Supporting the Analysis and Audit of Collaborative OAIS's Using an Outer OAIS-Inner OAIS (OO-IO) Model

Eld Zierau
Department of Digital Preservation
The Royal Library of Denmark
Søren Kierkegaards Plads 1
DK-1016 København K
ph. +45 91324690
elzi@kb.dk

Nancy Y McGovern
Curation and Preservation Services
MIT Libraries
77 Massachusetts Avenue
Cambridge MA 02139 USA
ph. +1 617 253 5664
nancymcg@mit.edu

ABSTRACT
This paper addresses the question: What would distributed digital preservation look like using the OAIS Reference Model? The challenge is the need for several organizations to cooperate to achieve distributed digital preservation; using replication, independence, and coordination to address the known threats to digital content through time. The main purpose of the paper is to present an *Outer OAIS-Inner OAIS* (OO-IO) Model that can support the analysis and audit of collaborative interactions between multiple OAIS's to enable distributed digital preservation. The paper provides extensive explanations and diagrams to demonstrate the ability of the OO-IO model to address distributed digital preservation conformance with the Open Archival Information System (OAIS) Reference Model. It is argued that the OO-IO model contributes a necessary extension to the literature of the digital preservation community to address the analysis and audit necessary for distributed digital preservation.

General Terms
Infrastructure, communities, preservation strategies and workflows, theory of digital preservation, case studies and best practice.

Keywords
OAIS Reference Model, Distributed Digital Preservation, Standards, Audits, Analysis, Collaboration.

1. INTRODUCTION
Digital preservation is the "active management of digital content over time to ensure ongoing access."[1] As good practice for digital preservation matures, organizations are naturally addressing more advanced strategic and operational aspects of the technology required to sustainable digital preservation program leading to distributed digital preservation.

Distributed digital preservation, a focus of this paper, is here defined as "the use of replication, independence, and coordination to address the known threats to digital content through time to ensure their accessibility" ([9] p. 78)[2]. Distributed digital preservation is a form of advanced digital preservation practice, which can be described as in the model for the development of a digital preservation program [4][3]. Here the most advanced stage in that model, externalize, is characterized by collaboration to achieve objectives. In general, it is common in distributed digital preservation for organizations to establish strategic collaborations to meet preservation.

The Open Archival Information System (OAIS) Reference Model is important for digital preservation and is the foundation for the *Outer OAIS-Inner OAIS* (OO-IO) model presented in this paper. The OAIS Reference Model is a core standard in good practice for digital preservation that was approved by the International Standards Organization (ISO) in 2003 and revised in 2012 [1].

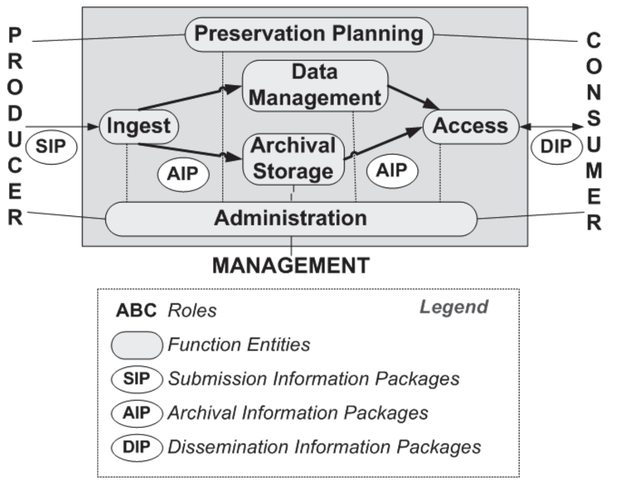

Figure 1. OAIS model

The functional entities and information packages in OAIS Reference Model are depicted in Figure 1, corresponding to

[1] There is no single, authorized definition of digital preservation. The authors cite this definition from the Library of Congress because it is succinct, effective, and often cited in the literature. Available at: http://www.digitalpreservation.gov/about/

[2] Definition is from the *Framework for Applying OAIS to Distributed Digital Preservation* mentioned later in this paper.

[3] The model is discussed in the first chapter of the *Aligning National Approaches to Digital Preservation* (ANADP) volume

Figure 4-1 in the OAIS Reference Model [1]. This will be referred to as a simple OAIS throughout the paper, by contrast to the complexities of the Inner and Outer instances of OASI the paper addresses. OAIS functional entities, functions and information packages will be written in Italic font in this paper.

The OAIS Reference Model provides a framework that has proven effective in guiding the development of sustainable digital preservation programs. "An OAIS is an Archive, consisting of an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community" (from [1] Section 1-2). References to the term organization in this paper are informed by this definition of an OAIS from the OAIS Reference Model document.

Although the OAIS Reference Model does briefly discuss interoperability for distributed digital preservation in section 6 [1], it needs to be more explicit in order to be usable for analysis and auditing purposes. The OO-IO model can support the analysis and audit of collaborative arrangements between multiple OAIS's, where this paper uses OAIS's as the plural form of an OAIS.

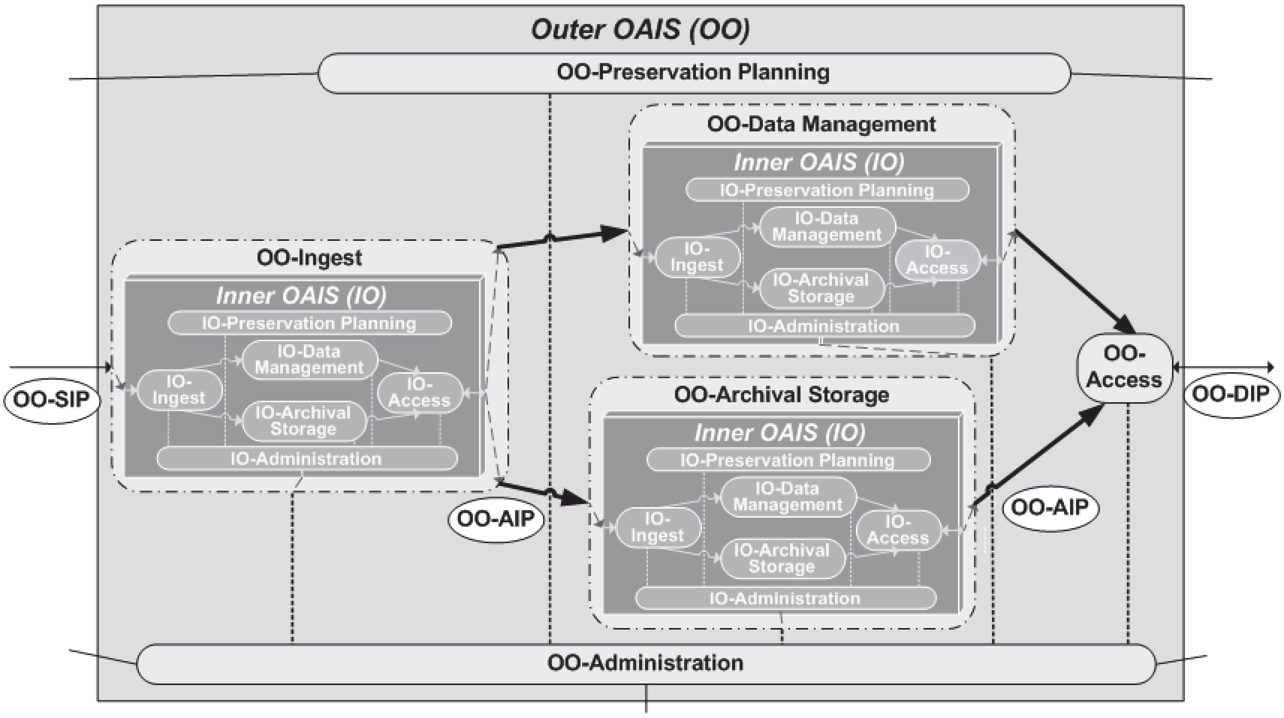

Figure 2. OO-IO model

The OO-IO model is depicted in Figure 2. Explanation of this model will follow later in this paper, as well as how this OO-IO model is building on a previous model (the IR-BR model [8][4]) and the work carried out in the international working group the *Framework for Applying OAIS to Distributed Digital Preservation (DDP)* [9][5],

The *Archival Storage* functional entity of OAIS was the starting point for developing the OO-IO model, just as storage partnerships have been a common starting point for distributed digital preservation. A core requirement of digital preservation is to maintain multiple, geographically-distributed copies of digital content. Meeting that requirement provides a natural opening for storage partnerships and services. The challenge is that the *Archival Storage* needs to be viewed as a distinct OAIS/OAIS's (the *inner OAIS* as the *Archival Storage* is within a separate *outer OAIS*). The reason is that the separated collaboration around *Archival Storage* will need portions of all OAIS functional entities, for example Preservation Planning for media migrations.

Means to support argumentation for conformance to OAIS are needed for distributed digital preservation solutions, which is where the OO-IO model can assist. A decade ago, the majority of organizations in the digital preservation community were focused on determining what it meant to conform to the OAIS Reference Model. The community now includes a growing number of organizations that are engaged in distributed digital preservation. Those organizations have a need to demonstrate conformance with standards through good practice, also for distributed digital preservation.

Section 2 of this paper, as background for the discussion, provides a brief history of the OO-IO model, and places the model into the context of standards and practice for digital preservation, noting developments that informed or led the need for this supplement to further address interoperability in the OAIS reference model. Section 3 explains and illustrates the components of the OO-IO model, and demonstrates the OO-IO model's conformance with a simple implementation of OAIS. Section 4 describes how the OO-

[4] The IR-BR model originates from the pre-study of the Danish BitRepository.org

[5] Preliminary information about the DDP Framework is available at www.metaarchive.org/ddp/index.php/Main_Page.

IO model can support the documentation and audit of collaborative OAIS's.

2. CONTEXT AND NEED

The emergence of good practice for distributed digital preservation that this paper addresses is grounded in the overall context of the development and promulgation of standards and practice for digital preservation. This section traces the development of relevant standards and practice to demonstrate the community-based need for the OO-IO model, as well as the activities that led to its development.

2.1 Standards and Practice

The OO-IO model contributes to the existing foundation of community standards and practice for digital preservation. The model can be used to demonstrate how the complexities of distributed digital preservation use cases can be specified and implemented.

Though digital content has been preserved by some organizations since the 1960s, the digital preservation community traces its roots to the seminal 1996 *Preserving Digital Information report* [7] that defined the problem of digital preservation, specified the challenges that organizations face in managing digital content across generations of technology, considered relevant roles and responsibilities for digital preservation, and framed a set of recommendations to guide the establishment of good practice.

There are several noteworthy things about the 1996 report. The authors of the report represented the domains of the community – libraries, archives, museums, and others – from multiple countries, a rare occurrence at the time, if not a first for the community. The report specified features of digital objects that need to be addressed to ensure the objects' integrity: content, fixity, reference, provenance, and context. In addition, the 1996 report specified the need for the certification of digital repositories that manage digital content to demonstrate good practice and called for "fully distributed storage"[6] of digital objects, a reference to the current challenges of distributed digital preservation. In the nearly twenty years since the *Preserving Digital Information* report was published, a growing set of standards and practice has emerged, as discussed in this section, that provides a frame for good digital preservation practice.

In 1995, the Consultative Committee for Space Data Systems (CCSDS) established the work package that led to the OAIS Reference Model. The OAIS Reference Model references the integrity features, the need for audit and certification of preservation repositories, and the requirement for distributed storage that were specified in the 1996 report [7]. OAIS was approved as an international standard in 2003 and updated in 2012. Most organizations that are responsible for the long-term preservation of digital content refer to and in many cases build and implement their repository systems to align with the principles and concepts of OAIS. These activities demonstrate that OAIS is being maintained and is in use within a significant portion of the community, two measures of success for standards that have a demonstrated impact on practice.

The OAIS Reference Model is not a standalone standard, but the anchor for a family of OAIS-related standards. One of the characteristics that has enabled OAIS to endure is the standards roadmap that has been included since the early drafts of the document that addresses the ways in which OAIS needs to be extended and applied.[7] Examples of standards that are called for in the OAIS standards road map and that are cited the 2012 update of the Reference Model include: the Producer-Archive Interface Method Abstract Standard (PAIMAS [3]); and preservation metadata, e.g., Preservation Metadata Implementation Strategies (PREMIS [5]).

Another standard in the OAIS family addresses the need for audit and certification to enable digital repositories to demonstrate good practice. Audit and Certification of Trustworthy Digital Repositories (ISO-16363 [2]) is a standard that was built on the Trustworthy Repository Audit and Certification (TRAC) requirements [6]. The audit and certification requirements for digital preservation stipulate that organizations provide evidence to demonstrate how they conform to the ISO 16363 requirements.

Demonstrating good practice for distributed digital preservation is complicated by the need to accumulate and consolidate evidence across collaborating OAIS's. The OO-IO model supports audit requirements within distributed digital preservation environments by elaborating the relationships and roles of functional entities and their functions within and between relevant OAIS's.

Section 6 of the OAIS Reference Model serves as a reference point for the OO-IO model within the current framework of digital preservation standards, That discussion in OAIS considers issues pertaining to interoperability between archives and levels of interaction between OAIS Archives (Section 6-1) and Management issues with federated archives (Section 6-2). This portion of OAIS acknowledges the need for interoperability in digital preservation, but the discussion is not extensive and does not specify an approach for achieving interoperability. Practitioners of distributed digital preservation are developing an understanding of how interoperability can be realized [9]. The methodology of the OO-IO model, informed by that deepening understanding, involved systematic analyses of common use cases for distributed digital preservation that are described in Section 2.2 and elaborated in Section 3.

2.2 Provenance of the OO-IO model

The development of the OO-IO model was initially motivated by the complexities of good practice for distributed digital preservation that were identified by organizations that have become engaged in distributed digital preservation. In practice, distributed digital preservation involves a range of use cases to address the specifics of interoperability between multiple OAIS's.

It was an investigation of such complexities that led to the development of the Institution Repository–Bit repository (IR-BR) model [8], the starting point for the OO-IO model. The IR-BR model emerged during work on the open source BitRepository.org framework that is used for bit preservation in Danish Cultural institution. Later in this paper, the correlations between the IR-BR model and the *Archival Storage* component of the OO-IO model are explained. In the IR-BR model, the Institution Repository is an Outer OAIS as an organization using a Bit Repository that is an Inner OAIS.

[6] Citation from the 1996 Report [7].

[7] OAIS section 1.5.

The IR-BR model informed and influenced the development of the *Framework for Applying OAIS to Distributed Digital Preservation (DDP)* [9][8], a result of a project established to address the growing awareness of the need to adapt and extend current standards to address distributed digital preservation, models and auditing methodologies to support DDP. The DDP Framework addresses the roles, functions, and use cases that build a layer upon section 6 of the OAIS Reference Model to begin to specify how interoperability and federation might work. The DDP Framework has been developed by a working group with representatives from both North America and Europe that included the authors of this paper and representatives from some major DDP examples, including MetaArchive, the Danish BitRepositorty.org, Chronopolis, Data-PASS, DuraCloud, Internet Archive, UC3 Merritt and Archivematica. Variations within this range of cases pointed to the need to focus on other OAIS functional entities that require distribution over more organizations requiring a generalization of the IR-BR model into the OO-IO model. The results of the DDP Framework project will be shared when available.

Developing the OO-IO model provided the means to analyze the functionality of an inner OAIS and provided common terminology between inner and outer OAIS's. The generalization in the OO-IO model applies not only to the *Archival Storage* functional entity that can be seen as a separate Inner OAIS, but also *Data Management* and *Ingest*. The following section explains and demonstrates the feasibility and validity of this generalization.

3. THE OUTER OAIS-INNER OAIS MODEL

The primary purpose of the Outer OAIS–Inner OAIS (OO-IO) model is to simplify the challenges – organizational (what needs to be done) and technological (how it can be done) - of engaging in distributed digital preservation that involves several organizations. An Outer OAIS refers to an entire OAIS implementation – a simple OAIS – that supports distributed digital preservation, including all of its Inner OAIS's. An Inner OAIS is an OAIS that is distinct from the Outer OAIS and is implemented to manage one OAIS functional entity - *Ingest*, *Data Management* or *Archival Storage*. Each inner OAIS is managed as a complete OAIS, though it is dedicated to managing a single functional entity in the Outer OAIS, as depicted in Figure 2. One example of a case that requires the OO-IO model rather than a simple OAIS is when the functional entity (e.g., *Archival Storage* as a bit repository) is separated and managed by one or more external organizations (OAIS's), as is often the case in distributed digital preservation.

Note that the sample *Inner OAIS* cases that have been specified in this paper for the OO-IO model (i.e. *Archival Storage*, *Ingest*, and *Data Management*) focus on functional entities that require storage because an inner OAIS without storage would not be necessary. The functional entities that require storage are those that interact directly with *SIPs*, *AIPs* and *DIPs*. These information packages are pictured in Figure 1. That figure illustrates Submission Information Packages (*SIPs*) being received via the functional entity *Ingest*. The *Ingest* functional entity then creates Archival Information Packages (*AIPs*) and the related data management information that are parsed to the *Archival Storage* and *Data Management* functional entities, respectively. In response to an *Access* request for Dissemination Information Packages (*DIPs*), the *AIPs* and related data management information required to create the *DIP* are delivered via the *Access* functional entity. There is no sample case for the *Access* functional entity in the OO-IO model because *Access* generates/re-generates *DIPs* based on information received the *Data Management* and the *Archival Storage* functional entities. Thus there are no obvious cases for risk of loss or need of cooperation in relation to the *Access* functional entity

The 'archive interoperability' discussion in Chapter 6 of the OAIS Reference Model states that an OAIS may be geographically distributed. It lists possibilities of all components being under the same *Management*, or spread over OAIS Archives with separate *Managements* that work cooperatively. The OO-IO model builds upon Chapter 6 and in doing so, the elaboration of the OO-IO model aligns with the existing OAIS reference model. The OO-IO model specifies an approach for using the OAIS model to achieve archive interoperability that Chapter 6 does not provide.

A strength of the OO-IO model is that the analysis required to develop the model demonstrates the need for the parsing of OO functional entities into OO and IO functional entities, as does the analysis of the interface between the OO and the IO. The systematic process for developing the OO-IO model identified the prefix for terms (OO or IO), making clear distinctions between inner or outer OAIS functions and information. This specification verifies that the OO-IO model conforms to the OAIS Reference Model. Therefore, the inner OAIS scenario is detailed to demonstrate the case for *Archival Storage*, *Ingest* and *Data Management* in the below sections.

3.1 The OO-IO *Archival Storage* Component

For distributed digital preservation, one use case for the *Archival Storage* component of the OO-IO model is the need to operate a separate standalone bit repository to meet the requirements of the *Archival Storage* functional entity of the *Outer OAIS*. The standalone bit repository itself is managed as an *Inner OAIS* and incorporates some of all of the functional entities of an OAIS.

The *Archival Storage* component of the OO-IO model addresses only the *Archival Storage* functional entity of the OAIS Reference Model. It is an inner OAIS as depicted in Figure 3.

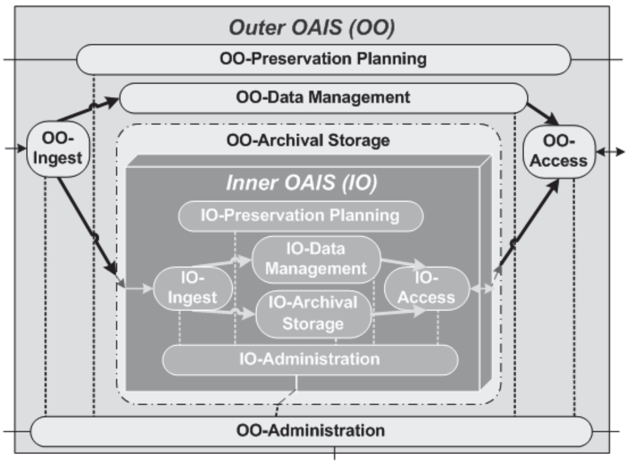

Figure 3. Archival Storage in the OO-IO model

[8] Preliminary information about the DDP Framework is available at www.metaarchive.org/ddp/index.php/Main_Page.

Distributed digital preservation implementations of OAIS require more OAIS functions, e.g., *Ingest* including the *Receive Submission* function, in addition to *Archival Storage* functions (depicted and described later in Figure 4[9]). The bit repository must be treated as an Inner OAIS where parts of *all* of the OAIS functional entities are required by the inner OAIS.

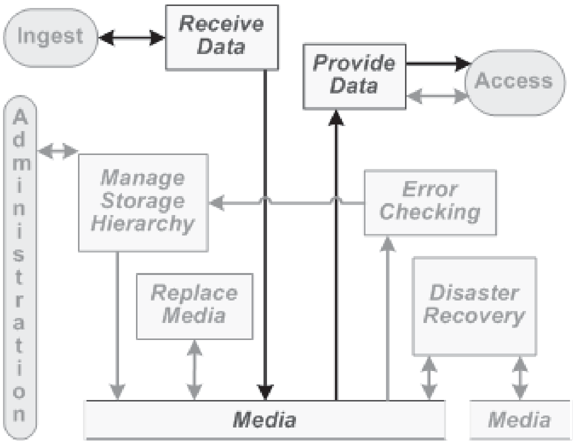

Figure 4. Functions of OO-IO *Archival Storage*

3.1.1 Flow for the Archival Storage component

In developing the OO-IO model, an investigation of the flow of information from *Ingest* to *Access* of the Outer OAIS confirmed the validity and utility of the *Archival Storage* component of the OO-IO model.

Figure 5 illustrates a high-level flow of the information packages with focus on the *Archival Storage* component of the OO-IO model. The dotted lines in the figure indicate that there are more functions involved in the path.

In this flow, an Outer OAIS *Submission Information Package* (*OO-SIP)*[10] is received from an *OO-PRODUCER*[11] and passed to the *OO-Ingest* functional entity. All of the internal *OO-Ingest* functions are executed, resulting in the transformation of *OO-SIPs* to *OO-AIPs*. The difference for Archival Storage in the OO-IO model occurs during the transfer of an Outer OAIS *Archival Information Package* (*OO-AIP)* to the *OO-Archival Storage.*

When the *OO-AIP* is transferred to *OO-Archival Storage*, it takes an alternative path from a simple OAIS implementation when it is ingested into the Inner OAIS within the *OO-Archival Storage*. Thus an *OO-AIP* becomes an *IO-SIP* and runs via the *IO-Ingest* functions before it is transformed into an *IO-AIP* in the *IO-Archival Storage*. Likewise, the receipt/storage confirmation for accepted data and completed storage is returned to the Outer OAIS from the *IO-Ingest*, as the inner OAIS acts as an OAIS.[12]

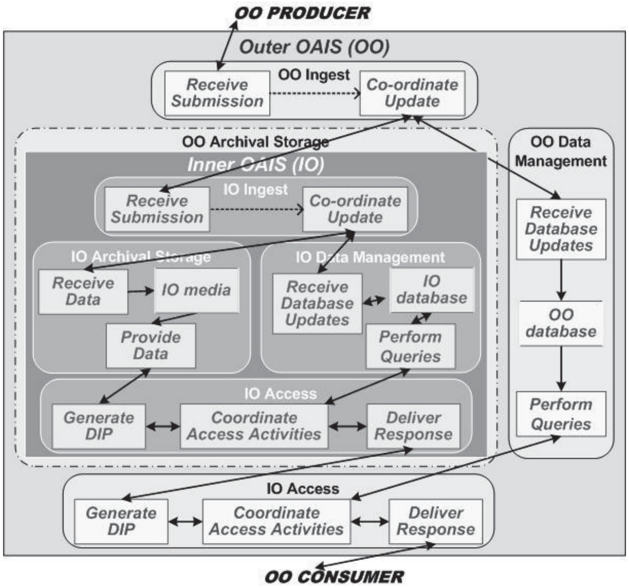

Figure 5. Information path for *Archival Storage* component

3.2 The OO-IO *Ingest* Component

The *Ingest* component of the OO-IO model addresses only the *Ingest* functional entity of OAIS, It is an Inner OAIS as depicted in Figure 6.

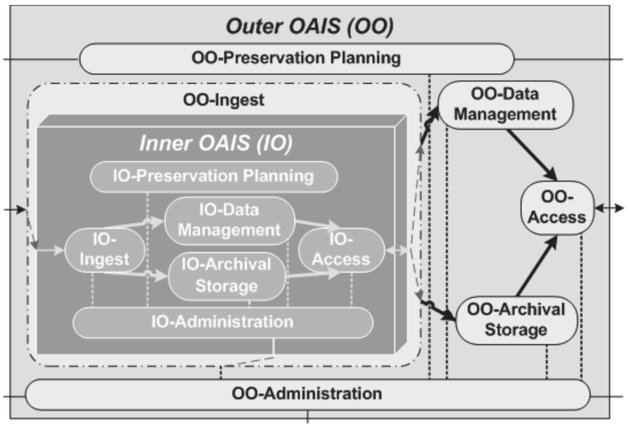

Figure 6. *Ingest* component of the OO-IO model

There are two use cases that demonstrate the need for Ingest as an Inner OAIS: distributed ingest and delayed processing in ingest, as discussed below.

Distributed ingest is a scenario that was identified in several cases developed for the DDP project. In particular, micro-service-based solutions like UC3-Merritt and Archivematica had examples of

[9] From the OAIS Reference Model Figure 4-3: Functions of the Archival Storage Functional Entity [1].

[10] Submission Information Package – see Figure 1.

[11] OAIS roles appear in all capitals in this paper.

[12] A similar argument can be made for the Access component. The full explanation and supporting justification can be found in the “Cross Institutional Cooperation on a Shared Bit Repository” [8]. This also includes examples of a possible split of OO-Data Management and IO-Data Management, OO-Administration and IO-Administration and OO-Preservation Planning and IO-Preservation Planning.

using the distribution of micro-services to manage many simultaneous loads of ingest processing.[13]

Preliminary archiving of *SIPs*, is a scenario where *SIPs* are secured in order to mitigate risk of losing them due to risk of loss caused by delays in the ingest process. The Royal Library of Denmark has focused on this as a result of a general risk analysis completed for the digital preservation program at the library. One of the reasons for a delay in archiving can be that it takes time to get all the information needed to generate *AIPs*. For instance a computer game may need trailers and digitized user guides before it can be archived. Another reason can be that large digitization project may have interdependent data that needs to be connected before archiving can proceed. A third reason can be that digital material can require a lot of work before becoming an *AIP*, for instance hard drives from deceased authors, which must be analyzed and restructured before becoming an *AIP*.

Like *Archival Storage*, the *Ingest* component of the OO-IO model is an Inner OAIS that functions as a separate ingest mechanism, including its own *Archival Storage*, within the *OO-Ingest*. The Inner OAIS must also include portions of all of the OAIS functional entities, not only Ingest.

3.2.1 Information flow for the Ingest component

In developing the OO-IO model, an investigation of the flow of information from *Ingest* to *Access* of the Outer OAIS demonstrated the validity and utility of the *Ingest* use case of the OO-IO model.

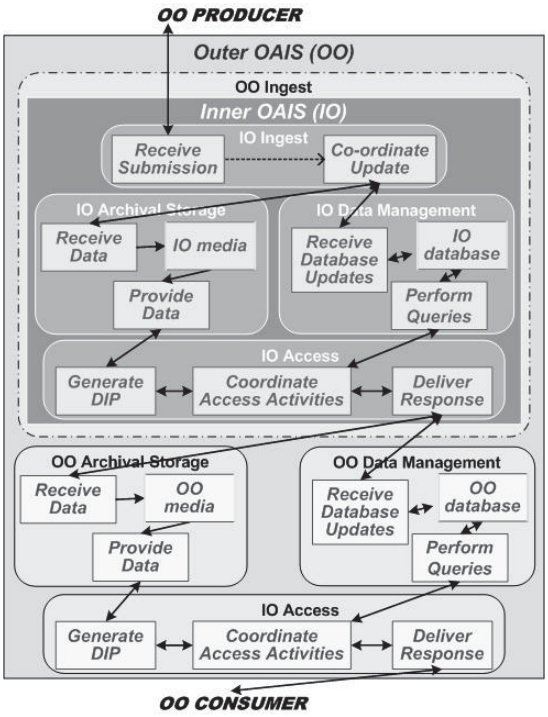

Figure 7. Information path for *Ingest* component

Figure 7 illustrates the information flow for *Ingest*, in the same way that Figure 5 did for *Archival Storage*. The flow is somewhat more complex for *Ingest* because this functional entity delivers information to both *Archival Storage* and *Data Management*.

In this flow, an *OO-SIP* is received from an *OO-PRODUCER* and passed to the *OO-Ingest* functional entity. Already there is a change because the *OO-SIP* takes alternate path by being ingested to the *IO* instead of being passed to *OO-Ingest* functions. An *OO-SIP* becomes an *IO-SIP* and runs via the *IO-Ingest* functions becoming an *IO-AIP*, which are secured in *IO-Archival Storage*. A closer look at the *Ingest* functions in Figure 8 makes this clearer.

The *Ingest* component of the OO-IO is also more complex than the *Archival Storage* component of the OO-IO because the *Ingest* component is ***not*** only a question of the information taking another path before reaching a destination (like the *OO-AIP* taking another path before reaching the *IO-Archival Storage)*. In the *Ingest* component, it is only the *IO-SIP*/*OO-SIP* that is sent to *IO-Archival Storage*, which means that *Ingest* functions corresponding to *OO-generate AIP* and *OO-Coordinate Update* are not expressed in the IO as they are in the OO. This means that the *OO-Ingest* must generate an *OO-AIP* and coordinate updates, while only the *OO-SIP* (possibly with a minimum of metadata) is secured in the *IO-Archival Storage*. The functions performed within the *IO-Access* functional entity are to generate the *OO-AIP* and coordinate updates.

Viewed from this perspective, it makes perfect sense that the *IO-DIP* can be associated with an *OO-AIP*, since a *DIP* is derived from an AIP to fit the request from a *CONSUMER*. For the *Ingest* component of the OO-IO model, an *IO-CONSUMER* (the *OO-Archival Storage/OO-Data Management*) gets an *IO-DIP* (or rather an *OO-AIP)* that is derived information from an *IO-AIP*. The *IO-Ingest* takes the *OO-SIPs* as *IO-SIPs* and transfers them to the *IO-Archival Storage* without transformations (although some minimum information may be added). This makes the *IO-AIP* equivalent (or very similar) to the *IO-SIP/OO-SIP*. This requires the *IO-Access* functional entity to operate like the *OO-generate AIP* and *OO-Coordinate Update* of an ordinary *OO-Ingest* functional entity.

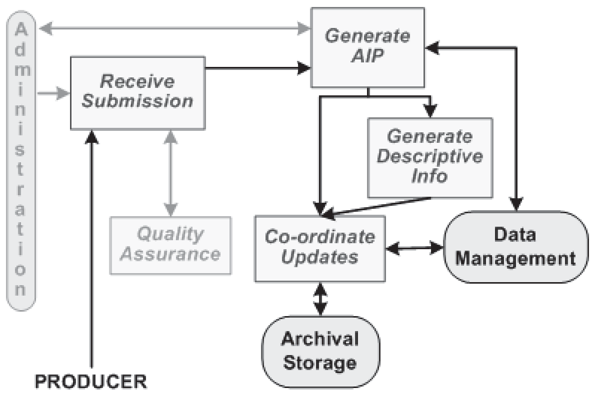

Figure 8. Functions of the *Ingest* functional entity

To further explain this portion of the analysis, the functions of the *IO-Access* functional entity are depicted in Figure 9.

[13] See UC3-Merritt at: https://merritt.cdlib.org/ and Archivematica at: https://www.archivematica.org/wiki/Main_Page.

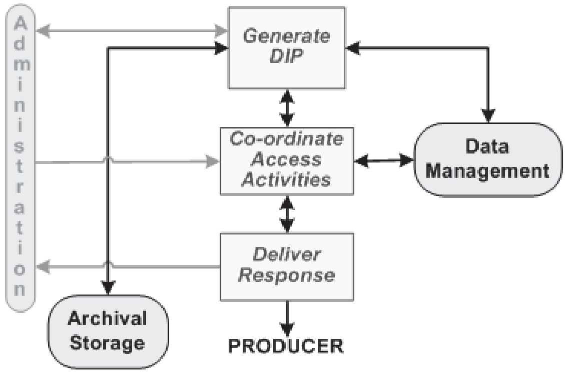

Figure 9. Functions of the *Access* functional entity

The *Ingest* functions that should correspond to *OO-Generate AIP* (and *OO-Generate Descriptive Info*) need to be included in the *Generate DIP function* of the *IO-Access* functional entity. As the redrawing[14] of the functions in Figure 8 and 9 show, there is similar flow through functions with similar names and meaning:

- the *PRODUCER* in Figure 8 matches the *Archival Storage* of Figure 9,
- the *Generate AIP* (together with *Generate Descriptive Info*) in Figure 8 matches *Generate DIP* of Figure 9,
- the *Archival Storage* in Figure 8 matches the *CONSUMER* of Figure 9, and
- the *Coordinate Updates* Figure 8 aligns with *coordinate Access Activities* of Figure 9.

In practice, it will be important to pay close attention to how this portion of the expected *OO-Ingest* functions map to these *IO-Access* functions.

For *Preservation Planning* in the *Ingest* component of the *OO-IO* model, the *IO-Preservation Planning* is only concerned with security of the ingested *IO-SIP* (corresponding to the *OO-SIP*). This means that *Preservation Planning* is split between the *IO* and the *OO-Preservation Planning* where *OO-Preservation Planning* covers all other *Preservation Planning* for the *OO-SIPs*. This may require coordination as in the example of the *Archival Storage* component of the OO-IO model.

For *MANAGEMENT* in the *Ingest* component of the OO-IO model, it is generally true – as it was for the *Archival Storage* component of the OO model – that requirements resulting in directions from *OO-MANAGEMENT* are dealt with by *IO-Administration*. From the *IO* perspective, the *OO-Administration* represents *IO-MANAGEMENT*. It is at the interface between *OO-Administration* and *IO-Administration* that the mapping of the requirements for the *IO* takes place.

For *Data Management* in the *Ingest* component of the OO-IO model, there may be specific *IO-Data Management* actions that are only relevant to the *IO*, but there will most likely also be elements of *IO-Data Management* that must be passed to the *OO-Data Management*. Examples include catalogs, inventories and audit trails. This portion of the OO-IO model works well as the *IO-DIP* (that becomes the *OO-AIP*, updating *OO-Archival Storage* and *OO-Data Management*) is generated from *IO-Archival Storage* as well as from the *IO-Data Management.*

3.3 The OO-IO *Data Management* component

The *Data Management* component of the OO-IO model addresses only the *Data Management* functional entity of the OAIS Reference Model. It is an Inner OAIS as depicted in Figure 10.

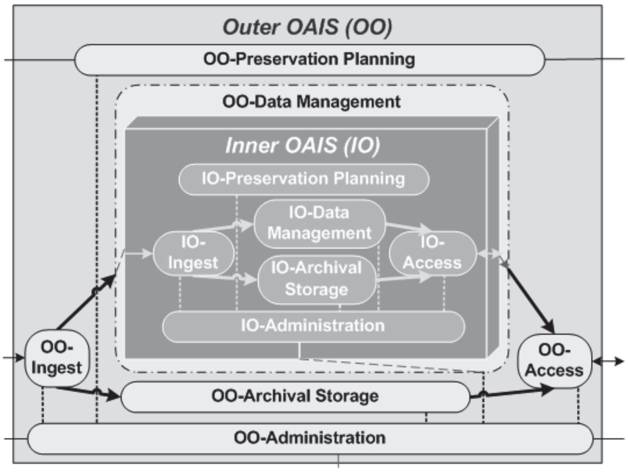

Figure 10. *Data Management* component of OO-IO model

There may be different use case scenarios, where it can make sense to have *Data Management* as an inner OAIS. The following scenarios are just examples:

- *Separate securing of data.* A situation that addresses multiple instances of content that has security requirements.
- *Distributed linked data representing database* A case where a database is represented by linked data that is distributed across multiple environments.

Separate secure data is a scenario similar to the one for *Ingest*. There may be portions of *Data Management* data that need to be secured for distributed digital preservation. This can occur when there is a need for an asynchronous update of *Data Management* in connection with ingests of digital materials or for the ongoing creation of collection information that later may be needed for preservation.

Distributed linked data represented in the database is a scenario where linked data are represented in the *Data Management*, implying that *Data Management* is distributed. This case could require descriptive elements from databases that are managed by different organizations. This is especially relevant for representation information, if for instance:

- One organization has descriptions of its preserved assets
- Another organization has the format registry used for the preserved assets
- A third organization has the environment registry used for the preserved assets

Distributed knowledge-bases like registries are individually maintained, and it make sense to have separate mechanisms for

[14] Simplification of duplicated arrows and moving the entities, functions and roles around compared to the illustrations in the OAIS Reference Model

preservation planning, e.g., policies for maintenance of registries that are used as shared resources.

Like *Ingest*, the *Data Management* component of the OO-IO model is an Inner OAIS that functions as a separate data management mechanism, including its own *Archival Storage* as well as portions of all the functional entities of an OAIS.

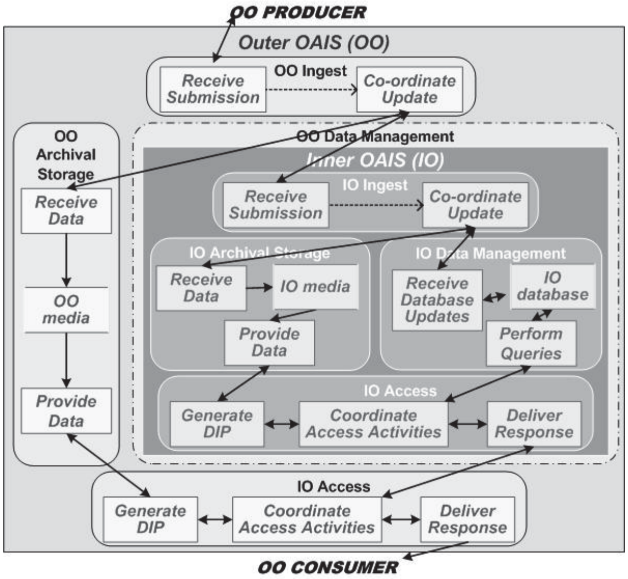

Figure 11. Flow for *Data Management*

3.3.1Flow for the Data Management component

In developing the OO-IO model, an investigation of the flow of information from *Ingest* to *Access* of the Outer OAIS validated the *Data Management* component of the OO-IO model. Figure 11 illustrates the information flow, as Figure 5 did for *Archival Storage*.

In the *Data Management* flow, an *OO-SIP* is received from an *OO-PRODUCER* and passed to the *OO-Ingest* functional entity. All the internal *OO-Ingest* functions are executed, transforming *OO-SIPs* to *OO-AIPs* and belonging to *OO-Data Management* information. The change occurs when *OO-Data Management* information, e.g., reports and update information, is transferred to the *OO-Data Management* because it takes an alternate path, being ingested into the Inner OAIS instead of being passed to *OO-Data Management* functions. This data management information becomes an *IO-SIP* and runs via the *IO-Ingest* functions before it ends as an *IO-AIP* in the *IO-Archival Storage.* Here, the *IO-Archival Storage* containing the *IO-AIP* (*OO-Data Management* information) may be seen as equivalent to the *OO-Data Management* database. The *IO-Access* acts as the *Perform Queries* function of the *OO-Data Management* functional entity.

Like the *Ingest* component of the OO-IO, the *Data Management* component of the OO-IO model is more complex than the *Archival Storage* component. Similarly, this is because the *OO-Data Management* functions – in a simple OAIS implementation - are not just taken over by the *IO-Data Management* functions, but have to be interpreted in terms of other IO-functional entities and functions. However, the *Data Management* component is simpler than the *Ingest* component because the *Ingest* and *Access* information for the functions of *Data Management* are more similar to OAIS, than the *Ingest* and *Access* information for *Ingest*.

Administration functions are managed within the OO and the IO. However, *OO-Administration* report requests from *OO-Data Management* can be regarded as either a report request to the IO from *IO-MANAGEMENT* or from *IO-Access* (if *IO-DIPs* are considered to be a report result). It may also be a mix of these depending on the type of reports requested.

As with the *Ingest* use case, *Preservation Planning* for the *Data Management* use case is split between the OO and the IO. *OO-Preservation Planning*, among other things, covers the function *Develop Preservation Strategies and Standards.* For the *Data Management* component of the OO-IO model, the split follows the split of responsibilities. For example, a format registry in the IO will include the *Develop Preservation Strategies and Standards* function for this registry, while other functions of *Preservation Planning*, like the *Monitor Technology* function that addresses issues like media for *Archival Storage*, are included in the OO.

Also for the *Data Management* component of the OO-IO model, all requirements resulting in directions from *OO-MANAGEMENT* are dealt with by *IO-Administration*, as described in the end of the previous section on the OO-IO *Ingest* component.

4. USING THE OO-IO MODEL

There is a range of use cases for which the OO-IO model can be advantageous for distributed digital preservation.

First, the OO-IO model provides a means to explicitly express the OAIS functional entities and functions that are referred to by prefixing then with OO (for Outer OAIS) and IO (for Inner OAIS) for each component of the OO-IO model. Although this may seem trivial, the experience from the Danish use of the model is that it can improve communications. The use of Inner and Outer qualifiers for discussions that involve distributed digital preservation can avoid misunderstandings.

Second, the OO-IO model provides a basis for analysis of the interfaces between an *Outer OAIS* and an *Inner OAIS* that are essentially for understanding and implementing interoperability that is essential for distributed digital preservation. The OO-IO model diagrams make it be possible to explicitly map inputs and outputs that inform or produce required evidence for audit. In using the Archival Storage component of the OO-IO model, this analysis has proven to be extremely useful, both initially for the design and later the auditing of the Danish BitRepository.org.

Third, the OO-IO model can support and enable audit for distributed digital preservation. The development of the OO-IO model produced a generalized model that addresses distributed digital preservation and is grounded in standards and practice. Though it can and should be extended, this version of the model can provide a framework self-assessment and audit processes for distributed digital preservation.

A challenge for audit within distributed digital preservation environments is mapping responsibilities and accumulating evidence across multiple OAIS's to cumulatively demonstrate compliance with digital preservation requirements as specified in ISO 16363. The OO-IO model supports audit for distributed digital preservation:

- By allowing the paths (roles, functions, inputs, and outputs) between Outer and Inner OAIS's to be mapped,
- By providing a framework, based on that mapping, to determine which components of Inner OAIS's and Outer OAIS's address specific ISO 16363 requirements, and
- By directly supporting the completion of a gap analysis, using the ISO 16363 requirements, in preparation for an audit (peer review or external) of a distributed digital preservation environment.

Summing up the OO-IO model can support the analysis and audit of collaborative interactions between multiple OAIS's to enable distributed digital preservation. This section has highlighted the benefits of the OO-IO model to improve communication, for developing and managing Inner and Outer OAIS's, and for supporting the audit of collaborative OAIS's

5. DISCUSSION AND FURTHER WORK

A challenge, though not insurmountable, in using and applying the OO-IO model is the complexity of the cases that detail the roles, functions, interactions, and outcomes of the interoperability between and within OAIS's that is required to manage distributed digital preservation environments. Therefore, working with the OO-IO model requires a deeper familiarity with and understanding of the workings of OAIS than is required for more simple use cases and implementations.

The different components of the OO-IO model have varying degree of complexity. For example, the *Ingest* component introduces an additional complexity by defining results from the *OO-Ingest* as the result of *IO-Access* of the Inner OAIS, i.e. the product *OO-AIP* for the Outer OAIS is part of the *IO-DIP* of the *Inner OAIS*, but also the *OO-Data Management i*nformation is part of this *IO-DIP*.

The example of using the OO-IO model to support and enable audit for distributed digital preservation also highlights further work that is needed on the model to elaborate use cases that illustrate and document audit processes. The productive discussions that occurred in DDP cases while developing the DDP Framework suggest that:

- an increasing number of practitioners are interested in and need to use DDP use cases,
- DDP cases contribute timely implementation examples to the literature of the digital preservation community, and
- DDP cases provide examples that can be used for academic and continuing educational purposes.

Now that the OO-IO model and the DDP Framework have been specified and both will be shared with the community, the OO-IO model and the DDP framework would benefit from an evaluation by more DDP organizations and by the broader digital preservation community.

Although the paper makes the case that the OO-IO model only makes sense for OAIS functional entities that involve storage of information packages, use cases may emerge that indicate the need to extend the model to also address *Preservation Planning* and *Administration*. These entities do not have obvious cases, because these functional entities do not come into direct contact with information packages in an OAIS. However, these functional entities could be investigated further and added to the OO-IO model, if relevant cases arise. The same applies for the *Access* functional entity if relevant cases arise.

6. CONCLUSION

In summary, the *Outer OAIS-Inner OAIS* (OO-IO) Model is needed to support the specification and audit of collaborative interactions between multiple OAIS implementations for distributed digital preservation.

The paper has provided extensive explanations and diagrams to make evident the ability of the OO-IO model to address distributed digital preservation conformance with the OAIS Reference Model.

The need for and utility of the OO-IO model as a supplement to the literature documenting current standards and practice for digital preservation was discussed then demonstrated using a sample of use cases for distributed digital preservation.

7. ACKNOWLEDGMENTS

We would like to thank participants in the DDP Framework group, since the evolved OO-IO model is based on work done in this group. We would also like to thank Sébastien Peyrard from The National Library of France, Kari René Smith from The Massachusetts Institute of Technology, and Ulla Bøgvad Kejser from The Royal Library of Denmark for useful input and discussions.

8. REFERENCES

[1] ISO 14721:2012. 2012. *Space data and information transfer systems - Open archival information system (OAIS) - Reference*. Retrievable via http://www.iso.org/iso/catalogue_ics.

[2] ISO 16363:2012. 2012. *Space data and information transfer systems - Audit and certification of trustworthy digital repositories*. Retrievable via http://www.iso.org/iso/catalogue_ics.

[3] ISO 20652:2006. 2006. *Space data and information transfer systems -- Producer-archive interface -- Methodology abstract standard*. Retrievable via http://www.iso.org/iso/catalogue_ics.

[4] McGovern, N.Y., Skinner, K. (Editors). 2012. *Aligning National Approaches to Digital Preservation*. Publisher: Educopia Institute Publications, Atlanta, Georgia. Available at http://metaarchive.org/public/publishing/Aligning_National_Approaches_to_Digital_Preservation.pdf, retrieved March 2014.

[5] *Preservation Metadata: Implementation Strategies (PREMIS)*, hosted at http://www.loc.gov/standards/premis/. Retrieved March 2014

[6] RLG/NARA Task Force on Digital Repository Certification. 2007. *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Publisher: Chicago, CRL.

[7] Waters, D., Garrett, J. 1996. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Available at http://www.clir.org/pubs/reports/pub63watersgarrett.pdf, retrieved March 2014

[8] Zierau, E., Kejser, U.B. 2013. *Cross Institutional Cooperation on a Shared Bit Repository*, In: Journal of the World Digital Libraries, vol. 6, no. 1.

[9] Zierau, E., Schultz, M. 2013. *Creating a Framework for Applying OAIS to Distributed Digital Preservation*, In: Proceedings of the 10th International Conference on Preservation of Digital Objects, Lisbon, Portugal, 2013.