contents and others don't, some provide full-text versions for indexing and others don't - it often leads to re-negotiating the deposit procedure with publishers and upgrading the quality of deposit submissions. For publishers, this interaction helps them to redesign their publishing process according to higher quality standards.

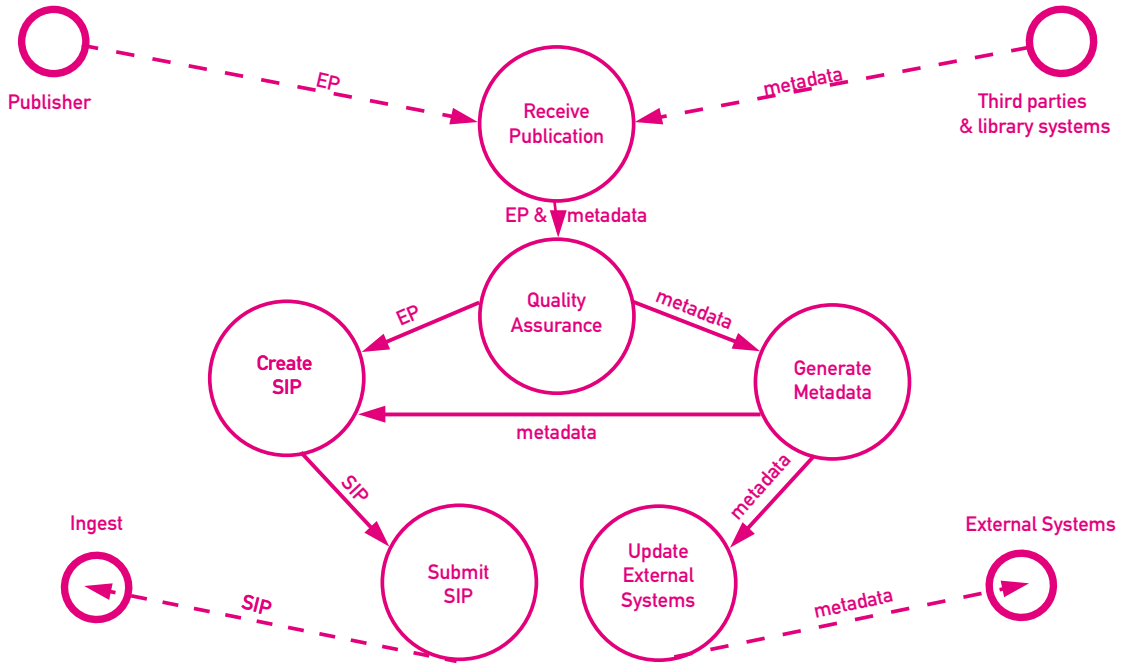


Figure 8 Process Delivery&Capture

The Delivery&Capture process is responsible for receiving deposit publications from the publishers and for submitting them to the DSEP system as a standard submission package.

This process is complementary to the library acquisition process. It manages the electronic deposit transaction, on the basis of agreed procedures with publishers. These procedures may vary with each publisher and each publication type.
Once the publication has been transferred successfully to the *deposit transfer area* within the library system, the publication is (re)packaged according to the specifications of a Submission Information Package (SIP) and submitted to the DSEP.

Data Flow

Input: one or more EP's [publisher]

An electronic publication may consist of any of the following:

- Content data object, provided by the publisher/author
- Descriptive metadata, provided by the publisher (e.g. title information)
- Structural metadata, provided by the publisher (e.g. specifying the entry-point and the navigation structure of a publication, such as the index.html file for a web publication)
- Access metadata, provided by the publisher (e.g. specifying an embargo on the publication)
- Technical metadata, provided by the publisher (e.g. specifying the system requirements for installing the publication)
- Software, provided by the publisher to run or/and view the publication
- Electronic publication definition file, specifying the component parts and logical units of the publication.

Input: External Metadata [third parties; library systems]

- Bibliographic metadata, provided by third parties (e.g. identifiers provided by identification agencies such as the national ISBN/ISSN agencies, CIP-descriptions provided by bibliographic agencies)
- Bibliographic record, provided by the library cataloguing system (e.g. the full bibliographic description made for the National Bibliography)
- Metadata, provided by the library acquisition system (e.g. publisher information).

Output: one or more SIPs [Ingest]

A SIP may contain any of the following:

- the electronic publication
- tagged and standardised metadata
- a full bibliographic description coming from the library cataloguing system
- (a pointer to) software, to run or/and view the publication.

A SIP always contains:

- packaging information
- SIP-identifier

Output: Standardised Metadata [library systems, third parties, publisher]

Metadata collected and standardised during the process Delivery&Capture may be copied to external systems.
The SIP specifications are established by the DSEP Administration Process and are known to the outside world. Only SIPs coming from the Delivery&Capture process can enter the DSEP. Deposit agreements with publishers may encourage and enable publishers to submit SIPs directly, however it is recommended to use Delivery&Capture as a safety valve.
The process Delivery&Capture includes the sub-processes Receive Publication, Quality Assurance, Generate Metadata, Update External Systems, Create SIP and Submit SIP.

Sub-process Receive Publication

Deposit procedures and schedules may vary depending on the agreements made with publishers and on the type of publications (e.g. serial or monographic publications).
Whatever the original distribution medium of an electronic publication, we can distinguish between offline and online deposits. In some instances the library will fetch the publication, in others the publisher will deliver it. In general, publications can be deposited in several ways:

- online delivery by the publisher via file transfer protocol (FTP) or email or any other online transfer protocol,
- online capture by the library via web transfer protocol (HTTP), file transfer protocol (FTP), via scheduled harvesting (web-harvest) or mirroring (site-mirror) or any other agreed means,
- offline deposit via postal delivery on a portable medium (diskette, CD-ROM, DVD, etc.). This includes deposit of offline publications, such as CD-ROM publications, and offline batch deposits of online publications.

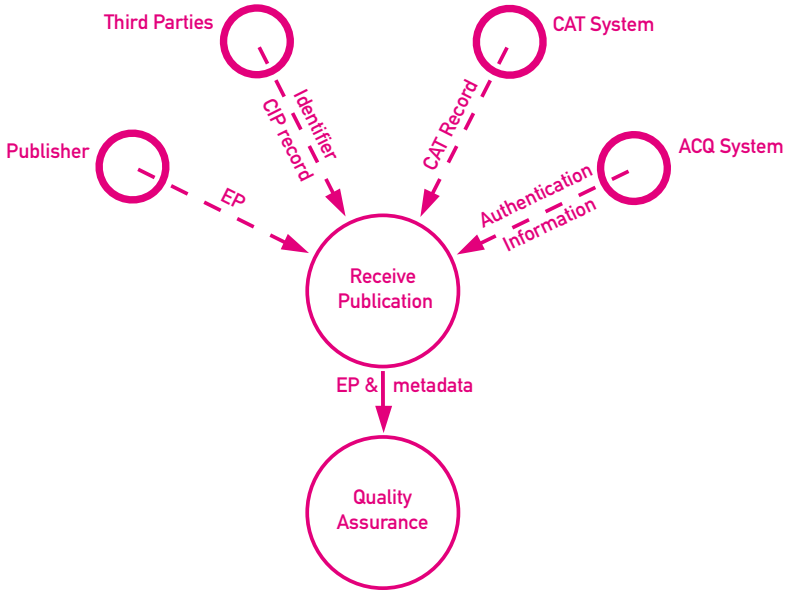


Figure 9 Process Receive Publication

The sub-process Receive Publication is responsible for receiving or fetching deposit publications from the publishers and transferring them to the deposit transfer area.
The sub-process Receive Publication activates deposit procedures from the library side. It fetches publications from agreed transit areas (e-mail post-boxes, ftp-directories, etc.) outside or inside the library's domain and copies them on to the deposit transfer area. Whatever the medium/protocols/procedures by which a publication is received or fetched, the deposit copy is always copied onto the *deposit transfer area*. This is done by way of secure transactions.
Control routines for checking data transfer, file integrity and for authenticating the publisher are performed. Automated filters may collect relevant metadata, such as the source of a publication received by e-mail. All metadata collected or generated during the transfer, including file integrity checksums and authentication data, are recorded and added to the deposit item as accompanying metadata.

When the transfer is successful, the sub-process Receive Publication notifies the library acquisition system and the deposit library employee of a new deposit transfer.
In some cases, for example when the publisher has not been identified, the deposit library employee needs to assess if the publisher and the publication meet the deposit conditions and selection criteria.

The *deposit transfer area* is a temporary storage space for clearing incoming deposit publications. It is configured as a managed workspace with separate areas (e.g. directories) for clearing different types of deposits. Only authorised (library) user-groups have access to it. It can be accessed from dedicated workstations by pre-defined deposit employee user-groups with special permissions (such as read, write, delete, create) to prepare the deposits for ingest. Very well defined deposit procedures and standards may allow for highly automated pre-processing.

Data Flow
Input: EP
Input: External Metadata

Output: EP + metadata generated or collected by the process Receive Publication [Quality Assurance]
- file integrity checksums
- authentication metadata (source/publisher)

Sub-process Quality Assurance

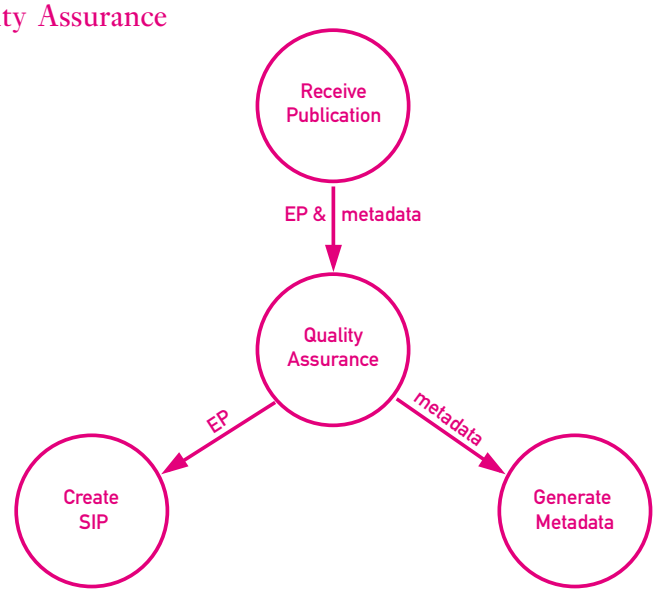


Figure 10 Process Quality Assurance

During the Quality Assurance sub-process the logical integrity of the deposit items is checked and the component parts are identified. This leads Quality Assurance to accept or reject the publication for deposit.

The deposited items are screened and validated. Virus checks are performed. The publication is checked on completeness: are all files present? The logical integrity is checked against the electronic publication definition file agreed with the publisher. This could include for example, the table of contents listing all the component parts of the publication, such as the list of all the articles of a journal issue.
The Quality Assurance sub-process parses and extracts the metadata accompanying the publication. It identifies the publication contents, the bibliographic metadata provided by the publisher, special access controls placed on the contents, full-text indexes and other additional data accompanying the publication such as system requirements. The different parts of the publication and the different types of metadata are tagged. The metadata are of differing quality levels and not always highly standardised. During the following sub-processes the metadata are enriched and standardised.

Once Quality Assurance has accepted the publication for deposit, a deposit stamp can be generated. It can be used to mark the publication as the digital original received from the publisher under the deposit regime. It can also be used to indicate that the publisher cannot be held responsible for changes occurred to the deposit copy during the preservation process.

Data Flow

Input: EP + generated or collected metadata [Receive Publication]
Output: EP [Create SIP]
Output: tagged metadata [Generate Metadata]
- deposit stamp
- tagged metadata

Sub-processes Generate Metadata

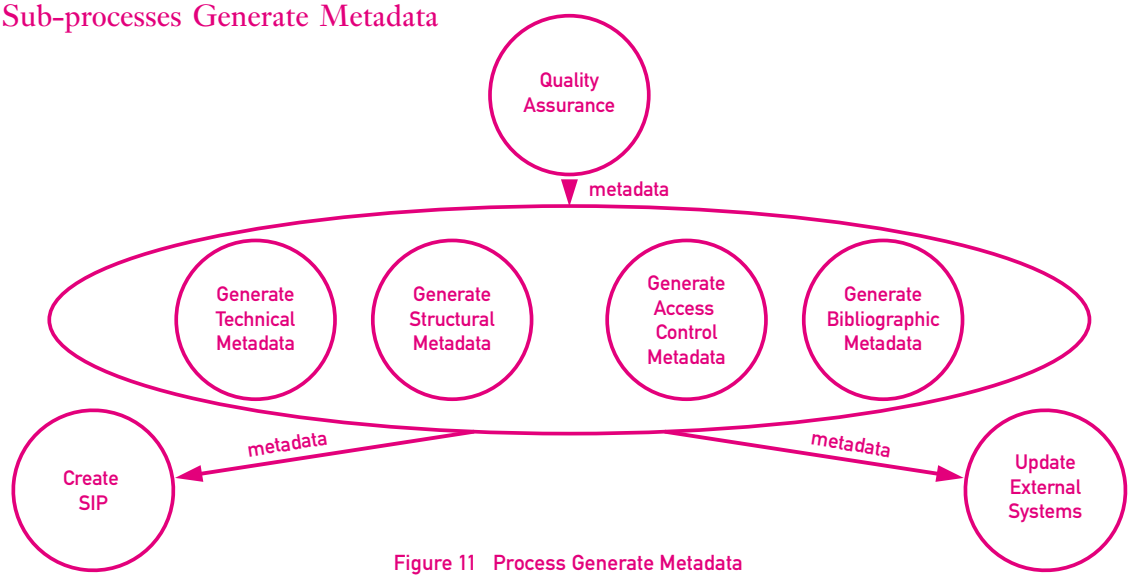


Figure 11 Process Generate Metadata

If Quality Assurance is completed successfully, the following different 'Generate Metadata' sub-processes are activated. In case mandatory accompanying metadata, such as technical metadata, are not provided by the publisher, the deposit library employee needs to enter this metadata him/herself.

Sub-process Generate Technical Metadata

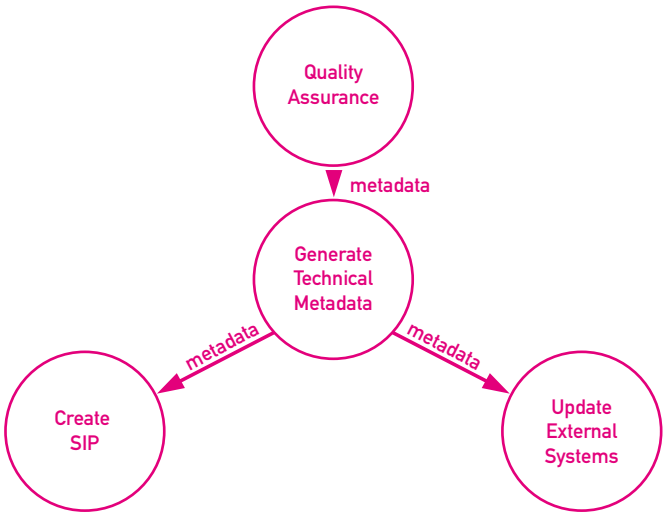


Figure 12 Process Generate Technical Metadata

The Generate technical metadata sub-process records all relevant technical metadata and verifies the installation requirements. The sub-process assigns a file type to each file on the basis of a controlled list of file types and versions. Dependencies on software required for viewing the publication are defined. A pointer to the *Reference Platform* is generated on the basis of the defined system requirements.
In some cases, such as in the case of CD-ROM publications, the publication is installed to ensure that it runs properly. Installation is the execution of the set-up procedures. The installation steps may be recorded as a script for future automatic installation. An image of the installed publication may be made and kept in a separate SIP.
The *Reference Platform* is a standard computer system configuration designated by the deposit library for installing and running electronic publications. The Reference Platform is suited for most publications appearing on the consumer market, during a given period of time.

Data Flow

Input: tagged metadata [Quality Assurance]

- system requirements

Output: standardised technical metadata [Create SIP, Update External Systems]

Technical metadata can consist of any of the following:

- file formats and format versions
- software dependencies
- installation metadata
- pointer to appropriate Reference Platform

Sub-process Generate Structural Metadata

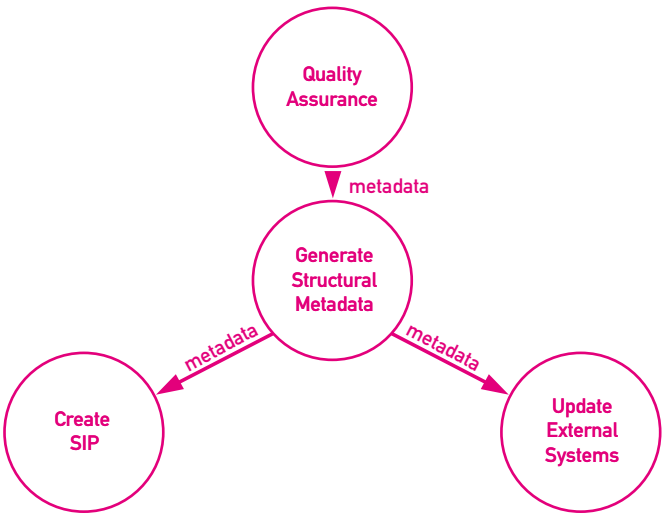


Figure 13 Process Generate Structural Metadata

The Generate structural metadata sub-process records all structural metadata necessary for presenting and navigating the publication at access time. This is particularly important for hypertext publications. For this type of publications it is necessary to know what the entry-point of the publication is and what its browsing structure is. By contrast, executable publications do not require separate structural metadata.

Data Flow

Input: tagged metadata [Quality Assurance]

- structural metadata

Output: standardised structural metadata [Create SIP, Update External Systems]

- entry-point file
- browsing structure file

Sub-process Generate Access Control Metadata

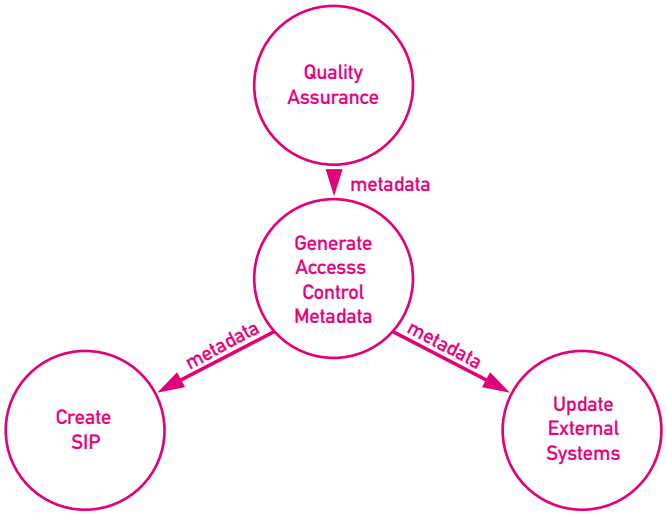


Figure 14 Process Generate Access Control Metadata

The Generate access control metadata sub-process records access metadata accompanying the publication, including for example access controls (embargo) placed on the publication. This metadata may be re-used at access time.

Data Flow

Input: tagged metadata [Quality Assurance]

- access control metadata

Output: standardised access control metadata [Create SIP, Update External Systems]

- encoded access control metadata

Sub-process Generate Bibliographic Metadata

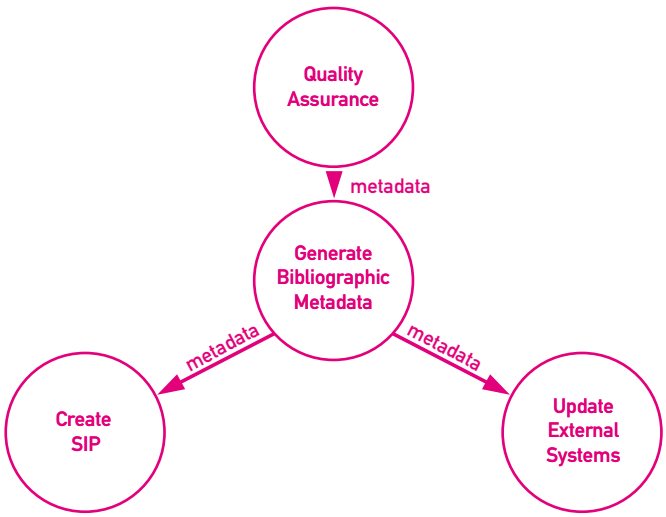


Figure 15 Process Generate Bibliographic Metadata

Descriptive metadata provided by the publisher and bibliographic metadata provided by third parties (e.g. identifiers provided by identification agencies such as the national ISBN/ISSN agencies, CIP-descriptions provided by bibliographic agencies) is reformatted into the library's bibliographic record format and copied to the library cataloguing system.

The unique identifier of the logical publication unit to be stored in DSEP and described in the catalogue is allocated.

Data Flow
Input: tagged metadata [Quality Assurance]
- title, author(s), date of publication, version, identifier, publisher, place of publication, …

Output: standardised descriptive metadata [Create SIP, Update External Systems]
- title, author(s), date of publication, version, identifiers including NBN, publisher, place of publication, keywords, …
- unique identifier of the logical publication unit

Related aspects

Granularity

The disassociation of content and physical carrier in the digital world has created definition problems regarding the unit of publication and levels of granularity. The library employee needs to be able to define the logical publication unit that will be the basis for description and retrieval through the library search systems. The granularity of publications differs with each type of publication (e-journal, web-publication, database-like publication, etc.) and has to be established according to well-documented rules. For example, the logical publication unit of a journal can be defined at the journal title level (open-ended unit because the journal usually continues to appear for years to come), at the issue level (bundling several articles together) or at the article level.
Whatever unit is chosen as the basis for description, this unit should be uniquely identifiable across the library's systems, in particular the catalogue and DSEP systems.

Identification and Linkage

Traditionally, the unique identifier of the publication is provided by the publisher or by an identification agency. For electronic publications this is however not (yet) a well-established practice, mainly due to the granularity issues raised above.
At the deposit stage the National Bibliographic Number (NBN) associated with the deposit copy of the publication is assigned. This identifier can be used to link the logical publication unit stored in DSEP with its bibliographic description in the catalogue.
Whichever identifier is used, unique identification of the logical publication unit should take place. The resulting identifier should in all cases be associated with the publication stored in DSEP and its description in the catalogue.

Bibliographic Record

When a full bibliographic description has been made (sometimes requiring installation of the publication to verify and enrich the descriptive metadata) this authoritative record may also be copied to the DSEP system to be kept along with the publication it describes. There are some practical workflow issues here because installation and description are best done once the publication has been ingested and stored successfully into DSEP. Incorporating the full bibliographic description to the AIP afterwards may prove cumbersome. Another issue is of course the synchronisation problem: how to update the same bibliographic record in different databases.

Sub-process Update External Systems

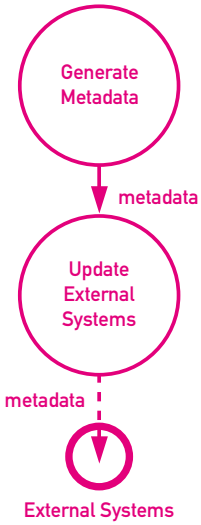


Figure 16 Process Update External Systems

The sub-process Update External Systems is responsible for updating external digital library systems with metadata collected, generated and standardised during the previous Generate Metadata processes. It copies, for example, descriptive metadata and identifier information to the library acquisition and cataloguing systems.

Data Flow
Input: standardised metadata [Generate Metadata]
Output: standardised metadata [External systems]

Sub-process Create SIP

Several publications from a given publisher may be packaged together into one SIP or on the contrary split into different SIPs. For example, an offline batch deposit of online journal issues may be packaged in SIP units at the issue level (bundling several articles together) or at the article level. If on the contrary the journal is seen as the unit, the deposit transfer is open ended and pragmatic choices should be made to package the incoming issues in subsequent SIPs. In that case more than one SIP constitute a logical publication unit.
In general the SIP is a packaging unit and there need not be a one-to-one relationship between SIPs and logical publication units. One or more SIPs can correspond to a logical publication unit. Several logical publication units can also be packaged into one SIP.

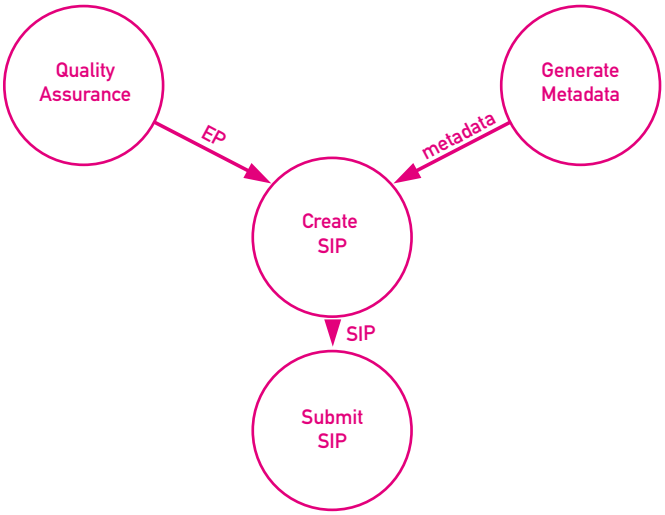


Figure 17 Process Create SIP

The sub-process Create SIP is responsible for packaging deposited publications into standard submission packages.
When activated the Create SIP sub-process scans and structures all directories and files belonging to and accompanying the publication. It combines all components into one or more SIPs.
The SIP package is assigned an identifier, the SIP-id.
A checksum is computed for integrity checking of the SIP package. Both SIP-id and checksum are added to the package itself.

Data Flow
Input: electronic publication [Quality Assurance]
Input: standardised metadata [Generate Metadata]
Output: one or more SIPs [Submit SIP]

Sub-Process Submit SIP

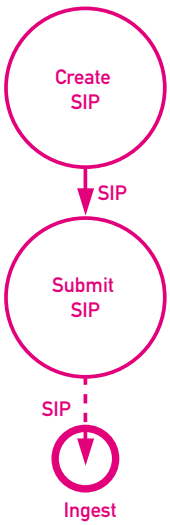


Figure 18 Process Submit SIP

The sub-process Submit SIP is responsible for transmitting standard submission packages (SIP) to Ingest.

Data Flow
Input: SIP [Create SIP]
Output: SIP [Ingest]

4.2 INGEST

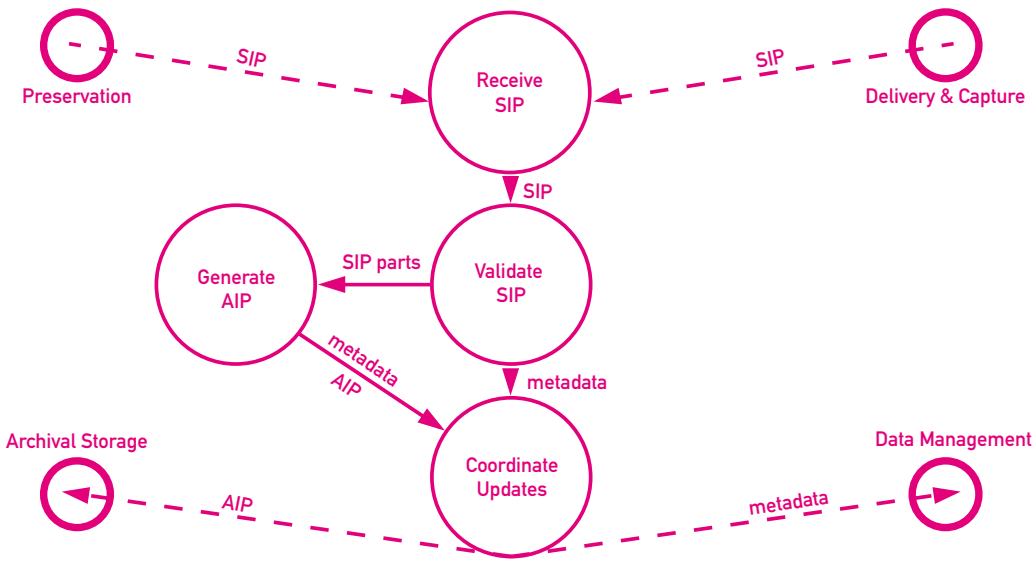


Figure 19 Process Ingest

Ingest receives Submission Information Packages (SIPs) from DSEP's pre-processing interface Delivery&Capture or from the Preservation Process. Ingest disassembles the SIP into its component parts and transfers the data to Archival Storage and Data Management.

Data Flow
Input: SIP [Delivery and Capture]
The SIP specifications are established by the DSEP Administration Process. Different SIP types may be defined with corresponding SIP-qualifiers (example: Publication SIP, Reference Platform SIP, Rendering Software SIP, etc.).
Output: AIP [Archival Storage]
An AIP may contain any of the following:
- the electronic publication
- accompanying metadata
- a full bibliographic description coming from the library cataloguing system
- software, to run or/and view the publication.

An AIP always contains:
- packaging information
- AIP-identifier

Output: Metadata [Data Management]
Metadata for Data-Management can be any of the following:
- Technical metadata
- Structural metadada
- Access control metadata
- Preservation metadata
- Descriptive metadata
- Unique identifier(s) of the logical publication unit(s)
- AIP-identifier
- AIP-qualifier
- Linkage information between related/dependent AIPs

The process Ingest includes the sub-processes Receive SIP, Validate SIP, Generate AIP and Coordinate Updates.

Sub-process Receive SIP

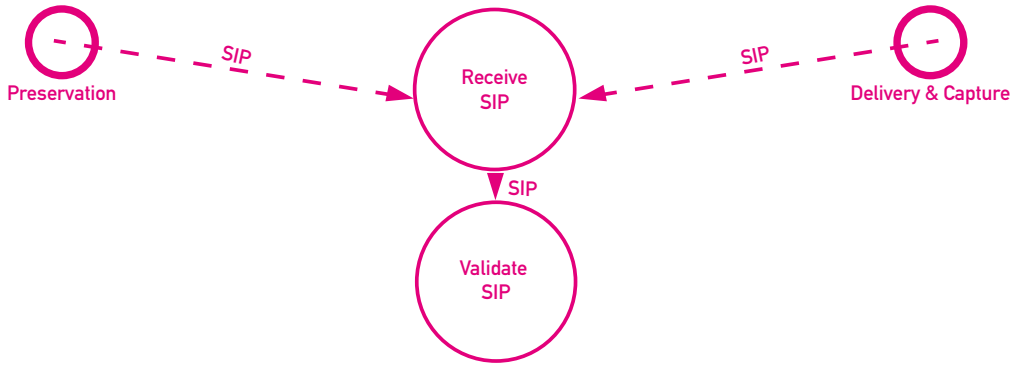


Figure 20 Process Receive SIP

The sub-process Receive SIP receives Submission Information Packages (SIPS) from DSEP's pre-processing interface Delivery&Capture or from the Preservation Process.

Data Flow
Input: SIP [Submit SIP]
Output: SIP [Validate SIP]

Sub-process Validate SIP

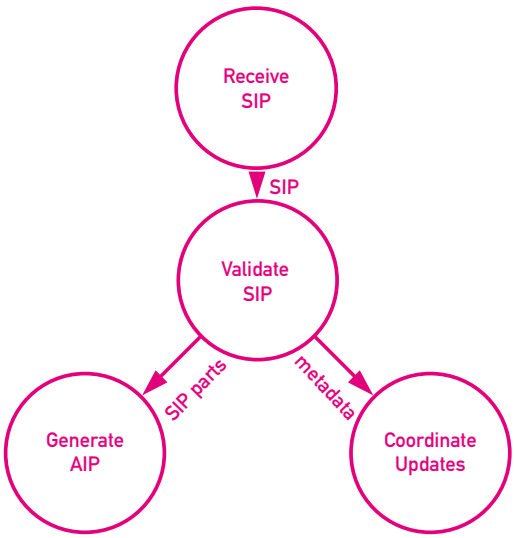


Figure 21 Process Validate SIP

Sub-process Validate SIP performs validation and integrity checks of the package to check if it conforms to the SIP standard. In the process the SIP is unpacked according to the packaging specification and is decomposed into its component parts: e.g. the application software, the data, the metadata, etc.

Data Flow
Input: SIP [Receive SIP]
Output: SIP parts [Generate AIP]
Output: Metadata [Data Management]

Sub-process Generate AIP

The AIP specifications are established by the DSEP Administration Process. Different AIP types may be defined with corresponding AIP-qualifiers (example: Data object AIP, Reference Platform AIP, Rendering Software AIP, Load Image of Installed Version AIP, etc.).

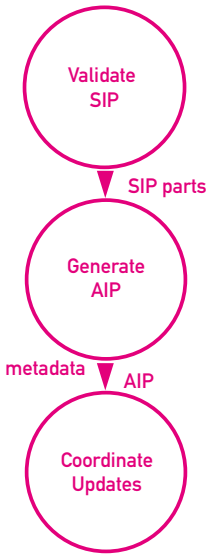


Figure 22 Process Generate AIP

The sub-process Generate AIP repackages component parts of the SIP into one or more Archival Information Packages complying with the archive's information requirements and data packaging standards.

Data Flow
Input: SIP parts [Validate SIP]
Output: AIP [Coordinate Updates]
Output: AIP related metadata [Coordinate Updates]

Packaging strategies

The original bit-stream of the publication's content information is packaged in an AIP.
The application software accompanying the publication can be a component part of the AIP or it can be packaged as a separate AIP. In the latter case the AIP carrying the data should carry a link to the AIP containing the application software. This is usually done in case the application software is common to more than one publication, for example Acrobat Reader version 4.0.
The system disk image of the reference platform is also packaged in a separate AIP, because it is used to run more than one publication.
The load image of the system disk + installed version of the publication may also be kept as a separate AIP, because it can be anticipated that the installed version is accessed more often than the uninstalled one.

Each AIP package is formatted according to documented standards and rules. The content parts are all identified and labelled.
Each AIP package is assigned a unique identifier, the AIP-id and carries a qualifier, the AIP-qualifier.
A checksum is computed for integrity checking of the AIP package.
Links between related/dependent AIPs are established.
AIP-id, AIP-qualifier and AIP-linkages are added to the package itself but are also tagged as AIP-related metadata for Data Management.

Sub-process Coordinate Updates

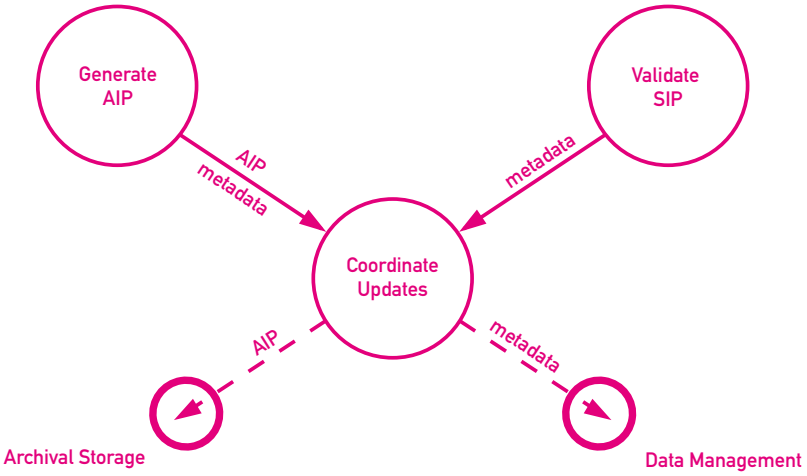


Figure 23 Process Coordinate Updates

The sub-process Coordinate Updates is responsible for co-ordinating updates to Archival Storage and Data Management.
The AIP(s) are sent to Archival Storage. After the transfer is completed and verified, Archival Storage returns a storage confirmation with an indication of the storage identification information for the AIP. The Coordinate Update process then adds the AIP storage identifier and the other AIP related information to the metadata for Data Management. The complete set of tagged metadata is then sent to Data-Management.

Data Flow
Input: AIP [Generate AIP]
Input: AIP related metadata [Generate AIP]
Input: AIP storage identifier [Archival Storage]
Input: Metadata for Data Management [Validate SIP]
Output: AIP [Archival Storage]
Output: Metadata [Data Management]

4.3 ARCHIVAL STORAGE

Archival Storage can handle any type of data object: the electronic publication, software for rendering the publication, image of the installed publication, image of the installed reference platform for running the publication, etc. For Archival Storage only the bitstream of the data object, packaged in the AIP, matters. Archival Storage functions are not interested in the contents of AIPs. They just handle bitstreams. Likewise metadata stored in AIPs are 'meaningless bitstreams' and not functional: Archival Storage cannot act upon them. By contrast, other DSEP processes such as Ingest, Data Management, Preservation, Administration and Access can act on metadata (query, retrieve, evoke processes and other metadata-driven routines).

It should be noted that in the DSEP Model, Archival Storage does not support any update functionality. This conforms with the OAIS-RM. The starting point of the Model is that AIPs are to be kept forever and should not be deleted or modified (replaced by new AIP versions). This starting point may however need to be revisited during practical implementations. In such cases clear rules need to be established for archival storage updates (storage strategies are specified in Administration).

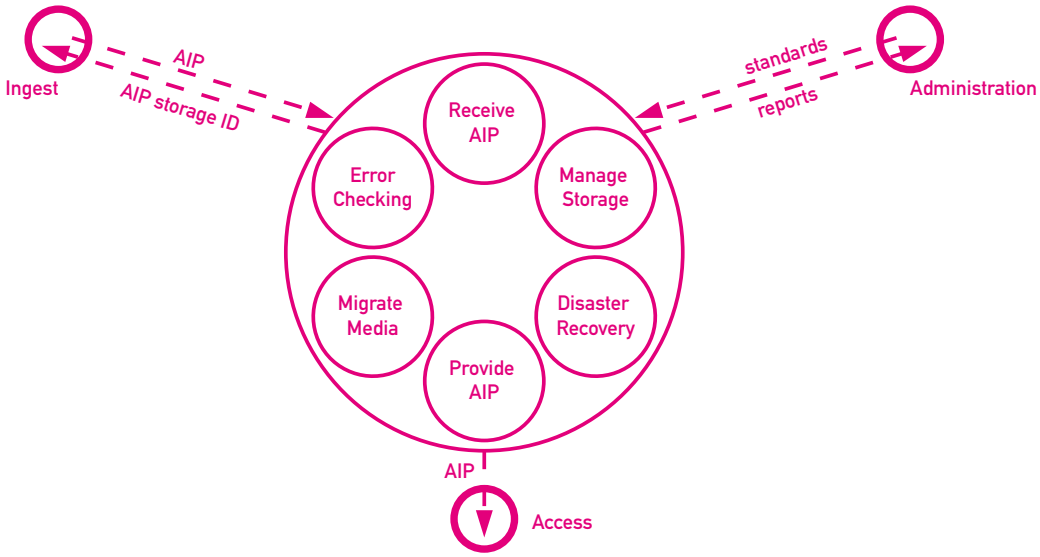


Figure 24 Process Archival Storage

The process Archival Storage is responsible for storage handling and retrieval of the Archival Information Packages (AIP).
Archival Storage only accepts AIPs. The process is responsible for all procedures necessary for secure storage, including storage management procedures, quality assurance, backup, disaster recovery, etc. It also includes regular medium migration (Refreshment and Replication), in order to safeguard the bitstream from decaying carriers. Finally Archival Storage provides AIPs to Access to fulfil access requests.
The Archival Storage process includes the sub-processes Receive AIP, Manage Storage, Error Checking, Migrate Media, Disaster Recovery and Provide AIP.

Data Flow
Input: AIP [Ingest]
Input: Standards [Administration]
Output: Reports [Administration]
Output: AIP storage identifier [Ingest: Coordinate Update]
Output: AIP [Access]

Sub-process Receive AIP

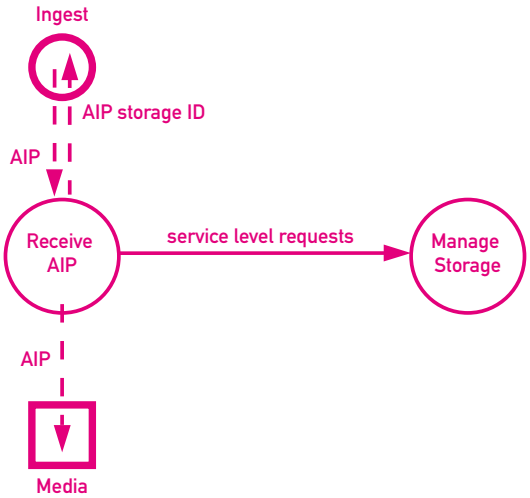


Figure 25 Process Receive AIP

The Receive AIP process receives a storage request and an AIP from Ingest and moves the data to permanent storage within the deposit system. The transfer request may indicate the storage service level requested for the AIP (based for example on the anticipated frequency of utilisation of the data objects comprising the AIP) to allow the appropriate storage devices or media to be selected for storing the AIP. The process will first validate the AIP and then select the media type, prepare the devices or volumes and perform the physical transfer to the Archival Storage volumes. On completion of the transfer this function sends a storage confirmation message to Ingest including the storage identification of the AIP.

Data Flow
Input: AIP [Ingest]
Input: storage requests [Ingest]
Output: AIP [Media]
Output: AIP storage identifier [Ingest]
Output: Service level requests [Manage Media]

Sub-process Manage Storage

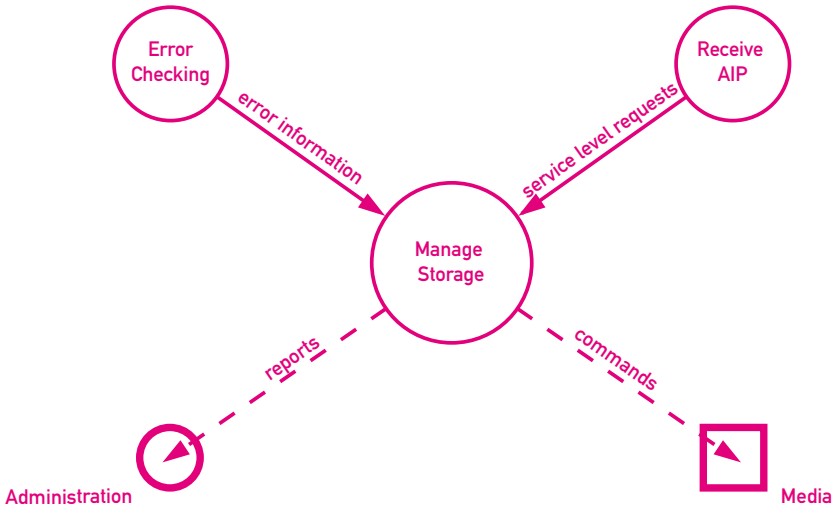


Figure 26 Process Manage Storage

The process Manage Storage positions the contents of the AIPs on the appropriate media based on storage strategies, operational statistics or directions from Ingest via storage request. It applies the storage service level and security measures required for the AIP. It ensures the appropriate level of protection for the AIP. The different levels of service include for example online, offline or near-line storage. They may also include required throughput rate, maximum allowed bit error rate, or special handling procedures.
The process takes care of standard disaster recovery procedures, such as backup and restore procedures. The backup procedures may differ according to the agreed levels of service (e.g. daily or weekly incremental backup of new AIPs).
Finally, the process Manage Storage provides reports on operational statistics to Administration, such as inventory of media, storage capacity and usage statistics.

Data Flow
Input: service level requests [Receive AIP]
Input: error information [Error Checking]
Output: commands [Media]
Output: reports [Administration]

Sub-process Migrate Media (Refreshment or Replication)

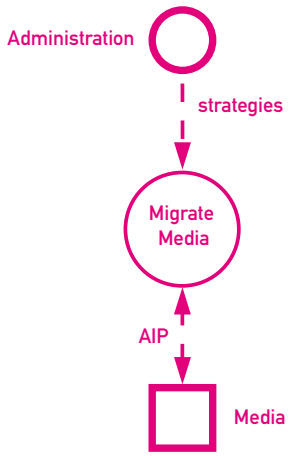


Figure 27 Process Migrate Media

The process Migrate Media is responsible for safeguarding AIPs from decaying media, according to well-defined media migration strategies (defined by Administration). The strategies establish which storage medium should be chosen, taking into account the expected and actual rates of error encountered in various media types, their performance and costs. Media technology developments (higher capacity media, lower costs) may also trigger media migration strategies. The Migrate Media process selects the AIPs that have to be migrated and performs either Refreshment or Replication. These types of migrations do not affect the content of the AIPs. When migration to different storage architectures affect the packaging or the content information of the AIP, the Preservation process is tasked to perform the migration.

Data Flow
Input: strategies [Administration]
Input: AIP [Media]
Output: AIP [Media]

Sub-process Error Checking

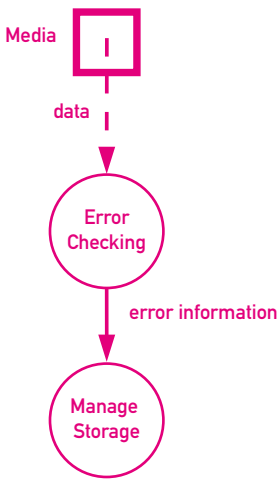


Figure 28 Process Error Checking

The process Error Checking provides statistically acceptable assurance that no components of the AIP are corrupted during any internal Archival Storage transfer. A standard mechanism for file error checking (e.g. CRCs) may be established for random verification of the integrity of AIPs.

Data Flow
Input: data [Media]
Output: error information [Manage Storage]

Sub-process Disaster Recovery

Figure 29 Process Disaster Recovery

The disaster recovery process provides the services for duplicating the digital contents of the deposit collection and storing the duplicate in a separate facility elsewhere. It may for example also be outsourced to a third party providing such a service. Disaster Recovery is a special form of backup. The details of disaster recovery strategies are specified by Administration.

Data Flow
Input: AIP [Media]
Output: AIP [Other Media]

Sub-process Provide AIP

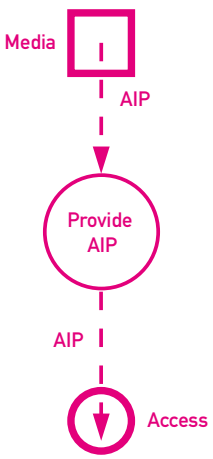


Figure 30 Process Provide AIP

The Provide AIP process receives a request for AIP from Access and retrieves a copy of a stored AIP. It transfers the copy of the AIP to an Access staging area. The process sends a notice of data transfer to Access.

Data Flow
Input: AIP [Media]
Input: AIP request [Access]
Output: AIP [Access]
Output: notice of data transfer [Access]

4.4 DATA MANAGEMENT

Data-Management is responsible for handling all DSEP metadata. We distinguish between two types of metadata:

- metadata associated with the electronic publication, such as descriptive metadata, structural metadata, technical metadata, access control metadata, preservation metadata, etc.
- metadata associated with the administration of the DSEP, such as operational statistics, status report information, etc.

The metadata associated with the publication may also be duplicated in other (external) systems. This raises however the implementation issue of synchronisation of metadata updates. The DSEP model takes as starting point that duplication of metadata should be avoided. There is a clear separation of functional metadata types. The cataloguing process, which creates a title-description of the electronic publication and involves subject indexing, takes place in the cataloguing environment of the digital library system. It may re-use primary metadata provided by the publisher. It may also return full bibliographic descriptions to the DSEP system, through the Delivery&Capture interface.
In principle, however, Data Management does not support cataloguing and it does not provide search & discovery functionality.

The metadata associated with the administration of the DSEP consist of pre-defined statistical information about system usage, capacity, up-time, etc. In addition there are also system bound operational data, recorded in flat ASCII files, that cannot be structured in database fields. These data provide important feed-back about the system's operational flow to the system administrator, recording all activities and the status of routines performed. They are stored by Data Management as well, but cannot be handled as metadata.

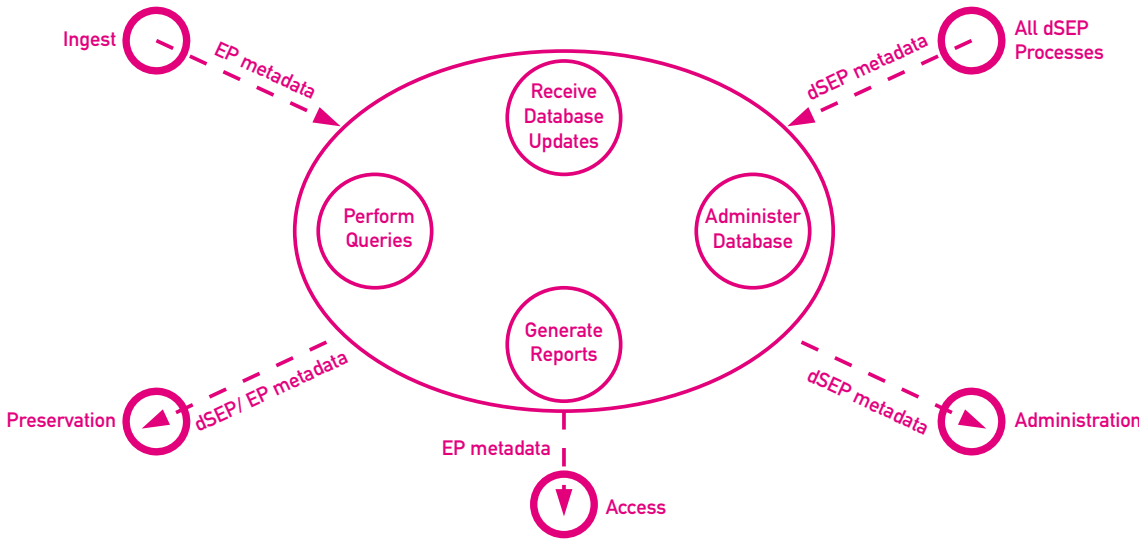


Figure 31 Process Data Management

Data-Management is responsible for storing and retrieving metadata necessary for the handling of electronic publications within the deposit system and for the daily operations of the deposit system.

Data Flow
Input: metadata associated with the EP [Ingest]
Input: metadata associated with DSEP [all DSEP-processes]
Output: metadata associated with the EP [Access, Preservation]
Output: metadata associated with DSEP [Administration, Preservation]

The process Data Managment includes the sub-process Receive Database Updates, Administer Database, Perform Queries and Generate Reports.

Sub-process Receive Database Updates

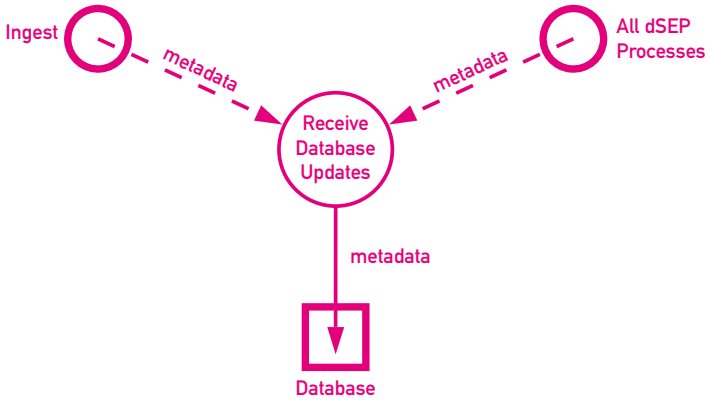


Figure 32 Process Receive Database Updates

The process Receive Database Updates is responsible for metadata storage handling. Through this process it is possible to add, modify or delete information in the Data Management storage. The main sources of updates are:

- Ingest, which provides metadata related to newly ingested publications and the new AIPS,
- All DSEP-processes, providing a regular stream of operational statistics.

In all cases Receive Database Updates checks if the metadata adhere to DSEP standards and formats.

Data Flow
Input: metadata associated to the EP [Ingest]
Input: metadata associated with DSEP [all DSEP-processes]
Output: metadata [Database]

Sub-process Administer Database

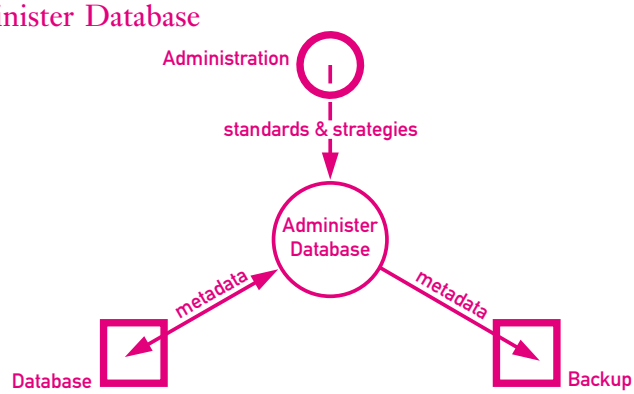


Figure 33 Process Administer Database

The process Administer Database is responsible for maintaining the integrity of the Data Management database. It checks if the metadata conforms to standards specified by Administration (e.g. tagged

metadata, standardised metadata, mandatory link between metadata and identifier, etc.). It maintains metadata schema definitions and format specifications, it translates between schemas by use of semantic mappings and between formats by use of data-field-mappings, it ensures linkage integrity between metadata, etc. The process also takes care of standard disaster recovery procedures, such as backup and restore of the database.
The process carries out its tasks in accordance with the strategies and standards specified by Administration.

Data Flow
Input: standards and strategies [Administration]
Input: metadata [Database]
Output: metadata [Database, Backup]

Sub-process Perform Queries

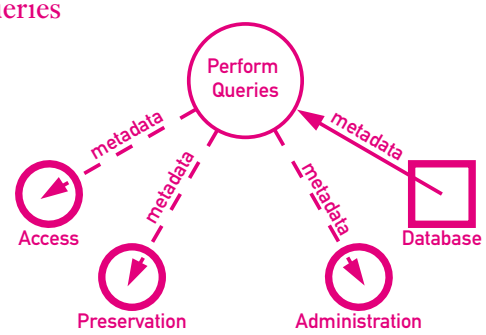


Figure 34 Process Perform Queries

The process Perform Queries receives a query from Access, Administration or Preservation and executes the query to generate a result set that is transmitted to the requester.

Data Flow
Input: request [Access, Administration, Preservation]
Input: metadata [Database]
Output: metadata [Access, Administration, Preservation]

Sub-process Generate Report

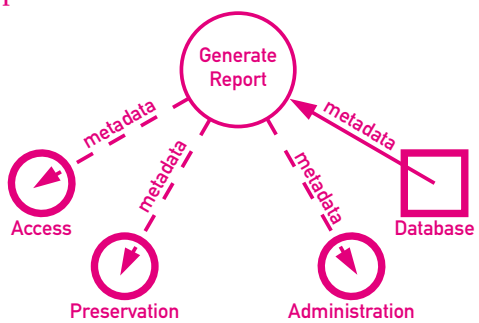


Figure 35 Process Generate Report

The process Generate Report receives a report request from Administration, Preservation or Access. Typically, reports are produced on the basis of pre-defined templates to present result sets generated from (pre-defined) queries on metadata.
Reports to Access usually are standard forms with access metadata about a given AIP.
Reports to Administration include reports about the deposit holdings, usage statistics, system functioning, etc.

Reports to Preservation typically provide preservation metadata about collections of AIPs.
The reports can be generated according to a pre-defined schedule, but also on-the-fly and on an ad-hoc basis. The reports can be viewed, printed and downloaded to spreadsheets, etc.

Data Flow
Input: report request [Access, Administration, Preservation]
Input: metadata [Database]
Output: report [Access, Administration, Preservation]

4.5 ACCESS

In the DSEP model the scope of Access is much more limited than in the OAIS model, because many related processes, such as creating finding aids, registering library users, applying access controls, etc. belong intrinsically to the wider digital library context and not specifically to a DSEP.
DSEP does not interface directly to the library end-user but to the library's access-system via the post-processing interface 'Packaging&Delivery'.

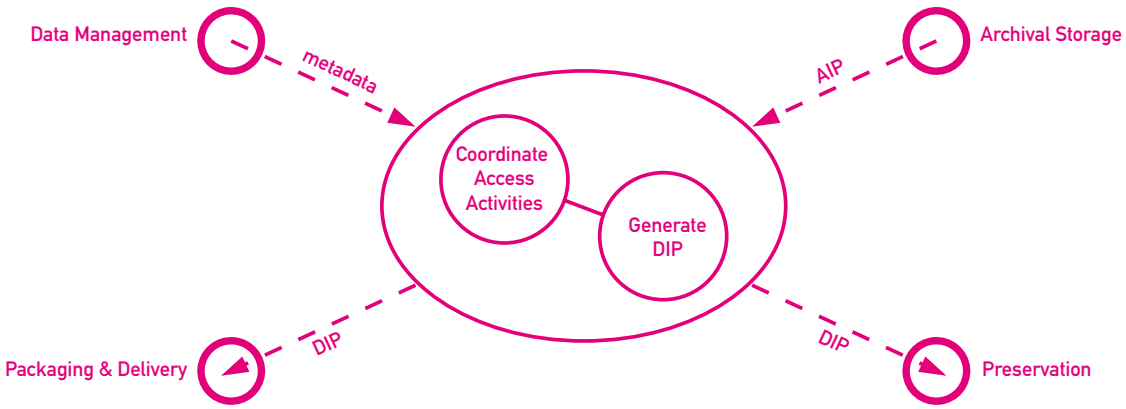


Figure 36 Process Access

Access makes electronic publications available from Archival Storage and associated metadata from Data Management. It generates Dissemination Information Packages (DIPs) and submits these upon request to DSEP's post-processing interface Packaging&Delivery or to the Preservation Process.

Data Flow
Input: AIP [Archival Storage]
Input: metadata [Data Management]
Output: DIP [Packaging&Delivery, Preservation]
A DIP may contain any of the following:
- The electronic publication
- Tagged and standardised metadata
- Software, to run or/and view the publication.

A DIP always contains:
- Packaging information
- DIP-identifier

The process Access includes the sub-processes Coordinate Access Activities and Generate DIP.

Sub-process Coordinate Access Activities

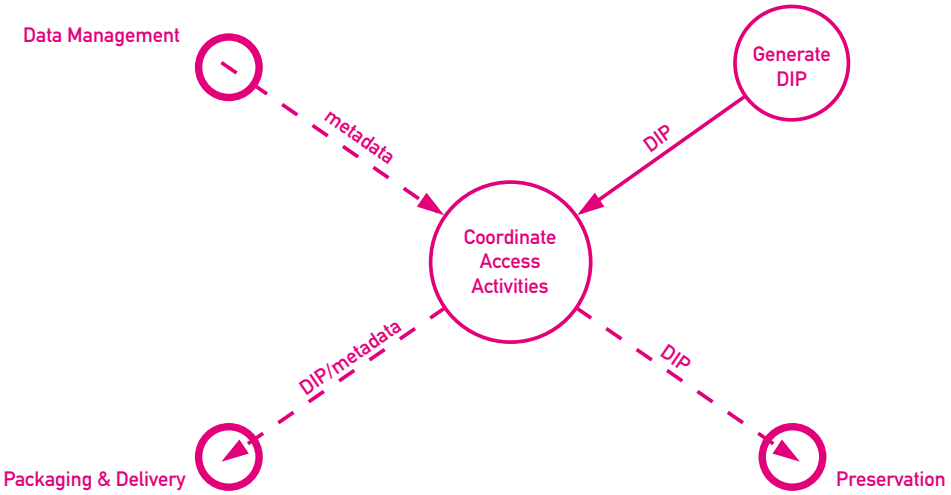


Figure 37 Process Coordinate Access Activities

The sub process Coordinate Access Activities handles requests for DIPs, coming from Packaging&Delivery or Preservation. It assures that the request comes from one of these two authorised processes. It then transfers the request to Data Management or to the Generate DIP process for execution, depending on the type of request.

Access requests can lead to negotiations in order to define which item from the deposit holding is to be delivered. This is typically the case when a request for access comes from Packaging&Delivery. This request usually generates first a report with access metadata related to the requested deposit item before it is effectively retrieved. This access metadata provides information about the requested original version and all related versions resulting from the Preservation process. This information includes information on systems requirements, retrieval time, access controls, etc. and enables Packaging&Delivery to decide which version suits best the user's demand and technical environment.
An access request from Preservation, on the contrary, is final and unambiguous because it can query Data Management directly and has knowledge of the preservation characteristics of the deposit holdings on beforehand.

When an access request is final this means that the requested version of the deposit item has been identified. Coordinate Access Activities then generates the final metadata report request to accompany the AIP-request and combines both in a DIP request to Generate DIP.

Coordinate Access Activities manages the status of access requests and activates Generate DIP to package the request results into standard packages. Finally it delivers the resulting DIP to the requesting process.

Data Flow
Input: DIP request [Packaging&Delivery, Preservation]
Input: DIP [Generate DIP]
Input: metadata [Data Management]
Output: metdata request [Data Management]
Output: DIP request [Generate DIP]
Output: DIP [Packaging&Delivery, Preservation]

Sub-process Generate DIP

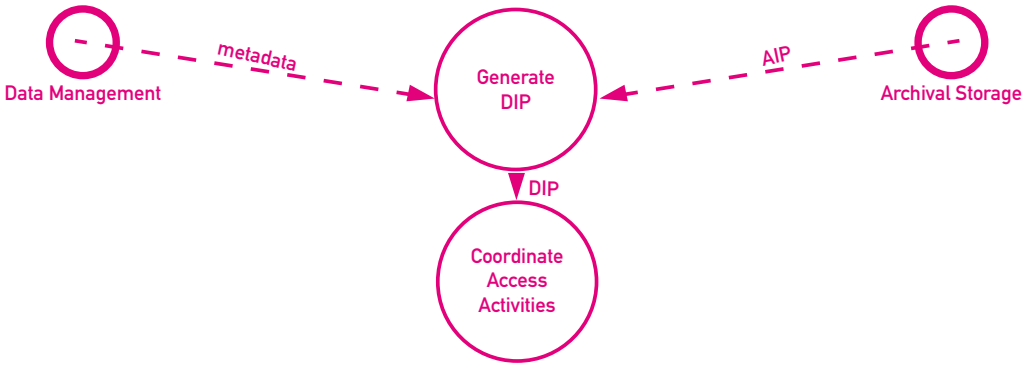


Figure 38 Process Generate DIP

The process Generate DIP is responsible for packaging the request results into standard Dissemination Information Packages. When activated by Coordinate Access Activities it processes the request. It passes on metadata report requests to Data Management and/or AIP requests to Archival Storage. It then packages the metadata report prepared by Data Management and the requested AIP(s) from Archival Storage into a DIP, as specified by Coordinate Access Activities.

Data Flow
Input: DIP request [Coordinate Access Activities]
Input: AIP [Archival Storage]
Input: metadata [Data Management]
Output: DIP [Coordinate Access Activities]

4.6 PACKAGING&DELIVERY

Packaging&Delivery is conceptualised as the 'post-processing' interface to the DSEP. It is needed because deposit libraries cannot anticipate all access modes and variables, such as user authorisation, access rights, publisher license access conditions and other access controls. Presently, for example, most deposit license agreements with publishers permit installation of publications onsite only, on a library workstation, and access by registered library users only.
These conditions will change through time with evolving business models, technological possibilities and user demands.

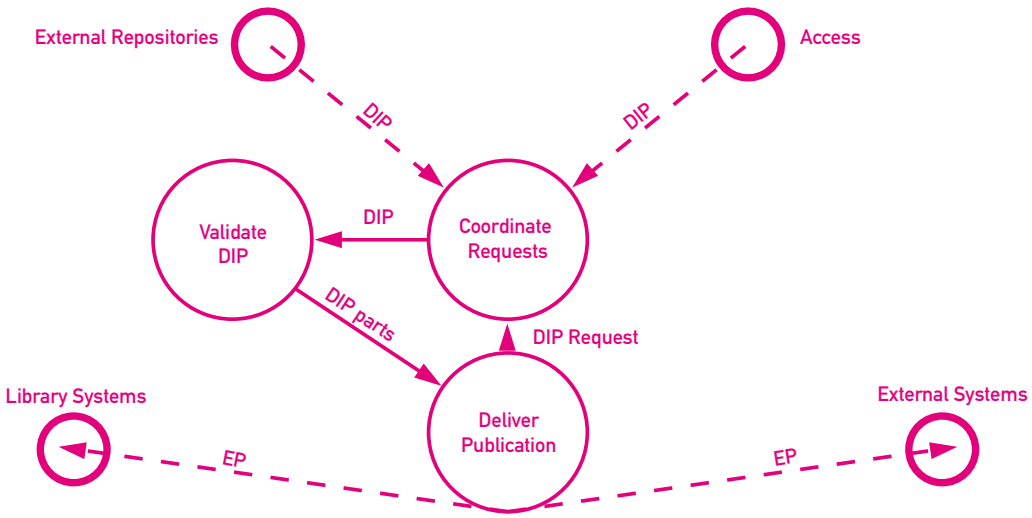


Figure 39 Process Packaging&Delivery

The Packaging&Delivery process is responsible for delivering deposit publications upon request to the library access systems or external systems. It takes care of the request negotiations with DSEP and of all processes needed to unpack a DIP and to make it fit for use by the requesting system.
This may entail extracting parts of the electronic publication, or adding a full-text index to it, or converting (parts) of the publication into appropriate formats for viewing, printing or downloading. It may involve providing a viewer or even interpreter and emulation software for displaying the original version of the publication.

Packaging&Delivery is the only process that can request and accept DIPs (Dissemination Information Package) from the Access module of a DSEP. The DIP consists of the requested publication in one of the available formats with accompanying software and/or metadata needed to install and display it, to assess its authenticity or to reconstruct the original copy.

Packaging&Delivery may also request DIPs from Representation Information Repositories, with which the deposit library has made contractual agreements. This enables the library to use a common pool of Representation Information, shared by a variety of deposit libraries and other memory organisations. These Repositories provide open standard format specifications, common viewing software, conversion software, interpreters, emulators, etc.

Data Flow
Input: DIP [Access, External Repositories]
Output: EP [Library systems, External systems]

The process Packaging&Delivery includes the sub-processes Coordinate Requests, Validate DIP, Deliver Publication.

Sub-processes Coordinate Requests

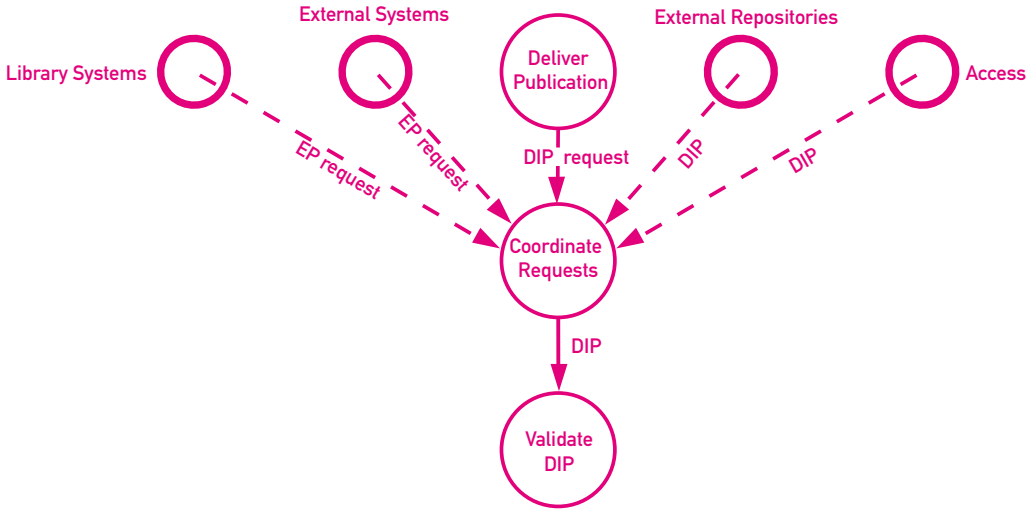


Figure 40 Process Coordinate Requests

The sub-process Coordinate Requests is responsible for the co-ordination of all requesting activities needed to successfully deliver a publication from DSEP. The process is initiated by a request for an electronic publication. The request may come from the library access-system or from an external system. The process supports identification, authentication and authorisation functionality in order to be able to react to these requests.
Before the requested publication is presented to the requesting system several questions need to be answered first, such as:

1. Does the user have access right to the publication?
2. Which version of the publication does the user need?

3 What is the user system configuration on which the publication should be viewed/run?

The Coordinate Requests process starts a negotiation session with the requesting party in order to get the right answers to all these questions. To support this negotiation it needs to get user-related access information from the library-access system. It also needs to get publication-related access control and technical metadata from DSEP's Data Management via Access.

Once the process is assured of the user permissions and has identified which (version of the) publication is requested, it activates the requests for the actual AIP(s) from Archival Storage via Access.

The process may also request DIPs from external repositories (see Representation Information Repositories).

Data Flow

Input: request publication [Library systems, External systems]

Input: DIP [Access, External Repositories]

Output: request DIP [Access]

Output: DIP [Validate DIP]

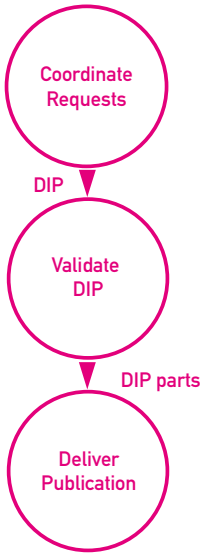


Figure 41 Process Validate DIP

Sub-processes Validate DIP

The sub-process Validate DIP receives the requested DIP, performs validation and integrity checks. In the process the DIP is unpacked and de-composed into its component parts.

Data Flow

Input: DIP [Coordinate Requests]

Output: DIP parts [Deliver Publication]

Sub-processes Deliver Publication

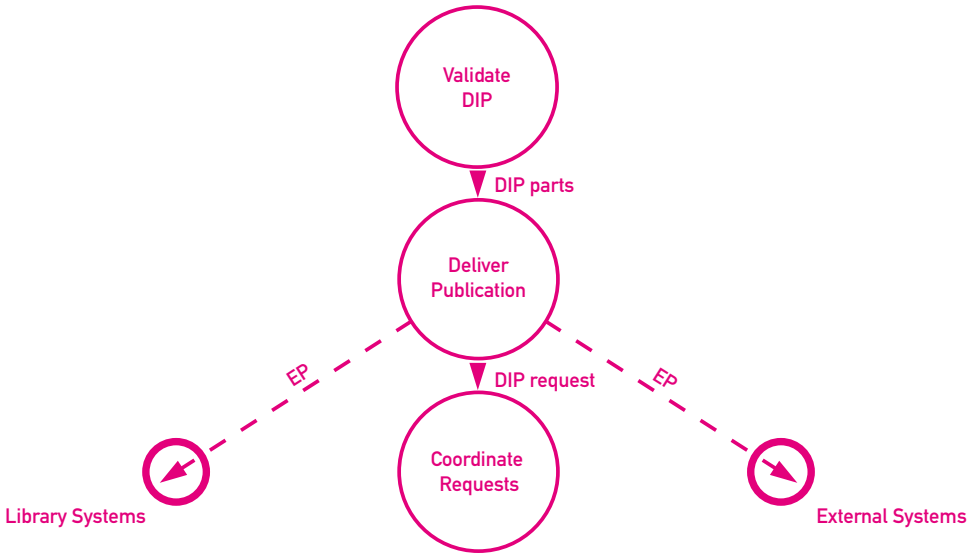


Figure 42 Process Deliver Publication

The sub-process Deliver Publication is responsible for delivering deposit publications to the requesting system. Sometimes only request forms providing additional information about the publication are delivered. As a rule this process delivers the actual publication. This is done on the basis of the unique identifier of the logical publication unit requested. The process ensures that all data necessary to deliver the requested item are successfully retrieved: the accompanying metadata, the publication itself, if required the associated viewing software or the corresponding Reference Platform.
The technical metadata accompanying the publication activate pointers or linkages to associated software AIPs or to external software repositories. This results in requests to Access or to Representation Information Repositories.
Finally the publication is 'installed' on the library reference platform for access. If the user is not using a library reference platform, the process can either transfer the electronic publication and all its accompanying data to the user workstation or it can install the publication on the library server and provide remote viewing facilities in terminal mode.

Data Flow

Input: DIP parts [Validate DIP]
Output: request [Coordinate Requests]
Output: EP [Library systems, External systems]

4.7 ADMINISTRATION

In the DSEP model Administration is scoped to the management of DSEP operations and DSEP process & quality control. Administration does not manage the content and all content related aspects such as AIP transformations, monitoring electronic publishing formats and digital preservation technologies. In particular it does not manage change required to keep the content accessible through time. These aspects are tackled by the Preservation process.

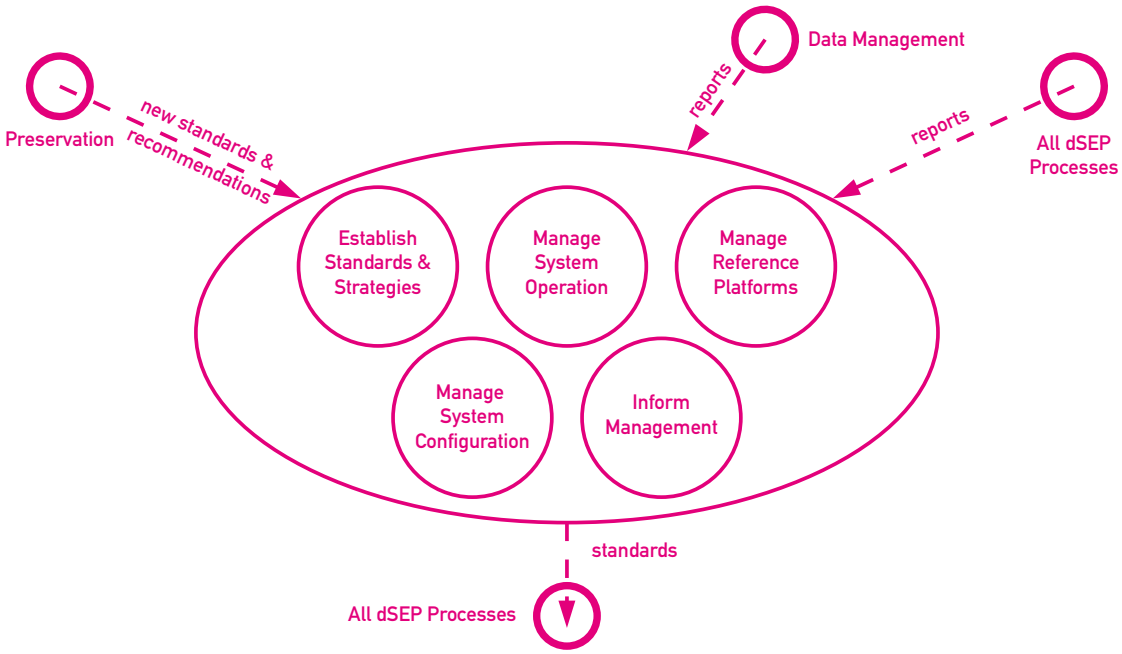


Figure 43 Process Administration

The Administration process is responsible for the overall operation of the deposit system. It performs quality control, based on documentation of DSEP standards, procedures, strategies and service levels, management of the DSEP system configuration, continuous monitoring of all DSEP-processes, provision of regular feed back on operational statistics to support deposit operation improvement, etc.
Administration reports to the deposit system administrator. The deposit system administrator is process-owner therefore this internal output is not visible in the data-flow.
Administration produces management information to the deposit library management, including those responsible for collection management and deposit policy making.

Data Flow
Input: policies [Management]
Input: reports [all DSEP-processes]
Input: new standards [Preservation]
Output: management information [Management]
Output: formats, standards, procedures, strategies [all DSEP-processes]

The process Administration includes the sub processes Manage System Operation, Manage System Configuration, Manage Reference Platforms, Establish Standards&Strategies and Inform Management.

Sub-process Manage System Operation

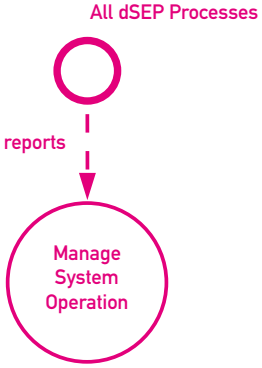


Figure 44 Manage System Operation

The process Manage System Operation monitors system operation, system performance, and system usage. It monitors the status reports coming from all DSEP processes and operational statistics coming from Data Management and Archival Storage. It reports back to the administrator, either upon request or by notification.

Data Flow

Input: reports [all DSEP-processes]

Sub-process Manage System Configuration

Figure 45 Process Manage System Configuration

The process Manage System Configuration is responsible for systematic control of changes to the system configuration. It maintains the integrity and tractability of the system hardware and software configuration. It records system migrations and manages system upgrades through time.

Data Flow

Not Applicable

Sub-process Manage Reference Platforms

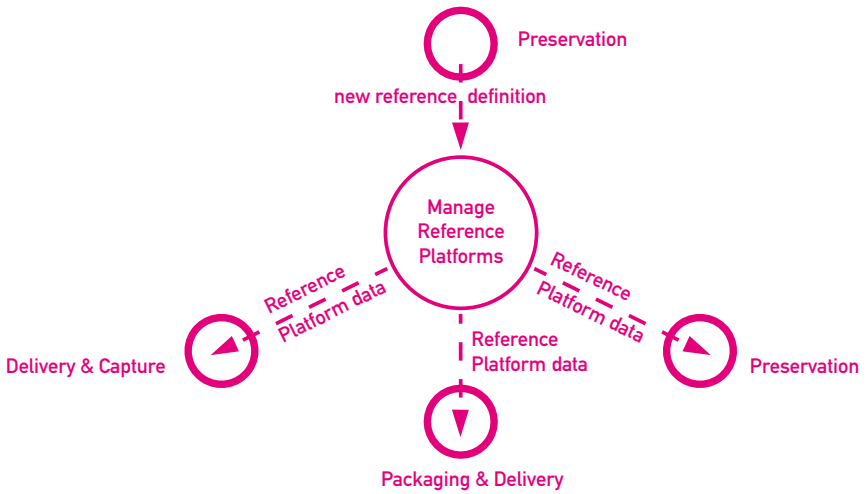


Figure 46 Process Manage Reference Platforms

The process Manage Reference Platforms is responsible for adding, deleting, updating Reference Platform configurations in the basis of definitions of new Reference Platform definitions provided by Preservation. It maintains the integrity and tractability of the Reference Platforms supported by and ingested into DSEP.

Data Flow

Input: definitions of new Reference Platforms [Preservation]
Output: Reference Platform Information [Delivery&Capture, Packaging&Delivery, Preservation]

Sub-process Establish Standards&Strategies

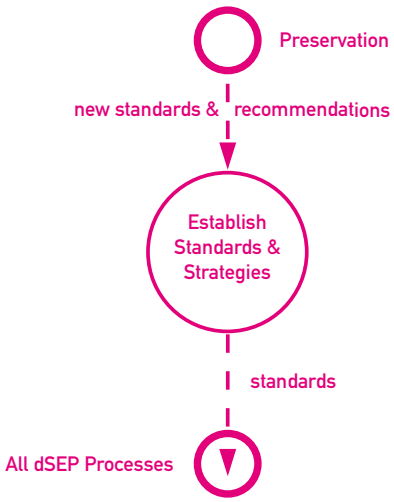


Figure 47 Process Establish Standards&Strategies

The Establish Standards&Strategies process is responsible for documenting and maintaining DSEP standards (DIP, AIP, SIP, tagged metadata categories and formats), procedures, schedules, submission agreements with publishers, storage and access strategies, security levels, system user permissions, service level agreements, report definitions (deposit holdings inventory reports, access negotiation reports, etc.). These are all established by the responsible deposit administrator and based on deposit policy development by the library management. The standards and strategies are sent to all affected DSEP processes.

Data Flow
Input: policies [Management]
Input: new standards and recommendations [Preservation]
Output: formats, standards, procedures, strategies [all DSEP-processes]

Sub process Inform Management

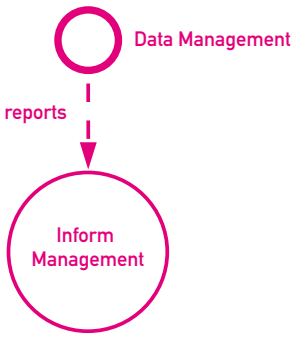


Figure 48 Process Inform Management

The process Inform Management is responsible for producing management information to the deposit library management, including those responsible for collection management and deposit policy making.

Data Flow
Input: reports [Data Management]
Output: reports [Management]

4.8 PRESERVATION

At the time of writing, the OAIS model does not explicitly include a preservation module.
However, as a result of the ISO-review process in the first half of 2000, the current draft CCSDS recommendation for an OAIS-RM [21] does propose to add a 'Preservation Planning' entity. This version of the DSEP model is based on the latest recommendation document.

As formats become obsolete and the viewers needed to interpret and render these formats also become obsolete, it will be necessary to take measures to preserve the content of a publication and all related aspects such as data, layout, structure and functionality. To this end, several strategies may be followed, such as document format specification, conversion and emulation.
The deposit library preservation policy defines the preservation strategies that are to be applied by the DSEP. Both transformation and emulation approaches are worked out in some detail in this DSEP model. The resulting output is either a new version of a formerly deposited publication, in which case it is ingested anew in the system, or it is a set of specifications for building emulators that can render a whole generation of publications on a future (unknown) reference platform. In both cases, new SIPs are ingested into the DSEP with new preservation metadata fed into Data-Management.

21 op.cit. 18

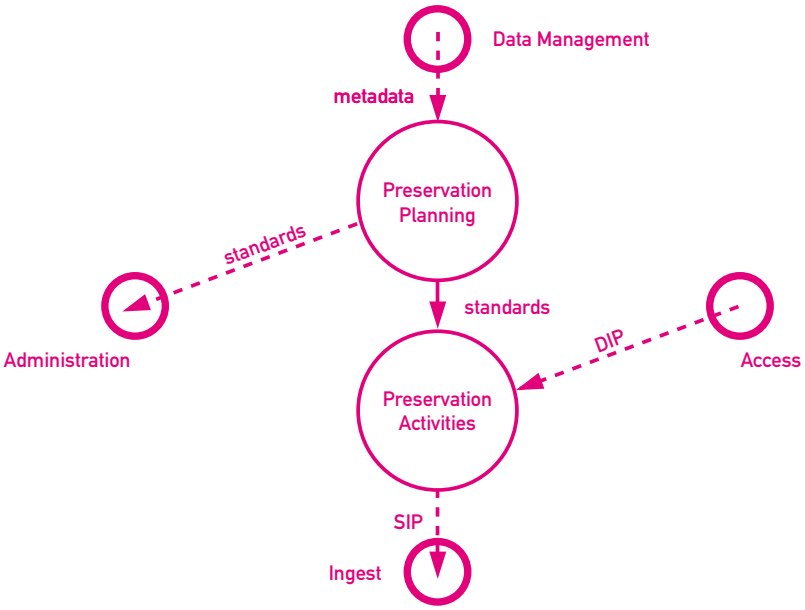


Figure 49 Process Preservation

The Preservation process is responsible for managing the accessibility of the deposit content through time, even if the original computer environment becomes obsolete.

Data Flow
Input: preservation policies [Management]
Input: metadata [Data Management]
Input: DIP [Access]
Output: formats, standards, procedures, strategies [Administration]
Output: SIP [Ingest]
The Preservation process consists of two sub-processes: 1) *Preservation Planning*, which takes care of research and development activities prior to implementing preservation strategies and 2) *Preservation Activities*, which takes care of actual preservation activities and ensures that the conditions for preservation are met within DSEP.

Sub-process Preservation Planning

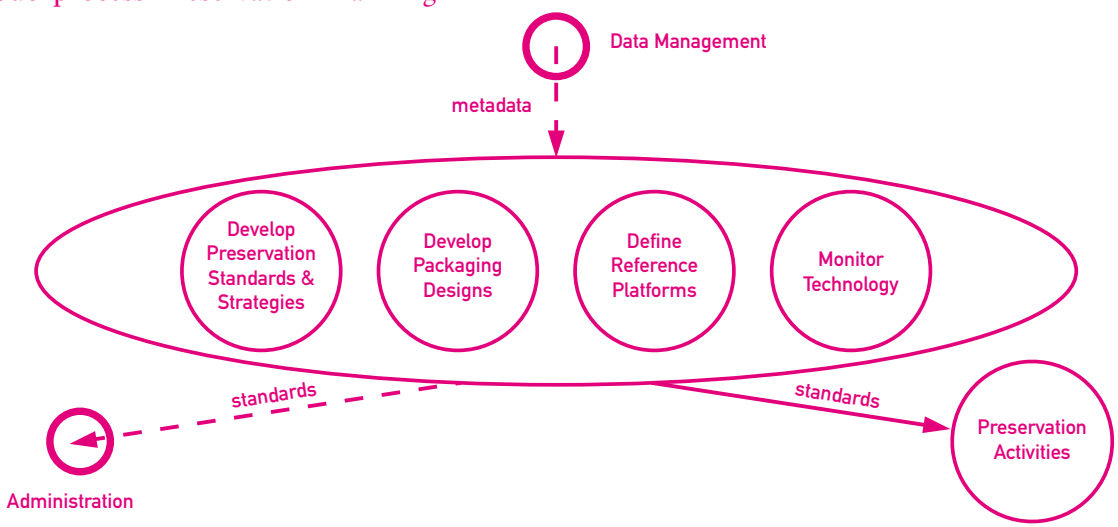


Figure 50 Process Preservation Planning

The Preservation Planning process is responsible for research and development necessary as a continuous effort to ensure accessibility and readability of the deposit holdings through time.

Data Flow
Input: preservation policies [Management]
Input: metadata for preservation [Data Management]
Output: formats, standards, procedures, strategies [Preservation Activities, Administration]

Preservation Planning includes the sub-processes Develop Preservation Standards&Strategies, Develop Packaging Designs, Define Reference Platforms and Monitor Technology.

Sub-process Develop Preservation Standards&Strategies

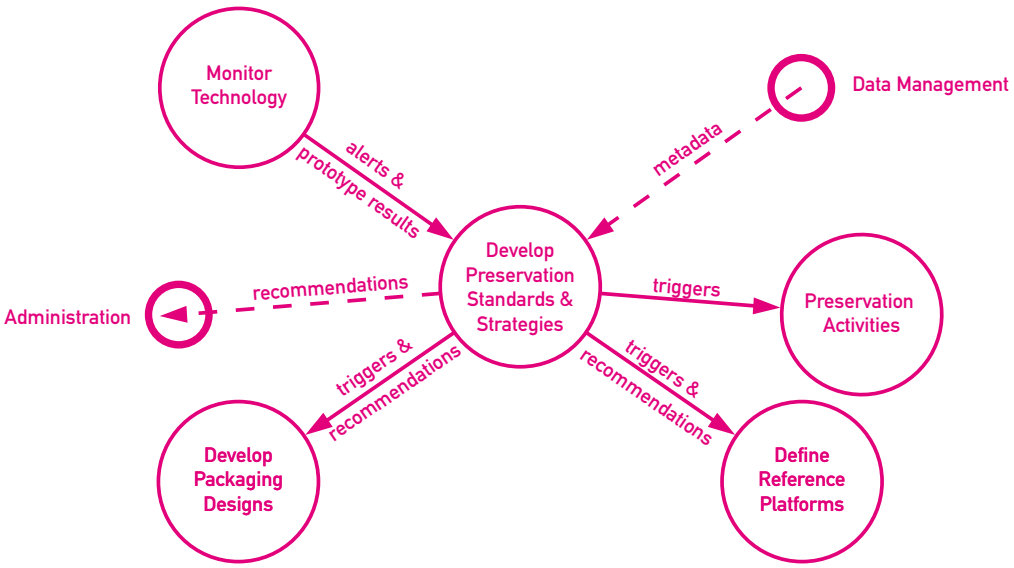


Figure 51 Process Develop Preservation Standards&Strategies

The process Develop Preservation Standards&Strategies is responsible for developing and recommending strategies and standards to enable the archive to better anticipate future changes in electronic publishing, user demands, technology trends, etc.
It receives and evaluates reports from Monitor Technology, establishes new requirements and sends requests for prototyping back to Monitor Technology. On the basis of prototyping results it recommends adoption of new standards, strategies and preservation technologies to Administration.
On the basis of technology alerts it recommends the definition of new reference platforms through time and/or develops (agreements with) *Representation Information Repositories* for the provision of open standard format specifications, common viewing software, conversion software, interpreters, emulators, etc. In this context it co-ordinates the definition, creation and management of reference platforms.
It recommends the development of packaging designs and co-ordinates upgrades and migrations in close co-operation with Administration. It triggers Preservation Activities to perform retrospective upgrades of old AIPs in Archival Storage.
Finally it can request metadata from Data Management to support the preservation planning process.

Data Flow
Input: deposit policies [Management]
Input: metadata for preservation [Data Management]
Input: technology alerts [Monitor Technology]
Input: prototype results [Monitor Technology]
Output: recommendations [Administration, Develop Packaging Designs, Define Reference Platforms]
Output: triggers [Develop Packaging Designs, Define Reference Platforms, Preservation Activities]

Sub-process Develop Packaging Designs

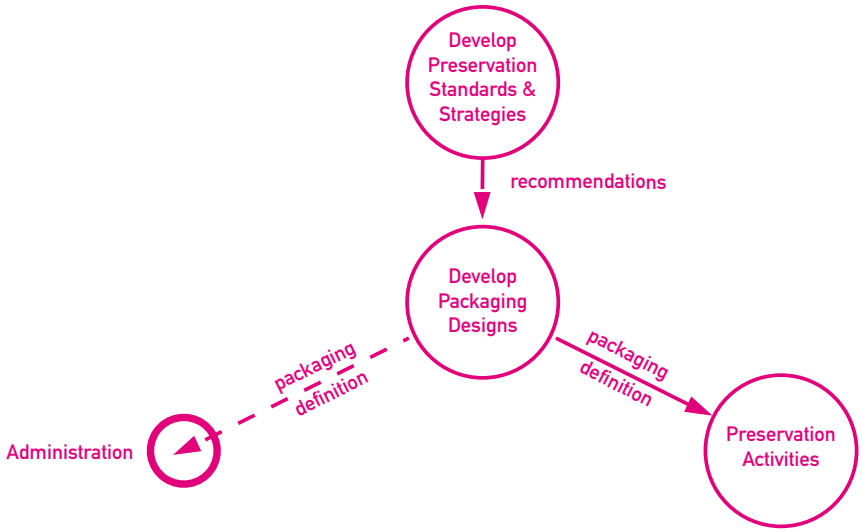


Figure 52 Process Develop Packaging Designs

The process Develop Packaging Designs is responsible for data re-design through time: it defines new AIP, SIP and DIP packaging designs for DSEP. It provides recommendations on specific application of these designs to specific deposit holdings, submissions and disseminations. It works on the basis of current packaging standards and formats as managed by Administration and applies these standards to evolving preservation requirements and technological changes. It provides new packaging designs to Administration and Preservation Activities.

Data Flow

Input: recommendations [Develop Preservation Strategies and Standards]
Output: Packaging Definition [Preservation Activities, Administration]

Sub-process Define Reference Platforms

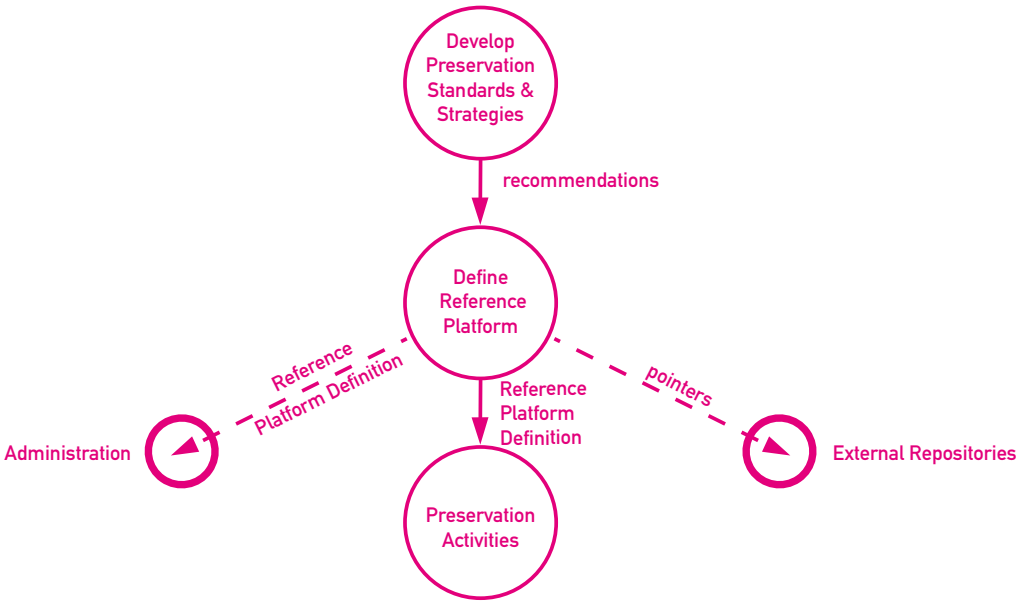


Figure 53 Process Define Reference Platforms

The process Define Reference Platforms is responsible for defining new Reference Platforms for DSEP. This is done when new electronic publications do not run on the current reference platform and

requires different/newer software versions or hardware parts. A reference platform is defined in terms of its consisting parts: the platform (hardware system, operating system and /or emulator) and a number of software applications (typically viewers).
This process interacts interacts closely with external Representation Information Repositories.

Data Flow
Input: recommendations [Develop Preservation Strategies and Standards]
Output: Reference Platform Definition [Preservation Activities, Administration]
Output: Pointers [External Repositories]

Sub-process Monitor Technology

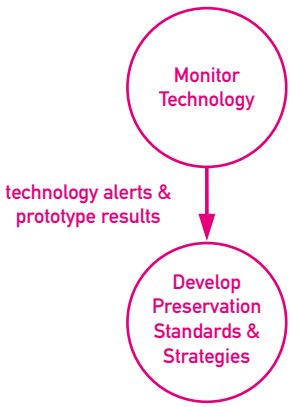


Figure 54 Process Monitor Technology

The process Monitor Technology functions as a technology watch in the areas of digital preservation, electronic publishing and end-user ICT-configurations.
It tracks emerging digital technologies, information standards and computing platforms to identify technology which could cause obsolescence in the deposit's environment and prevent access to some of the deposit holdings.
This process manages the Preservation Layer Model (PLM). This is a model of the different layers of Representation Information, based on the layered Information Model from Annex E of the OAIS-RM and adapted by NEDLIB.[22] Monitor Technology keeps track of all software and hardware dependencies of electronic publications that appear on the market.
This process may include a prototyping capability for better evaluation of emerging technologies.

Data Flow
Input: prototype request [Develop Preservation Strategies and Standards]
Output: technology alerts and prototype results [Develop Preservation Strategies and Standards]

22 op. cit. 12

Sub-process Preservation Activities

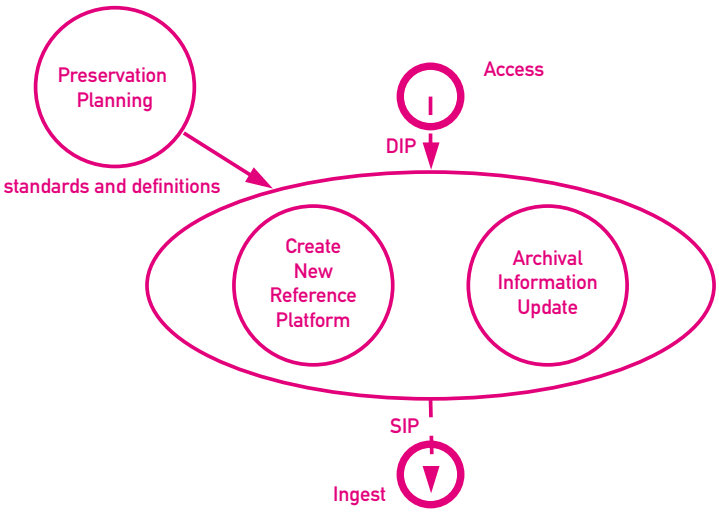


Figure 55 Process Preservation Activities

The Preservation Activities process takes care of the actual preservation processing activities and ensures that the conditions for preservation are met within DSEP.

Data Flow
Input: standards [Preservation Planning]
Input: DIP [Access]
Output: SIP [Ingest]

The Preservation Activities process includes the sub-processes Create New Reference Platform, and Archival Information Update.

Sub-process Create New Reference Platform

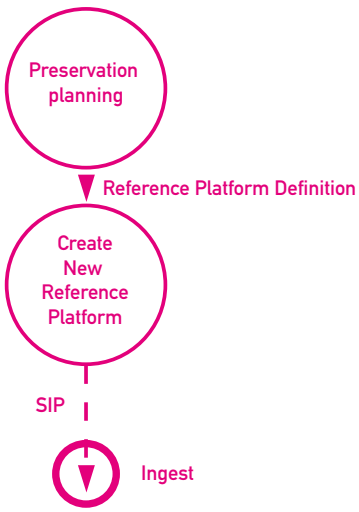
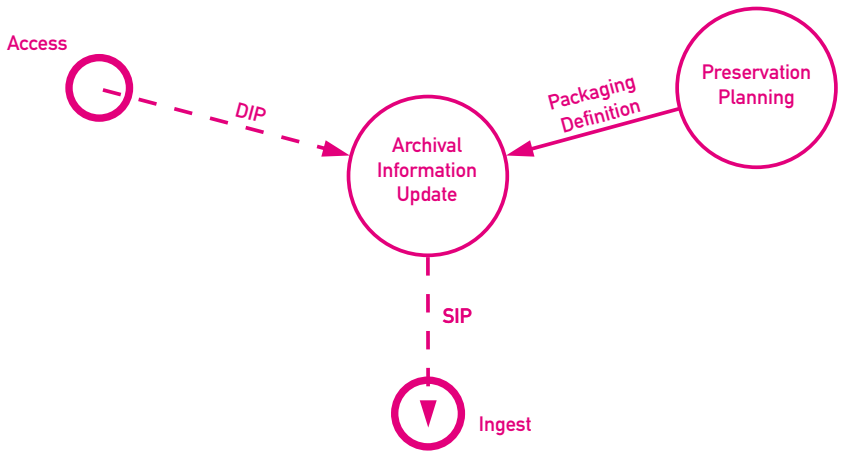


Figure 56 Process Create New Reference Platform

The process Create New Reference Platform is responsible for creating a newly defined reference platform and ingesting it into DSEP. This is done when the deposit library decides to replace the current reference platform by a new one or to add a new current platform configuration. The reference platform

is specified by its technical metadata and is uniquely identified. The system software and configuration data of the new reference platform is packaged in a new SIP, together with the technical metadata.
Viewer programs may be ingested as separate SIPs or as part of the reference platform. The image of the installed reference platform may also be stored in a separate SIP.
When such a new reference platform has been created, all new deposit publications that run on it carry a reference to this reference platform in their technical metadata.
When parts of the reference platform are in danger of becoming obsolete, emulation strategies can be applied by creating new versions of the reference platform that include emulator specifications and/or software. To this end the process may also provide a 'Reference Platform Update' mechanism.

Data Flow
Input: Reference Platform Definition [Preservation Planning]
Output: SIP [Ingest]



Sub-process Archival Information Update

Figure 57 Process Archival information Update

The process Archival Information Update is responsible for upgrading AIPs that are already in Archival Storage. This is done when the deposit library decides to replace the current packaging design by a new one (Repackaging process) or when publication data formats are in danger of becoming obsolete and the library decides to convert them to newer data formats (Transformation process).
The process Archival Information Update requests (a collection of) DIPs from Access, unpacks them, performs changes, records the changes made as preservation metadata, adapts the authenticity information if necessary, and packages them into new SIPs. The SIPs are ingested into DSEP.

Data Flow
Input: Packaging Definition [Preservation Planning]
Input: DIP [Access]
Output: SIP [Ingest]

ANNEX 1 The deposit library workflow

1 Selection The deposit library defines its own selection criteria concerning electronic publications.

2 Acquisition The electronic publication may be ordered from the publisher or automatically delivered.

3 Capture This is the electronic transfer of a copy of the electronic publication from the publisher's production/distribution system to the library's deposit system.

4 Registration The new incoming electronic publication is checked in the deposit system.

5 Verification Control routines are performed on the new incoming electronic publications to assess its physical and logical integrity.

6 Description A title entry is created in the library catalogue to ensure the new publication can be found in the library search systems.

7 Storage The new publication is stored in the digital store of the deposit system.

8 Preservation Preservation strategies are carried out to ensure the publication remains useable.

9 Delivery The publication is retrieved from the digital store and made fit for use on the library visitor's workstation.

10 Access The library user access environment enables use of the publication, according to well-defined access conditions.

11 Monitoring The whole workflow for handling electronic publications is monitored and quality-controlled.

ANNEX 2 The DSEP process tree

DELIVERY&CAPTURE
Receive Publication
Quality Assurance
Generate Metadata
    Generate Technical Metadata
    Generate Structural Metadata
    Generate Access Control Metadata
    Generate Bibliographic Metadata
Update External Systems
Create SIP
Submit SIP

INGEST
Receive SIP
Validate SIP
Generate AIP
Coordinate Updates

ARCHIVAL STORAGE
Receive AIP
Manage Storage
Error Checking
Migrate Media
Disaster Recovery
Provide AIP

DATA MANAGEMENT
Receive Database Updates
Administer Database
Perform Queries
Generate Report

ACCESS
Coordinate Access Activities
Generate DIP

PACKAGING&DELIVERY
Coordinate Requests
Validate DIP
Deliver Publication

ADMINISTRATION
Manage System Operation
Manage System Configuration
Manage Reference Platforms
Establish Standards&Strategies
Inform Management

PRESERVATION
Preservation Planning
    Develop Preservation Standards&Strategies
    Develop Packaging Designs
    Define Reference Platforms
    Monitor Technology
Preservation Activities
    Create New Reference Platform
    Archival Information Update

Glossary

AIP
Archival Information Package (from OAIS-RM).

CD-ROM Publication
An electronic publication originally distributed on CD-ROM. Not to be confused with offline batch deposits.

CIP
Cataloguing in Publication

Delivery&Capture
Pre-processing interface responsible for receiving deposit publications from the publishers and for submitting them to the DSEP system as a standard submission package.

Deposit transfer area
Temporary storage space for clearing incoming deposit publications.

Digital Library System
The ICT-infrastructure and automated library systems that support library business processes, in particular the handling of digital library collections.

DIP
Dissemination Information Package (from OAIS-RM).

DLS
Digital Library System

DSEP
Deposit System for Electronic Publications

EP
Electronic Publication

Electronic publication definition file
The electronic publication definition file specifies the component parts (e.g. metadata files, data files, software files, etc.) and the logical units (table of contents, part/whole relationships) of the publication. This file permits DSEP to perform logical integrity checks.

Logical publication unit
Electronic publications are defined more in terms of logical and structural boundaries then by physical carrier boundaries. A logical publication unit defines a given granularity level of the electronic publications and is highly dependent on the type of publication (e-journal, web-publication, database-like publication, etc.). This concept is similar to the OAIS concept of Archival Information Unit (AIU).

National Bibliographic Number (NBN)
Unique identifier associated with a deposited publication and its bibliographic description in the National Bibliography. The NBN is issued by a National Bibliographic Agency.

OAIS-RM
Open Archivals Information System Reference Model

Offline batch deposits
Batch-like deposit of online publications on a physical carrier such as CD-ROM.

OPAC
Online Public Access Catalogue

Preservation Layer Model (PLM)
This is a model of the different layers of Representation Information, based on the layered Information Model from Annex E of the OAIS-RM and adapted by NEDLIB [NEDLIB Report 2]. The PLM can be used as a tool to keep track of all software and hardware dependencies of electronic publications.

Reference platform
Standard computer system configuration designated by the deposit library for installing and running electronic publications. The Reference Platform is suited for most publications appearing on the consumer market, during a given period of time. Example.

Representation Information Repositories
Repositories containing a common pool of Representation Information, shared by a variety of deposit libraries and other memory organisations. These Repositories provide open standard format specifications, common viewing software, conversion software, interpreters, emulators, etc.

SIP
Submission Information Package (from OAIS-RM)

UBC
Universal Bibliographic Control