OAIS and Distributed Digital Preservation in Practice

An exploration of Danish and other use cases that contributed to the development of the *Outer OAIS–Inner OAIS Model* for *Distributed Digital Preservation*

E. Zierau
The Royal Danish Library
P.O. Box 2149
Denmark
elzi@kb.dk

ABSTRACT

The aim of the paper is to illustrate how the distributed aspects of digital preservation can be aligned in practice, with the concepts and principles of the Open Archival Information System (OAIS) Reference Model.

There has been a growing awareness within the digital preservation community of the need for cooperation between organizations to address digital preservation requirements. One common example is that replicas of preservation copies of digital objects need to be independently preserved (e.g., stored, managed, monitored, documented) to ensure that at least one correct replica will survive for as long as needed. Such independence can be achieved through distributed digital preservation that relies upon specific agreements between participating and contributing organizations. The OAIS Reference Model does not address the challenges of distributed digital preservation in detail, though it acknowledges the potential benefits and the options.

A model in form of an extension to the OAIS Reference Model was developed by a Danish bit repository project, as there was an urgent need for such a model. This model has evolved to the Outer OAIS–Inner OAIS (OO-IO) Model through the international project "Framework for Applying the OAIS Reference Model to Distributed Digital Preservation". Previous papers have presented the theoretical basis for the OO-IO model as a model to describe distributed digital preservation systems in a way that conforms to the OAIS Reference Model, but practical examples of applying the model have been sparse.

This paper provides detailed descriptions of how the need for the OO-IO model emerged, how it has been used for both design and audit of the Danish bit repository, how we plan to use it for minimal effort ingest, and what other use cases there are for applying the OO-IO model for distributed digital preservation purposes. This will illustrate how using the OO-IO model can assist in the analysis of complex digital preservation tasks of a distributed OAIS-conformant repository, where the OO-IO model provides terminology and contribute to break down analysis and audit questions.

KEYWORDS

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

1 INTRODUCTION

Awareness within the digital preservation community of the need for cooperation between organizations to address fundamental digital preservation requirements are traceable from the early days. One of the first digital preservation community reports from 1996 mentions distribution [14]. Up to the start of the 2010s most community discussions of distributed digital preservation have mainly limited to bit preservation [11,12,15]. However, since then cumulative community experience has demonstrated that collaboration, both technical and organizational, is necessary to address the challenges involved in achieving good practice for digital preservation for more and more aspects of digital preservation [9,17].

The Open Archival Information System (OAIS) Reference Model [4], is an international standard[1] that has proven to be extremely useful to the digital preservation community. However, this reference model does not address the challenges of distributed digital preservation in detail. To fill this community need, the Outer OAIS–Inner OAIS (OO-IO) Model provides an extension to the OAIS Reference Model.

The purpose of this paper is to provide detailed examples (primarily based on Danish experiences) that illustrate how the OO-IO model can be used, and thereby hopefully make it clearer and simpler to use the model to support analysis and audit of complex parts of an OAIS in handling distributed digital preservation in practice.

The OAIS Reference Model guides the development of sustainable digital preservation programs by providing a common vocabulary, an information model, and a high-level digital preservation architecture. The OAIS Reference Model defines an OAIS repository[1] as [4] p.1:

> "…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."

This means that the development and long-term management of a full OAIS repository requires not only technology, but also the skills and practices of people within organizations. At the functional level, an OAIS repository can be understood as a system in which information packages are processed by

[1]In order to distinguish between the OAIS Reference Model and an OAIS, this paper denotes an OAIS as an OAIS repository, unless it is qualified with Inner or Outer.

individual functions that have been abstracted into higher-level *functional entities* as depicted in Fig. 1 [2].

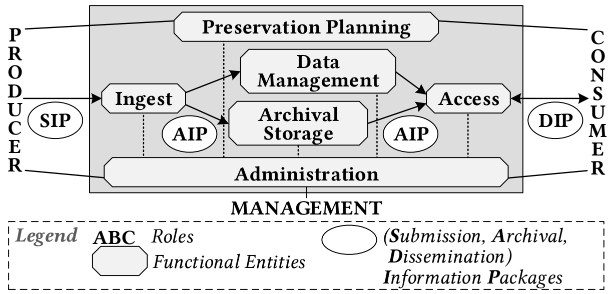

Figure 1: OAIS functional entities.

In the remainder of the paper, terms from the OAIS Reference Model and the names of OAIS functions, functional entities, and information packages appear in *Italic* font.

In section 6 on 'Archive Interoperability', the OAIS Reference Model briefly acknowledges the potential benefits of and the options for federation and distribution states that [4] (p. 6-2):

> "In general one OAIS [repository] is not interoperable with another; however, there are a number of reasons that some level of interoperability may be desirable, motivated for example by Users, Producers or Management"

In relation to distributed digital preservation, the OAIS Reference Model standard also notes that [4] (p. 6-1):

> "An OAIS [repository] may be geographically distributed but with all parts under the same Management, for example the Archival Storage Functional Entity could be divided over several separate locations to increase resilience against disaster. In other cases OAIS Archives with separate Managements may wish to co-operate as described below."

However, to describe the technical and organizational aspects of distributed digital preservation properly, it is necessary to elaborate upon and extend the OAIS Reference Model and that need motivated the design of the OO-IO model.

The Outer OAIS–Inner OAIS (OO-IO) model is an overlay enabling the description of distributed digital preservation systems in a way that conforms to the OAIS Reference Model. The main purpose of the Model is to simplify the challenges when several organizations are involved – both organizationally (what needs to be done) and technologically (how it can be done).

The Outer OAIS–Inner OAIS takes its name from the way the model is used. Here, the term 'Outer' refers to the OAIS repository at the outermost level, e.g., a repository using a bit preservation system, while the term 'Inner' refers to parts of the functional entity of the *Outer OAIS* that is represented by an individual *Inner OAIS*, e.g., a bit repository performing bit preservation for the *Outer OAIS*. In other words; an *Outer OAIS* refers to an entire OAIS repository implementation that supports distributed digital preservation, including all of its *Inner OAIS*'s[3].

The starting point for the OO-IO model was a model developed by a Danish bit repository project. An international project team expanded and broadened the use of the Danish model in "Framework for Applying the OAIS Reference Model to Distributed Digital Preservation" [17] that is referred to as the DDP project. In 2013, the DDP project started as a working group to adapt and extend current standards to address distributed digital preservation. The DDP project included representatives from both North American and European organizations that were engaged in distributed digital preservation, e.g., MetaArchive, the Danish BitRepository.org, Chronopolis, Data-PASS, DuraCloud, Internet Archive, UC3 Merritt, and Archivematica. Variations within these cases identified the need to address other OAIS functional entities, in addition to *Archival Storage*, that require distribution over multiple organizations. Fig. 2 provides an overview of the OO-IO model.

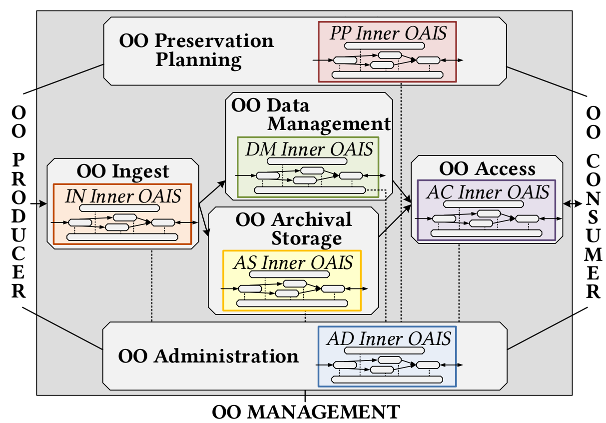

Figure 2: OO-IO model.

The OO-IO model builds upon section 6 of the OAIS Reference Model and, in doing so, the elaboration of the OO-IO model explicitly aligns with the current version of the OAIS Reference Model. By design, the OO-IO model specifies an approach for using the OAIS Reference Model to archive interoperability, in the form of distributed digital preservation, which section 6 does not provide.

Distribution as part of digital preservation will always be complex, and the OO-IO model deals with these complexities by deconstructing them into more manageable and simple components. The idea behind the OO-IO model is that each functional entity of an OAIS repository can be described as a complete OAIS repository (an *Inner OAIS*), that helps an organization (an *Outer OAIS*) to address that functional entity. Each *Inner OAIS* is managed as a complete OAIS repository,

[2]Fig. 1 corresponds to Figure 4-1 in the OAIS Reference Model [4] with a small deviation: In this paper, eight-angled boxes are used to represent functional entities to distinguish functional entities from functions.

[3]This paper uses OAIS's as the plural form of an OAIS repository.

though it is dedicated to managing a single functional entity in the *Outer OAIS*. Using the OO-IO model may only involve one of the *Outer* functional entities, for example *Archival Storage* as a bit repository separated and managed by an external organization.

Previously published work on the OO-IO model has focused on theory, with substantiated arguments to demonstrate that the OO-IO model does not violate the principles of OAIS Reference Model. The most recent is the iPRES paper "Supporting the Analysis and Audit of Collaborative OAIS's - Using an Outer OAIS–Inner OAIS (OO-IO) Model" from 2014 [16]. From a theoretical perspective, the OO-IO model may seem complex and even daunting because distributed digital preservation is a complex topic. This paper aims to make distributed digital preservation more approachable and clear by illustrating the application of the OO-IO model using a set of examples.

The Examples included in this paper span from detailed experience based examples to potential future use cases. Most of the detailed examples concern preservation storage. Preservation storage is often the use case that prompts organizations to recognize the need for a collaborated effort, and therefore this is where most examples can be found today. However, as the DDP project found, there is much more to distributed digital preservation than preservation storage. Therefore, this paper provides more speculative examples that are based in part on the findings of the DDP project.

The experience-based examples describe how the OO-IO model has been used for the Danish bit repository: *First*, as basis for definition of terminology, analysis and design of the bit repository. *Second*, to audit the operating bit repository partly based on the audit and certification standard ISO 16363 [5] that is references the OAIS Reference Model.

Another example reflects the deep immediate interest within the digital preservation community in preliminary preservation storage as part of (pre)ingest. Latest summary of recent discussions can be found on the OAIS Community DPC wiki [1] about pre-ingest. The best poster "Minimal Effort Ingest" from iPRES 2015 addressed this challenge [6], and the topic is also mentioned in the paper on the OO-IO model from 2014 [16]. This example is also about preservation storage, and in spite of the fact that it is not fully implemented, it is included due to the emerging community interest in this case.

Finally, there are short descriptions of distributed digital preservation use cases other than preservation storage and for other OAIS functional entities than *Ingest* and *Archival Storage*.

2 EXAMPLE: BIT REPOSITORY DESIGN

This example explains how the need of the OO-IO model emerged, and how the model was used for terminology, analysis and design for a Danish bit repository. The project started in 2009 as a national pre-study project to investigate the possibilities of creating a shared repository providing bit preservation for Danish cultural organizations.

The establishment of the project was based on the acknowledgement of the need for collaborate effort to create independence between replicas (e.g. to avoid loss of data due to fire or local natural disasters), or more formally to achieve proper bit preservation that requires [12].

- Several replicas of data that is being preserved.
- Independence between the replicas.
- Regular bit audit proving that the replicas are identical.

A great inspiration for this was the LOCKSS program [11], where David Rosenthal in several papers pointed to the importance of independence, even at an organizational level to avoid identical procedural errors being performed on a majority of replicas of data and thus causing loss of data.

The participants in this project were the Danish National Archives and the National Libraries. To avoid misunderstandings that could result from different uses of terms in archive and the library sectors, the participants agreed to apply terms from the OAIS Reference Model. This had the intended effect, at least to the point that bit preservation was part of the *Archival Storage* function entity for each user of a bit repository (illustrated in Fig. 3 as *Outer OAIS* for a user).

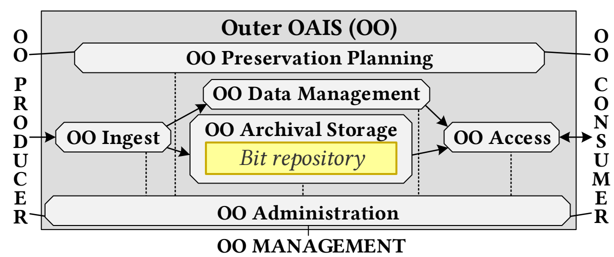

Figure 3: An *Outer OAIS* using a bit repository.

However, it also added great confusion because non-preservation IT staff working with storage facilities regarded *Archival Storage* as representing the bit preservation fully, and therefore in their perspective, bit preservation was solely a question of storage technology. On the other hand, digital preservation staff were convinced that *Preservation Planning* had to be part of the bit repository as well, and therefore bit preservation could not be limited to *Archival Storage*. The challenge was that it was hard to distinguish between users of a bit repository (the *Outer OAIS*'s) and the bit repository itself (as a separate organization). As long as the OAIS repository was seen as *one* repository, it became unclear whether *Preservation Planning* would cover both *Preservation Planning* for the bit repository and the repositories using the bit repository, for example. Furthermore, this approach would conflict with the aim of regarding the bit repository as a repository managed separately from the users' repository.

To mitigate the confusion, the Danish project created a model where the bit repository was depicted as a complete *Inner OAIS* within the *Outer OAIS Archival Storage* functional entity. In this way each of the different cultural organizations represented each their *Outer OAIS*'s using the same bit repository, which was seen as an *Inner OAIS* repository including all the OAIS functional entities. To verify that doing so did not violate the OAIS

Reference Model, analysis were made, and the final model was described and published [15]. Fig. 4 depicts this model in terms of OO-IO, the *Inner OAIS* is called *AS Inner OAIS* since it represents an *Archival Storage*. The names of the functional entities are prefixed with "AS-IO" to distinguish them from *Outer OAIS* functional entities names, which are prefixed with "OO".

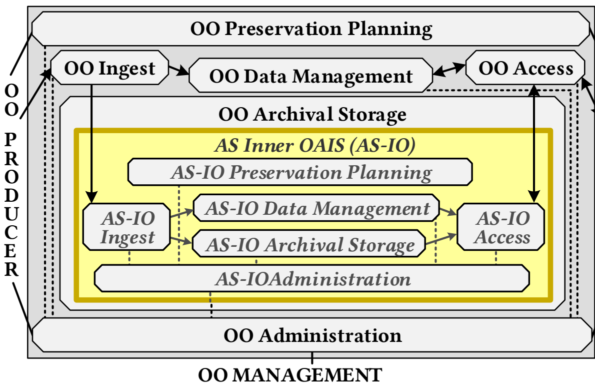

Figure 4: The OO-IO model *Archival Storage* component.

The model helped to avoid further confusion when referring to OAIS terms, simply by specifying whether it concerned the *Outer OAIS*'s (the users) or the *Inner OAIS* (the bit repository). In this way, everybody could agree that the new Danish bit repository should conform to a full OAIS repository (the same way as LOCKSS does [11]), and that this *Inner OAIS* would function as *OO Archival Storage* for each of the *Outer OAIS*'s representing the cultural organizations.

In this way, the AS-IO functional entities of the bit repository were covering parts belonging to bit preservation and thus moved out of the *OO* functional entities. For example, the *monitoring technology* for storage media used for the preservation storage was ***not*** part of the *OO Preservation Planning* but moved into the *AS-IO Preservation Planning*, while monitoring technology for file formats that exists in the OO remained in the *OO Preservation Planning* functional entities and thus not a concern needing attention by the shared bit repository. Likewise for the other functional entities.

It turned out that the model was helpful in many other aspects as well. First, to avoid confusion in architectural discussion by using specific references to whether the discussions concern the *IO* or the *OO* repositories. Second, to identify which parts were covered in *Inner* and *Outer OAIS* not just the functional entities but also information packages and roles. For example *IO AIP´s* (*Archival Information Packages*) were simply bit streams with minimal metadata for lookup, and *IO Management* had to be defined as a separate management dedicated for the bit repository.

The OO-IO model became essential in architectural discussions to analyze the stakeholders' requirements. The three organizations had very different requirements for bit preservation. One organization had only open non-restricted data, while another had requirements coming from the legal framework in connection with confidential materials [7]. From this analysis, it became clear that a flexible architecture was needed to meet varied requirements at different costs. Additionally, there were requirements for different levels of bit safety, where the number of copies has direct impact on cost. Also specific confidentiality requirements required special setups, which again had possible impact on costs and bit safety.

The requirements analysis feed back to the OAIS-based analysis regarding *IO Management*. The *IO Management* consisting of collaborative partners forming the bit repository. The IO Management were to ensure that requirements from *Outer OAIS*'s were met (i.e. requirements from *IO Producers* and *IO Consumers*) Here, the requirements to the Inner OAIS were identified to cover coordination of storing individual replicas on independent platforms services, and coordination of agreements that ensure continuous independence between replicas.

The OO-IO model also assisted in the analysis and definition of the interfaces of the *Outer OAIS* (*OO*) and the bit repository as the *Inner OAIS*. Since several *Outer OAIS*'s were to use the *Inner OAIS*, it was essential to have a common understanding and definition of the information needed for the *IO Ingest* to be able to persistently identify and retrieve ingested bits via *IO Access*. Another important aspect was definition of *audit trails*, where the *Outer OAIS*'s would need *audit trail* information on the bit level as supplementary information to the *Outer OAIS*'s *audit trails*. Digital materials on the logical level (e.g. as a book, where each page has *audit trails* per replica of bit streams representing the page) would need all *audit trail* information, including actions performed on the bit level. Thus, a part of the interface would need to include *IO audit trail* information to be passed as *IO report* information to the *Outer OAIS*'s through the *IO Administration* to the *OO Administration*.

From the analysis and specifications, the Danish project considered using LOCKSS or similar approaches. However, due to the specific requirement that confidential data must be stored off-line, these options were not possible. Instead the pre-study project initiated a new project to develop an actual implementation of an open source framework to set up a bit repository (bitrepository.org [7]).

3 EXAMPLE: BIT REPOSITORY AUDIT

After implementation, the OO-IO model was used to lay out audit points relevant for the bit repository. Since the OO-IO model conforms to the OAIS Reference Model, the ISO 16363 standard for OAIS based auditing was used, at least as inspiration. However, distribution aspects of digital preservation are not specifically addressed in ISO 16363. The lack of guidelines for distribution aspects by ISO 16363 has been addressed in several contexts. For example, as an issue raised in many of the interviews with organizations performing distributed digital preservation in the "Framework for Applying OAIS to Distributed Digital Preservation (DDP)" [17]. Another example is in the paper "Self-assessment of the Digital Repository at the State and University Library, Denmark - a Case

Study [2]. Using the OO-IO model helped identify many additional coordination related issues by defining the *Inner OAIS* responsibilities and functionality. However, the OO-IO model combined with bit preservation theories helped overcome some of these gaps.

When planning for the audit of the bit repository, the first observation was that the bit repository would only cover some types of digital material. For example, the digital materials resulting from digitization of physically preserved documents were not included. Instead these digital materials could be stored more cheaply with some acceptable risk of loss, because re-digitization is possible. Therefore, these materials are kept within the *Outer OAIS* and are ingested into the *OO Archival Storage* via the usual path through OO functions (such as *OO Receive data* and *OO Provide Data*) in parallel with *AS Inner OAIS* covering the bit repository (as depicted in Fig. 5).

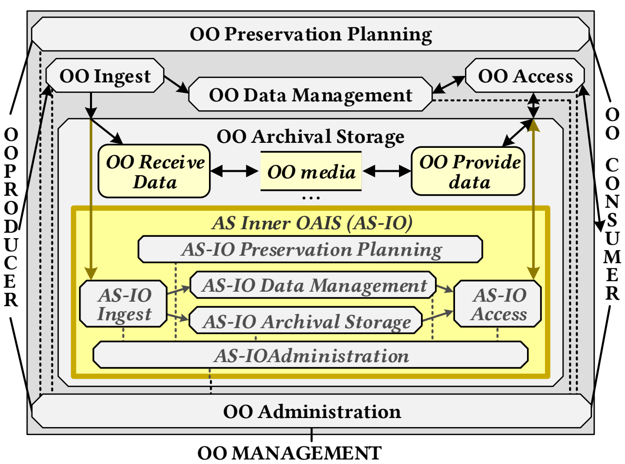

Figure 5: The OO-IO model *Archival Storage* component with parallel OO functions.

This split conforms to the OAIS Reference Model, which allows splitting functions according to needs of an implementation. Consequently, the bit repository audits do not cover digitized versions of physical materials, but only born-digital material or substitution digitization where no loss is acceptable.

In planning of the individual audits, the OO-IO model was used by mapping responsibilities between *Inner* and *Outer OAIS*'s and accumulating evidence across multiple OAIS's to cumulatively demonstrate compliance with digital preservation requirements. This was partly based on the audit and certification standard ISO 16363 and partly on analysis of audit points needed in order to ensure proper bit preservation ensuring continuous independence between replicas and regularly bit integrity checks across replicas. Before outlining the detailed audits, a full picture of *Outer OAIS* and *Inner OAIS*'s was required.

In the actual Danish implementation, the bit repository was not established as an individual organization; instead, each institution negotiated contracts for using each other's facilities for preservation storage of single replicas. Therefore, it became possible to use the OO-IO model in a different way. However, in the Royal Danish Library case, it was an advantage to keep the perspective of the *Inner OAIS* representing the full bit repository, so results from the analysis in the pre-study project could be used.

In the pre-study project, media technology was assumed to be placed in the *AS-IO Preservation Planning*. However, at the time of the launch, it became clear that an extra layer of *Inner OAIS*'s was needed. Replicas of data were in most cases placed on independent types of media in other organizations, where the requisite expertise about specific media existed, and could be monitored. So, all but the in-house replica was manage by a preservation program with a different *Management*. In accordance with the OAIS Reference Model, several OAIS

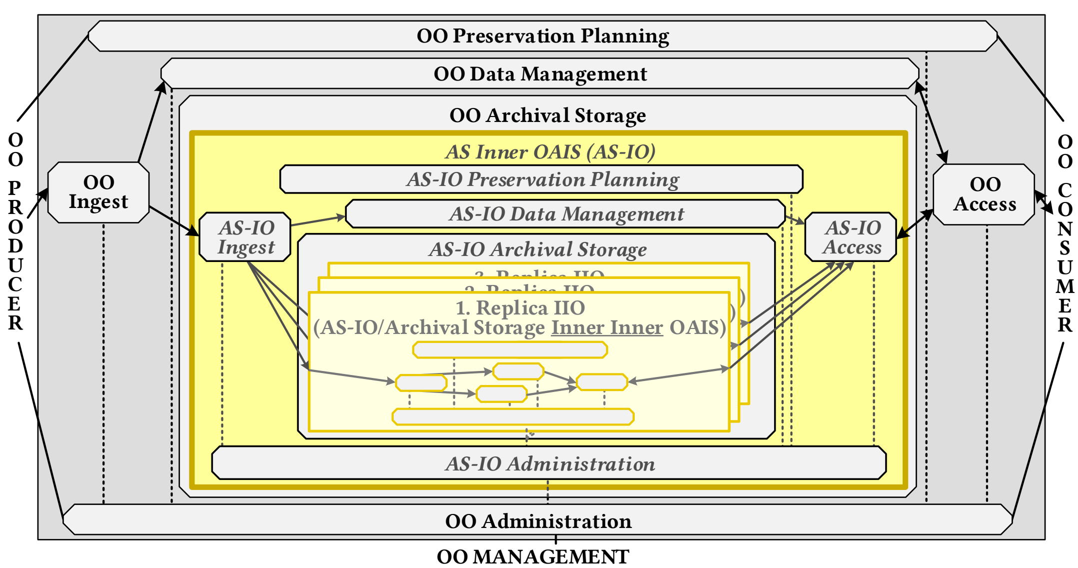

Figure 6: Example of a full picture for a bit repository with Inner (Inner) OAIS's for replica repositories.

repositories with separate *Managements* can work cooperatively, but they are not the same OAIS repository. Therefore each of the organizations with a replica were viewed as separate OAIS repositories. To handle all audits of individual replicas, the in-house replica was viewed in the same way. Fig. 6 illustrates the chosen perspective in terms of the OO-IO model, where the bit repository is the *AS Inner OAIS* to the *Outer OAIS*, and the three units for replicas are *Inner OAIS*'s to the bit repository, i.e. *Inner-Inner OAIS*'s to the *Outer OAIS*. The *Inner Inner OAIS*'s are here referred to as the *Replica IIOs*.[4]

The use of the OO-IO model has, in the Danish case, assisted in maintaining focus within the individual audits. There are a number of people who have multiple roles for parts of the *AS Inner OAIS* and/or a *Replica IIO*, but by explicitly specifying the audit target, the audit process has avoided confusion about which role a particular person should assume in a specific context. Furthermore, it has also been easier to spot when overlapping roles were causing challenges to the independence of replicas and security of data. Lessons learned from the audits are provided below.

3.1 Audit of the *Inner OAIS*

The audit was organized by specifying the roles and responsibilities that the *Inner OAIS* was to cover (excluding what was covered by the *Replica IIO*s). Most of the responsibilities are related to the coordination needed to ensure independence between replicas and uphold integrity across the replicas. One example is that media migration involves a risk of data loss. For example, audit of procedures to avoid simultaneous media migration, and continuous evaluations of whether new media should be included, possibly by including another options of storing replicas, e.g. on molecular DNA. The following describes various findings in the audits of the *Inner OAIS*.

One of the first audits revealed that one person had write rights to data that would enable him to edit data without leaving obvious traces of the change, losing the original in the process. The bit repository contained three replicas of data. Additionally, the bit repository contained votes (in form of checksums) to be able to point to a correct replica in cases where two replicas have errors. The person in question only had access to one of the three replicas, but in total, he had access to a majority of the votes. Therefore, he could change the copy and the checksums, and thereby make the repository software overwrite the original copies with the changed copy. So, the audit detected and mitigated this potential risk.

A more recent example was that an organization wanted to change their operating system on the servers holding one of the replicas. They wanted to switch from Windows to UNIX. However, another online replica was already on a Linux system. Since Linux and UNIX are strongly related operating systems, the conclusion was to stick to Windows to not jeopardize the principle of independence between operating systems.

[4] If given a systematic name, the name should be denoted AS-IO/AS-IIO, they are referred to as the Replica IIOs for readability reasons.

A very current example is that the Danish government decided to merge the Royal Library of Denmark (in Copenhagen) and the State & University Library (in Aarhus) into one organization from January 2017 (now called the Danish Royal Library). Before this merger, there was organizational independence between the replicas, but after the merger, this independence disappeared. A solution to this challenge has not yet been found. An additional potential challenge to independence is that the daily operation of servers will be centralized in this new organization.

In relation to *Ingest*, there were audit points regarding ensuring security in transfer of data to the *Replica IIO*s. In this relation there has been audit points about the state of storage for each replica before any receipt of proper ingest can be given. As described in the next section, one finding was that definition of when ingest is acceptable is not that straightforward, since a receipt of reception from a *Replica IIO* does not necessarily mean that the replica of data has reached its final destination as an independently stored replica.

3.2 Audit of the *Inner Inner OAIS*'s

The main advantage of regarding *Replica IIOs* as individual *OAIS's*, is that it became clear what kind of agreement had be created between the *AS Inner OAIS* and the individual *Replica IIOs*, and thus forming the basis for audits. When the audits of the individual *Replica IIO*s were prepared, the OAIS Reference Model and the ISO 16363 standard were used to identify the different audit point in relation to requirements. The tricky part was to make sure that all requirements were incorporated, since some of them related back to the requirements from the *Outer OAIS*'s, such as those regarding the status of bit integrity, confidentiality, access time, and costs.

The specific requirements regarding confidentiality were particularly hard to fulfill. Here, the *Replica IIO* with storage of data on an off-line tape was considered good for replicas of confidential data. However, challenges appeared already at a preliminary audit conducted before using the bit repository for confidential materials.

The main challenge was that the existing tape installation was tuned for data with no confidentiality issues, where large amounts of data were delivered daily. The new confidential data would only arrive in small portions, possibly monthly. When data arrived at the *tape Replica IIO*, it was placed on a temporary staging area and would not be written to tape before enough data had arrived to fill up a tape. For the new confidential data, this meant there was a high probability that data could remain in the staging area for up to half a year before it was actually written to a tape and placed off-line. This raised two questions:

One relates to bit integrity. The question was whether the data on a staging area could be considered ingested when it had not been placed on its final media. The problem is that risk analysis for data loss is based on the independence obtained from storage on different media. Given the timeframe of six months, this was not considered acceptable.

The second question is strictly related to confidentiality of the data. Since the Danish Royal Library has a policy that bit preserved data must not be encrypted, the question was whether the staging area could be considered safe for confidential data. However, the external tape technology supplier had access to the staging area to be able to assist in solving production problems for the large amounts of open data. Although this was a huge advantage and no problem when working with open data, it was inacceptable for the confidential data.

The conclusion from these findings was that the running tape solution was unfit for confidential data. An analysis showed that it would be quite expensive to bring this platform to an acceptable security level for confidential data. Furthermore, both cost and access time would rise unacceptably for the open data, if stored on a secure platform. Therefore, a separate and much smaller installation would have to be created specifically for the confidential data. In future audits, such special installations are regarded as a separate *Replica IIOs* and thus giving a separate audit.

The rest of the audit points and findings were not of particular interest for the use of the OO-IO model, but merely traditional as audit of a repository, although here limited to the roles and responsibilities of a *Replica IIO*.

4 EXAMPLE: OO-IO USED FOR INGEST

The need for preservation storage as part of *Ingest* is a topic that the Royal Danish Library has been working with since the beginning of this decade [6,16]. The motivation is that the library is not able to control what kind of data is ingested, and where capacity and collection issues may result in longer waiting periods before finalizing the *OO AIPs*. This corresponds to the former mentioned discussions about pre-ingest [1].

Regarding the types of ingested material, there are cases where it is impossible to predict a timeframe for data curation. One example is reception of hard drives from deceased authors to the manuscript collection. Here, there is a long process before the final *OO AIP*s can be created: Moving data from the hard drive, curation, restructuring, enriching and performing initial preservation actions. Furthermore, this process is complicated by numerous directory structures and file formats.

Regarding collection issues, another example is computer games; Video trailers must be harvested and user guides digitized before the final curation process can take place to create the final *OO AIP*s.

Regarding capacity issues, there are numerous examples of digitization projects that are brief and produce large amounts of data. For substitution digitization of fragile and/or deteriorating materials, there are no possibilities for re-digitizing the materials, and thus digital preservation of the new digital materials is crucial. In some cases, such digital materials had to be curated before the final *OO AIP*s could be produced (e.g. interdependent data that needs to be connected) and due to the amount of data produced, the curation period was stretched over a longer period. In other cases, it has been limitations in ingest to the bit repository that delayed the final securement of data.

The bit repository mentioned in the previous sections will be used for this preliminary archiving. Although there is still no formal setup with analysis and audits based on the OO-IO model, it is still an example that is close to be in use in practice.

Preservation storage as part of *Ingest* can be modelled by the *Ingest* component of the OO-IO model and is illustrated in Fig. 7[5]. Here the *Inner OAIS* is named *IN Inner OAIS* since it represents an *Ingest Inner OAIS*. It illustrates that the ingested *Outer OAIS Submission Information Packages* (*OO SIPs*) are preliminarily archived in the *IN Inner OAIS* before it is used for generation of *Outer OAIS Archival Information Packages* (*OO AIPs*) and parsed to the *OO Archival Storage*.

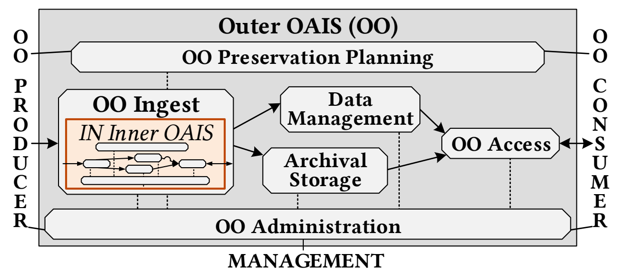

Figure 7: The OO-IO model *Ingest* component.

As for the *Archival Storage* case, the *IN-IO Archival Storage* represents preservation storage, and is therefore modelled as an *Inner Inner* OAIS (Corresponding to the *Inner OAIS* described in the previous sections). This is illustrated in Fig. 8, where the main *OO Ingest* functions are depicted as well, to indicate that there may be materials that will ***not*** be covered by the *IN Inner OAIS*, but take the usual path through the *OO Ingest* functions.

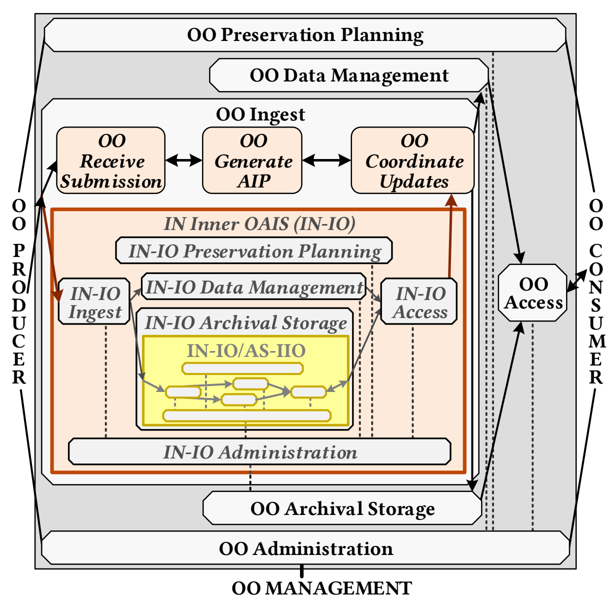

Figure 8: The OO-IO model *Ingest* component with *Inner Inner OAIS* and parallel *Outer OAIS* functions.

[5]Corresponding to Fig. 2, but where only the orange *IN Inner OAIS* is included.

Note that the *IN Inner OAIS* in Fig. 8 does ***not*** cover the full *OO Ingest* functional entity. Instead the *IN Inner OAIS* deliver *IN-IO DIPs* (*IN-IO Dissemination Information Packages*) to the *OO Coordinate Update* function, which cover the further delivery to the *OO Archival Storage* and *OO Data Management* functional entities. According to the OAIS Reference Model [4] (p.4-7):

> "the Coordinate Updates function is responsible for transferring the AIPs to Archival Storage and the Descriptive Information to Data Management"

In this case, the *IN Inner OAIS* will be responsible for receiving *OO SIPs*. It will then perform quality assurance on and generations of the *IN-IO DIPs*, which will have to include the *OO AIPs* plus *OO Descriptive Information* in order for the *OO Coordinate update* function to coordinate the updates at the *Outer OAIS* level. This is the way that we at the Royal Danish Library expect it to be modelled for the Danish case. However, it should be noted that the final detailed model for a specific case would depend on what is included in the *IN Inner OAIS*. For example, the reception of the *OO SIP* along with quality assurance may be placed at the *Outer OAIS* level. In such a case the *IN Inner OAIS* will only cover the *Generate AIP* function of the *OO Ingest* functional entity.

Terminology is an immediate benefit achieved by using the OO-IO model *Ingest* component. A common confusion when dealing with preliminary archiving is that an *AIP* may refer to the preliminary archived *AIP* with minimal metadata before processing (e.g. an image of an author's hard drive) or it may refer to the final *AIP*(s) in the repository (emails, research for different literary works etc.). Here, the OO-IO model can assist in terminology and visualization of what is referred to: *OO AIP* for the final AIPs and *IN-IO AIP* for the preliminary archived *AIP*.

As for the *Archival Storage* component of the OO-IO model, there are also advantages in analysis and delimitation of what is covered by the *IN Inner OAIS*, and thus how to audit.

In the Danish case, the *Inner OAIS* for are *Archival Storage* and the *Inner Inner OAIS* for the *Ingest* IO *Archival Storage* will to a large extend represent the same bit repository. Therefore, a complete illustration of the Danish preliminary archiving case would need to involve all levels down to the Replica level in the same way as illustrated in the previous sections. It should be noted that such an illustration will seem much more complex and voluminous than what has to be audited in practice, since there will be overlap in places where the bit repository and its *Replica IIOs* are the same.

As for the case of *Archival Storage*, all the *Inner OAIS*'s might not be represented by separate organizations with different management, but merely be organizations with relevant roles and responsibilities for the *Inner OAIS* that it represents. Still, defining the *Inner OAIS*'s has an important role in the breakdown of audits and recognition of roles and responsibilities.

5 OTHER DDP USE CASES

The examples in the previous sections are solely related to preservation storage with bit preservation. However, as mentioned in the introduction, the DDP project found other cases where distribution was needed for digital preservation. The following includes such examples as well as examples of use cases where the OO-IO model can be used for the remaining OAIS functional entities.

5.1 Other *Ingest* Use Cases

In the previous section, there was an example of using the OO-IO *Ingest* component for preservation storage. However, preservation storage is not the only example of distributed digital preservation needed in *Ingest*. In the DDP project there were found other use cases, in particular a need for distributed ingest. Examples of distributed ingest were micro-service-based solutions like UC3-Merritt and Archivematica who had examples of using the distribution of micro-services to manage many simultaneous loads of ingest processing.

5.2 Use Cases for *Data Management*

The *Data Management* functional entity covers information, such as catalogs and inventories.

The obvious case related to preservation storage, is when inventory information can be critical, e.g., for collection building or to be able to recreate an operational system quickly, if the catalogs and inventories are corrupted or lost.

Another example is the case of linked data where sources may be spread across several organizations. In these cases, there might be information that has been considered part of the *Data Management* functional entity, but in practice, this information is crucial for future access to *AIPs*. It can also be *Representation Information* that is distributed across organizations: one organization has descriptions of its preserved assets, another organization has the format registry used for the preserved assets, and a third organization has the environment registry used for the preserved assets.

It can be argued that if the OAIS Reference Model is strictly followed, all relevant information should be placed in *OO AIPs*. However, for linked data, it may not be possible. Preservation storage or cooperative solutions may therefore be needed at least for parts of the *Data Management* database, and thus specification of an *Inner OAIS* can be of help.

5.3 Use Cases for *Administration*

The *Administration* functional entity mainly contains services and functions needed to control the operations on a day-to-day basis, including the *Archival Information Update* function that provides a mechanism for updating the contents of an OAIS repository [4].

One use case is the case of complex and time-consuming updates of data related to crowd sourcing. This is a case of distributed digital preservation in the form of preservation storage. Preservation storage may here be required in the *Archival Information Update* function, before re-ingest. A practical example of collecting and securing updates is distributed collection of corrections in archived Optical Character Recognized texts from the public. Such crowd sourcing produces a continuous flow of new data that needs preparation, before it can be passed to *Ingest*. It should be noted that such cases could also be modelled as part of (pre-)*Ingest* as it has similarities to securing data as described in the previous section 4.

Another use case is for collection and maintenance of community based standards used in *Administration*. Standards are important when tracing operations on the long term. Community based standards are important to achieve good practice. On the other hand, digital preservation standards are improving over time, and particularly for community-based standards, there needs to be some control of local implementations to ensure persistent references to the actual standard used.

5.4 Use Cases for *Preservation Planning*

The *Preservation Planning* functional entity is responsible for planning to ensure that the information stored in the OAIS repository remains accessible and understandable.

The most obvious examples of preservation storage needed for *Preservation Planning* regards preservation planning information that is shared and maintained by a digital preservation community, such as format registries like PRONOM [10] and authorities for preservation metadata like event types used in preservation metadata [8]. The best current example is the need for establishment of a common software source library for emulators [13]. This is a huge task, and it is reasonable to assume that no single organization can manage it alone.

The OO-IO model can be used for analysis and auditing of included preservation storage or collaborate maintenance and preservation of the information needed in *Preservation Planning*.

5.5 Use Cases for *Access*

The *Access* functional entity provides services to make archived material visible to *Consumers*. At first glance, this does not seem to involve preservation storage or collaborative efforts. However, an example in a topic of current interest is how access is provided for materials where the emulation preservation strategy is chosen [13]. In such cases, *OO AIP*s will contain technical metadata about the original environments needed for access, but the setup needed for emulation will change over time according to present technologies.

Setting up emulation based solely on preserved data in the *OO AIP*s can be very time consuming (months or years) if all parts of the access platforms are lost (e.g. as a result of a natural disaster). Preservation of an access platform can be crucial for access, although the requirements for such a Preservation may be more short term than the actual digital materials.

There are also cases where responsibilities for access components are shared between different organizations in form of collections of components needed for an emulation platform. One example can be found in the paper "Exhibiting Digital Art via Emulation, Boot-to-Emulator with the EMiL Kiosk System" [3]. If such access platforms are regarded as crucial for an OAIS repository, this would lead to audit of this collaboration.

Both in the case of preserving an access platform and in the case of cooperative maintenance of an access platform, the OO-IO model can be of help in analysis and audits.

6 DISCUSSION AND FURTHER WORK

One of the first questions that comes to mind is whether the OO-IO model can be used in cases where an *Inner OAIS* does not comply with the OAIS Reference Model (e.g. one of the *Replica IIO*s used in the bit repository example). The examples provided have shown that using the OO-IO model in breaking down the complexity in distributed digital preservation can help analysis and auditing, without any reference to compliance. Thus, if achieving the benefits described is an objective, the answer will be yes. However, if the objective is certification then the question is not easily answered, but probably something that should be investigated further.

Another question is how the OO-IO model can and should be referred to by the OAIS Reference Model standard revision that is under way in 2017. As distribution is crucial for digital preservation, it is important that the OAIS Reference Model does address this much better than it does today, and the OO-IO model presently seems to be the best way to do so.

In case the OAIS Reference Model refers to the OO-IO model, this in turn leads to a consideration of how the ISO 16363 standard can be extended to cover distributed digital preservation. There are plenty of evidence that the ISO 16363 standard needs to be better to address distribution and coordination. The standard will most likely benefit from the OO-IO model, but will also need to include more bit preservation aspects especially regarding coordination and replica independence. Furthermore, if the ISO 16363 standard uses the OO-IO model, there has to be guidelines on how to evaluate the entire compliancy based on compliancy of all inherited *Inner OAIS*'s in order to become a certified OAIS repository.

A challenge 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 are 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 the OAIS Reference Model than is required for simpler use cases and implementations. However this complexity is inherent in distributed digital preservation, and not introduced by the OO-IO model. Furthermore, as seen from the examples, it is not an insurmountable task, especially if one step is taken at a time. Taking all steps to begin with will probably be incomprehensible and mind-blowing for most people, as for example introducing OO-IO by starting looking at the OO-IO *Ingest* component illustrated in Fig. 8, but expanded with *Inner Inner Inner OAIS*'s for replicas as done for the OO-IO *Archival Storage* component in Fig. 6. Introduction can also be simplified by use of names for the *Inner OAIS*'s as done for the *Replica IIO*s.

In addition to work mentioned that will be needed in relation to standards, further work on the OO-IO model could be to elaborate more use cases that illustrate and document audit processes of distributed digital preservation. Furthermore, the examples can be elaborated for use for academic and educational purposes.

7 CONCLUSIONS

This paper has presented a number of examples of the OO-IO model used in practice, and how this model supports the application of the OAIS Reference Model in relation to distributed digital preservation. Although the OO-IO model may seem complex at first glance, the examples have hopefully clarified how it can be used to support analysis and audit of complex parts of an OAIS repository.

In summary, this paper has provided detailed examples demonstrating how the Outer OAIS–Inner OAIS (OO-IO) Model supports the specification and audit of collaborative interactions between multiple OAIS repository implementations. The main advantages presented have been:

- Provision of terminology to distinguish between distributed parts and information at different stages.
- A method to conduct detailed analysis of complex distributed digital preservation, by breakdown into manageable components, where interfaces and contents, roles and responsibilities can be defined.
- A method supporting requirements analysis for stakeholders.
- A method to support identification of audit points, roles and responsibilities relevant for auditing.
- A method to enable detailed audits for all organizations involved in distributed digital preservation.

The provided examples and use cases have also demonstrated that the OO-IO model can be used in a flexible way adapted to the actual repository in hand. The flexibility regards:

- whether information is entirely covered by the Outer OAIS or by one or more Inner OAIS's
- whether the Inner OAIS covers the entire Outer OAIS functional entity or just subsets of functions
- whether the Inner OAIS covers a separate repository with separate Management or an internal organisation within the Outer OAIS.
- whether multiple levels of Inner OAIS's should be used

Finally, it has been discussed how the OO-IO model may supplement existing standards to address the important and needed distribution aspect of digital preservation.

ACKNOWLEDGMENTS

A huge thanks to Nancy McGovern, who contributed to an earlier paper on the OO-IO model and to this paper to a degree where we considered that she should be a co-author. This paper is primarily based on Danish examples that demonstrate the development and implementation of the model, so we determined that it made sense for me to be the sole author this time. I would like to acknowledge her contribution to both the present paper as well as to previous work on the theoretical parts, where she contributed to the DDP project and co-authored the iPRES 2014 paper [16]. Also a big thanks to Ulla Bøgvad Kejser, who have given useful input to this paper, but also previous work, where Ulla co-authored the paper about the initial OO-IO model [15]. Also thanks to Andrea Goethals from Harvard University for useful input and discussions on the more theoretical parts and to my Danish colleagues Jakob Moesgaard, Bolette Jurik and Asger Blekinge for useful reviews of this paper.

REFERENCES

References to web resources are here provided using the *drafted* Persistent Web Identifier URI (draft 00: pwid:archive.org:2016-12-12_14.42.12Z:https://tools.ietf.org/html/draft-pwid-uri-specification-00).

[1] DPC wiki – OAIS Community. 2017. *Pre-Ingest*. Persistent web identifier: pwid:archive.org:2017-03-05_11.58.41Z:http://wiki.dpconline.org/index.php?title=Pre-ingest.

[2] Elstrøm, G. V., Junge, J. G. 2014. Self-assessment of the Digital Repository at the State and University Library, Denmark - a Case Study. In *Proceedings of the 11th International Conference on Preservation of Digital Objects (iPRES 2014)*, 271-278.

[3] Espenschied, D., Stobbe, O., Liebetraut, T., Rechert, K. 2016. Exhibiting Digital Art via Emulation, Boot-to-Emulator with the EMiL Kiosk System. In *Proceedings of the 13th International Conference on Preservation of Digital Objects (iPRES 2016)*, 37-40.

[4] ISO 14721:2012. 2012. Space data and information transfer systems - Open archival information system (OAIS) – Reference model.

[5] ISO 16363:2012. 2012. Space data and information transfer systems - Audit and certification of trustworthy digital repositories.

[6] Jurik, B. A., Blekinge, A. A., Christiansen, K. F. 2015. Minimal Effort Ingest. In Proceedings of the 12th International Conference on Preservation of Digital Objects (iPRES 2015), 242-243.

[7] Jurik, B. A., Nielsen A. B., Zierau, E. 2012. Flexible Bit Preservation on a National Basis. In *Proceedings of the IS&T Archiving 2012*, 2-7.

[8] LC Linked Data Service. 2017. *Event Type*, Persistent web identifier: pwid:archive.org:2016-11-18_15.59.35Z:http://id.loc.gov/vocabulary/preservation/eventType.html.

[9] McGovern, N.Y., Skinner, K. (Editors). 2012. *Aligning National Approaches to Digital Preservation*. Educopia Institute Publications, Atlanta, Georgia. Persistent web identifier: pwid:archive.org:2017-03-05_11.12.20Z:http://metaarchive.org/public/publishing/Aligning_National_Approaches_to_Digital_Preservation.pdf.

[10] The National Archive. 2016. *The technical registry PRONOM*, Persistent web identifier: pwid:2016-11-19_12.15.07Z:http://nationalarchives.gov.uk/aboutapps/pronom/default.htm.

[11] Reich, V., Rosenthal, D. 2009. Distributed Digital Preservation: Private LOCKSS Networks as Business, Social, and Technical Frameworks. *Library Trends, vol. 57 no. 3*, 461-475. DOI doi:10.1353/lib.0.0047.

[12] Rosenthal, D.S.H. 2008. Bit Preservation: A Solved Problem?. In Proceedings of the 5th International Conference on Preservation of Digital Objects (iPRES 2008).

[13] Rosenthal, D. S. H. 2015. *Emulation & Virtualization as Preservation Strategies*, report commissioned by The Andrew W. Mellon Foundation, Persistent web identifier: pwid:2016-03-08_23.31.13Z/https://mellon.org/media/filer_public/0c/3e/0c3eee7d-4166-4ba6-a767-6b42e6a1c2a7/rosenthal-emulation-2015.pdf.

[14] Waters, D., Garrett, J. 1996. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Persistent web identifier: pwid:archive.org:2009-10-04_21.41.29Z:http://www.clir.org/pubs/reports/pub63watersgarrett.pdf.

[15] Zierau, E., Kejser, U.B. 2010. Cross Institutional Cooperation on a Shared Bit Repository. In *Proceedings of International Conference on Digital Libraries (ICDL 2010)*. (Later version in 2013 in *Journal of the World Digital Libraries, vol. 6, no. 1*, 25-36. DOI: doi:10.3233/WDL-120098.

[16] Zierau, E., McGovern, N.Y. 2014. Supporting the Analysis and Audit of Collaborative OAIS's - Using an Outer OAIS–Inner OAIS (OO-IO) Model. In *Proceedings of the 11th International Conference on Preservation of Digital Objects (iPRES 2014)*.

[17] 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 (iPRES 2013)*.