

Preservation Planning in the OAIS Model

Stephan Strodl and Andreas Rauber
Institute of Software Technology and Interactive Systems
Vienna University of Technology
{strodl, rauber}@ifs.tuwien.ac.at

Abstract

The OAIS model has gained widespread acceptance as the fundamental design reference model for an archival system by the digital library community. One of its core entities is Preservation Planning, for which central functional entities are given yet without every detail breaking down into activities requirements. The Planets Preservation Planning approach addresses this by providing a way to make informed and accountable decisions on which preservation strategy to implement in order to most suitably preserve digital objects for a given preservation context. It allows the explicit definition of requirements and goals and offers a systematic way to evaluate preservation strategies.

This paper shows the implementation of the *Develop Preservation Strategies and Standards* function of the OAIS model through the Planets Preservation Planning approach. Moreover, the paper examines the information flows within the OAIS Preservation Planning entity and with other functional entities with respect to the Planets model.

1 Introduction

A range of tools exist today to support the variety of preservation strategies such as migration or emulation. Heterogeneous content, complex preservation requirements and goals, and untested tools make the decision on which solution to implement very difficult. In the Planets project, the DELOS Preservation Testbed forms the basis for the Preservation Planning approach. It allows the evaluation of all kinds of preservation strategies against individual requirements, and the selection of the most suitable solution. It enforces the explicit definition of preservation requirements and supports the appropriate documentation and evaluation by assisting in the process of running systematic preservation experiments.

The Reference Model for an Open Archival Information System (OAIS) was published in May 1999 by the Consultative Committee for Space Data Systems (CCSDS) [4], in 2003 the OAIS Model was adopted as ISO 14721:2003 [6]. The primary goal of an OAIS is to preserve information for a designated community over an indefinite period of time. In the community of digital preservation the OAIS model has been widely accepted as a key standard reference model for archival systems.

We put the Planets Preservation Planning Approach in relation to the OAIS reference model. We therefore examine the information flows within the OAIS *Preservation Planning* entity

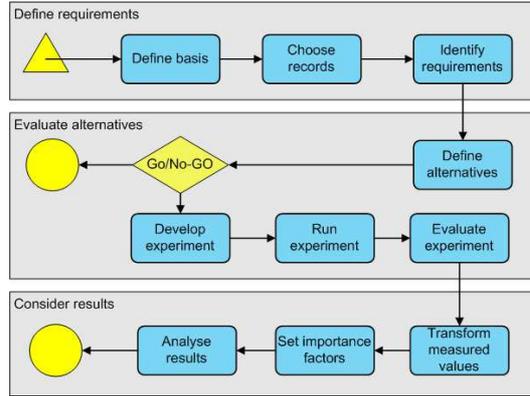


Figure 1: Overview of the Planets Preservation Planning workflow

and with other functional entities. The work presents alerts for preservation planning activities and the according information flow in the OAIS model. It further shows the implementation of the OAIS *Develop Preservation Strategies and Standards* function of the *Preservation Planning* functional entity throughout the Planets Preservation Planning approach. Furthermore, we analyse effects of new preservation strategies on the OAIS model.

2 Related Work

Figure 1 provides an overview of the Planets Preservation Planning approach workflow. The 3-phase process, consisting of 11 steps, starts with defining the preservation scenario, choosing sample records for experiments, and identifying the requirements and goals for the preservation scenario. In the so-called objective tree, high-level goals and detailed requirements are collected and organised in a tree structure.

The second part of the process consists of the definition and evaluation of potential preservation alternatives. Therefore, alternatives are identified, including technical settings and required resources for running experiments. The Go/No-Go-Decision enforces a review of the work in the previous steps. The experiments are set up and run. The last step of the second phase is the evaluation of the experimental outcomes against the requirements and goals defined in the first phase.

In the third part of the workflow, the results of the experiments are aggregated to make them comparable, the importance factors are set, and the alternatives are ranked. The stability of the final ranking is analysed with respect to minor changes in the weighting and performance of the individual objectives using Sensitivity Analysis. After this consideration, a clear and well justified recommendation for one of the alternatives can be made.

The final outcome of a preservation planning activity is a report, summarising the results of all steps. The reports are stored in a knowledge database. The database supports the definition of requirements for preservation activities. A detailed description of the workflow and a range of case studies for different preservation setting are presented in [3, 7].

The OAIS model [4, 6] has gained widespread acceptance as the fundamental design reference model for an archival system by the digital library community. A range of projects are based on the OAIS model, for example the AONS II project by the Australian National Library. The Automated Obsolescence Notification System version 2 (AONS II) [8] enables

to get informed when file formats existing in a repository are obsolete or at risk. The basic principle of the system is to harvest file format registries to receive obsolescence information. AONS II implements parts of the functionality of the Monitor Technology function of the Preservation Planning entity.

The DSpace Institutional Digital Repository System was developed in a cooperation between Hewlett-Packard and MIT Libraries in 2002 [1, 2]. DSpace is a digital repository addressing the long-term preservation concerns. This system is freely available as open-source software under the terms of the BSD distribution license. In 2007 more than 200 projects are using the DSpace software. DSpace is influenced by the OAIS Model in terms of methodology and functional plans. Where possible, DSpace adopts the OAIS model and vocabulary to articulate its own design and terminology. Primarily the data model and the metadata conform with the OAIS reference model. However, the architecture of DSpace does not fully match the functional model of OAIS model. For example, in the current version preservation planning is not available. An Architecture Review Group was organised to review the basic architecture of DSpace. Based on the result of the review a set of recommendations for developments were identified for DSpace 2 [5].

3 Planning in the OAIS Model

This section presents alerts for initiating preservation planning activities and the according information flow in the OAIS model. It further shows the implementation of the *Develop Preservation Strategies and Standards* function through the Planets Preservation Planning approach. Moreover, effects of new preservation strategies on the OAIS model are analysed.

3.1 Alerts

First of all, alerts that can trigger a preservation planning activity are identified. Primarily, three alerts can trigger a preservation planning activity,

New Format Alert (NFA)

If a new file format is accepted in the archive, a *New Format Alert* (NFA) is raised to identify and evaluate an appropriate preservation strategy for the new file format.

The *Monitor Designated Community* tracks the service requirements and the available product technologies of the designated community, producer as well as consumer. Changes in the designated community are for example the usage of new file formats. The usage of a new file formats in the designated community may require to accept the format in the archive. In this case a new preservation strategy has to be identified and evaluated. A *New Format Alert* (NFA) is sent to the *Develop Preservation Strategies and Standards* function.

Another indicator of the usage of new file formats on the producer side is frequent submission of unanticipated formats. In OAIS, the producer submits a submission information package (SIP) to the *Ingest* functional entity. The received SIPs are sent for audit to the *Audit Submission* function of *Administration* functional entity. There, the SIP is verified against the specification of submission agreements by the *Audit Submission* function according to the internal archive policies for audit. The audit is supported by AIP reviews from the *Preservation Planning* functional entity. SIPs with incomplete or missing information, unsupported file formats or standards fail the audit. In this case the *Audit Submission* function informs the producer and the *Ingest* about the negative audit. In order to take unsupported

file formats into account, we extend the OAIS model and the functionality of the *Monitor Designated Community* function. A notification *Unanticipated SIP* is sent for unsupported file formats to the *Monitor Designated Community* function of the *Preservation Planning* functional entity. An increase of submission of unsupported formats can be an indicator for a change of the environment on the producer side and a possible oncoming obsolete file format that is currently held in the archive. The analysis of unsupported submission file formats is necessary to guarantee the long term access and usage of the archival holding. The *Monitor Designated Community* function further analyses the unsupported file formats. Specification and guidelines for the analysis can be defined by archival policies. The analysis can be supported by surveys of the producer community. If too many unanticipated submission are received, the *Monitor Designated Community* function may reach the decision to support new formats. In this case a *New File Format Alert* (NFA) is addressed to the *Develop Preservation Strategies and Standards* function to identify, evaluate and select feasible preservation strategies for the new, unsupported file format.

New Requirement Alert (NRA)

Changes and development in technology and in the designated community can change the requirements for the preservation of digital objects. In this case a *New Requirement Alert* (NRA) is raised to evaluate existing preservation strategies against the new requirements.

The *Monitor Designated Community* tracks the product technologies of the designated community includes new computer platforms and new communication technologies. Underlying changes in the technology that affect the archive are reported to the *Develop Preservation Strategies and Standards* function as *New requirement Alert* (NRA). Existing preservation strategies are re-evaluated to ensure that they fulfil the new requirements.

The *Monitor Technology* function provides technological alerts of obsolete digital technologies including information standards and computing platforms (e.g. hardware and software). Changes in information standards and software may require new preservation strategies for the current holdings. Changes affecting the archive are reported as NRA to the *Develop Preservation Strategies and Standards* function.

Frequent submission of unanticipated formats can also be an indicator for a change of requirements on the producer side. For example, a change of product technologies may cause submissions in new and for the archive unknown formats. It may also imply a possible oncoming obsolete file format that is currently held in the archive. In this case the existing preservation strategy has to be re-evaluated to ensure that long term usage of the file form. A *New Requirement Alert* (NRA) is send from *Monitor Designated Community* function to the *Develop Preservation Strategies and Standards* function.

Revision Alert (REA)

Periodical reviews of existing preservation strategies help to improve and further develop existing strategies. A *Revision Alert* (REA) is raised to evaluate new developed preservation strategies to improve the existing strategies. The reviews are raised by the *Develop Packaging Design and Migration Plans* function on a scheduled basis as *Review Alert* (REA).

We further introduce a extension to the OAIS model to inform the *Monitor Designated Community* function about the operational statistic of the archive. The statistic is send as an inventory report from the *Manage System Configuration* function of the *Administration*. This information allows for periodical review depending on specific file formats or collections for example most frequently accessed file format. The review is raised by the *Monitor*

Designated Community function as REA. The review cycle is defined as policy and provided by the *Establish Standards and Policies* function of the *Administration*.

The *Monitor Technology* provides reports about developments of new file format and tools. New file formats and tools can improve existing preservation strategies. In this case a REA is raised by the *Monitor Technology* function to the *Develop Preservation Strategies and Standards* function.

An review of an existing preservation strategy includes the evaluation of a new version of used preservation tools or the migration to a new file format. This may improve the performance of preservation process or the output, for example improved support of embedded objects.

3.2 Preservation Planning Information Flow

In this section, we identify the information flows for each step of the Planets Preservation Planning approach with function entities in the OAIS model. Figure 2 shows the integration of the Planets Preservation Planning approach in the *Preservation Planning* functional entity. In the following, the information flows for the three types of preservation planning activities are presented.

3.2.1 Evaluation of New Preservation Strategies for New File Formats

When a new file format is accepted in the archival system, a new preservation strategy has to be identified, evaluated and implemented to ensure the long term usability of the objects.

The evaluation of a new preservation strategy is triggered by *New File Format Alerts*.

The information flows of the steps in the Preservation Planning approach, described in detail in [3, 7], are shown in Figure 2.

1. Define Basis

In the first step, the context of the planning activity is described including the collection, the types of objects and number of objects. Also legal and institutional requirements for handling the records are recorded. The legal information is provided by the *Establish Standards and Policies* function of the *Administration*.

2. Choose Records

In this step, a representative sample of records is collected, these can be provided either by the archive itself or by the designated community. For new file formats the designated community is able to provide a broader spectrum of sample records with a variety of file characteristics.

3. Identify Requirements

The following information are provided to support the building of the objective tree,

- **Objective Tree Templates (Preservation Planning Knowledge Database)**

The PP knowledge database provides template sub-trees for specific domains. Examples for template trees would be a content sub tree for pictures, or cost models for libraries. These templates can be integrated in the objective tree and adjusted regarding the specific setting. Alternatively, the objective to be reached by the preservation solution can be built form scratch.

- **Performance information (*Manage System Configuration* function of *Administration*)**
Performance information can raise preservation requirements for performance of preservation strategies for duration per megabit or workload of the CPU to be included in the objective tree.
- **Consumer comments (*Consumer Service* function of the *Administration*)**
The *Consumer Service* function collects feedback from the consumers about the access services and products of the archive. The comments can deduce requirements regarding the access, behaviour and usage of digital objects.
- **Emerging standards (*Monitor Designated Community* function)**
Emerging standards reports identify trends of future developments and can provide requirements that might have to be fulfilled in the future. The consideration of future requirements can improve the longevity of a preservation strategy.
- **Reports from the *Monitor Technology* function**
The reports about emerging digital technologies, information standards and computing platforms indicate future developments. Therefrom requirements for current preservation strategies can be detected.
- **Legal and Institutional Obligations**
Legal as well as institutional obligations for handling the records are provided from the *Establish Standards and Policies* function of the *Administration*. They define requirements for a preservation strategy.

All the information support the process of building the objective tree. This implies that experts from at least these domains as well as additional ones, usually technical, managerial and curatorial are required to participate in the workshops for requirements selection.

4. Define Alternatives

Possible preservation strategies are selected to evaluate them against the specific requirements of the preservation activity. The reports about developments of technology from the *Monitor Technology* function support the identification of new preservation strategies.

The *Monitor Technology* function offers the functionality to evaluate emerging technologies by prototype requests. The results form a first indication for closer consideration of new and untested tools and services. For example, the *Monitor Technology* function tests emerging preservation tools and services against a well defined benchmark corpus, the outcome can indicate the applicability of the technology.

Previous preservation planning activities performed within the given or similar institutions are also useful sources of identify preservation strategies. For example, planning activities with same file formats can be found in the Knowledge Database.

5.-8. Perform Experiments

These steps are performed according the methodology described in [7]. The Go/No-Go Decision is taken and the experiments are developed, run and evaluated against the requirements of the objective tree.

9. Transform Measured Values

The template trees from the PP Knowledge Database contain transformation functions. These functions are based on best practice and may be revised and adjusted to the requirements of the current setting. The transformation values have a strong influence on the result of the evaluation and have to be handled very thoughtfully.

10. Set Importance Factors

Requirements are weighted by importance factors which depend on the preservation setting and on the trigger event, in addition consumer comments and performance information can influence the factors. The importance factors are usually defined in a workshop with experts from different domains.

11. Analyse Results

The results, the documentation of each step and the conclusion of the evaluation activity are submitted to the PP Knowledge Database and provide support for future planning activities.

3.2.2 Revision of Existing Preservation Strategies for New Requirements

Changes and developments of technology induce a change in preservation requirements over time. Therefore, existing preservation strategies for the current holdings have to be revised according the new requirements. If necessary, new preservation strategies have to be identified, evaluated and deployed, this is triggered by *New Requirements Alerts*.

In order to identify the strategies affected by the new requirements, it is useful to identify the categories of collections being affected, for example collections of videos or collections of a specific file format. A collection profiling can support this task.

The evaluation for all affected preservation strategies have to be re-executed. The objective trees need to be adapted regarding the new requirements (Step 3 in the Planets Preservation Planning approach). Steps 1-2 and 4-5 remain unchanged. It might be necessary to develop and run new experiments to evaluate the new preservation requirements. Step 8 *Evaluate Experiments* has to be re-executed with special focus on the new requirements, transformation functions for the new requirements have to be defined. In some cases the importance factors may have to be adjusted.

In the last step of the workflow, the measured values are aggregated to one single comparable number for each alternative. The results have to be analysed whether the existing preservation strategy fulfils the new requirements.

There are the three possible results,

- *The deployed preservation strategy fulfils the requirements and performs best*
No change in preservation strategies is needed.
- *Another preservation strategy performs better*
A change in preservation strategies for this collection is recommended. Another preservation strategy is more suitable for long term preservation.
- *None of the existing strategies fulfils the new requirements*
New strategies have to be identified and compared to find the most suitable strategy.

The evaluation and the gained conclusions of the planning activity are submitted to the Knowledge Database. The database receives also reports from the *Monitor Designated Community* and from the *Monitor Technology* function of the OAIS Model. The reports and the

than an extension or specification in more detail, as has been done so far. We suggest sending the outcome only to *Establish Standards and Policies* function of the *Administration* and use the approved standards information flow to inform the *Develop Packaging Designs and Migration Plans* function about the new preservation strategy. The advantage of this flow is that the *Administration* can take a final confirmation of the new preservation strategy.

The *Develop Packages Designs and Migration Plans* function produces AIP/SIP review, AIP/SIP templates and customisation advices for the new preservation strategy. These are provided to the producer to support the creation of new submissions. The AIP/SIP reviews are used to audit new submission. For new accepted file formats in the archive, it is necessary to negotiate new submission agreements with the producers.

New preservation strategies may require an update of the *Administer Database* function of the *Data Management* entity. The Descriptive Information of objects may need to be changed for new file formats in the archive.

The *Ingest* entity is also affected by new preservation strategies. The *Establish Standards and Policies* function provides format standards, documentation standards and procedures to the *Generate AIP* function of the *Ingest* entity. This information is required to generate AIPs from SIPs. The update is necessary for revised as well as new preservation strategies. The *Develop Packages Designs and Migration Plans* function develops new Information Package (IP) designs and detailed preservation plans. The IP designs assist to apply and review SIPs and AIPs.

The preservation plans and the IP designs are submitted within a migration package to the *Manage System Configuration* function of the *Administration*. The function also receives system evolution policies from the *Establish Standards and Procedures* function. Based on these two inputs, plans for the system evolution are developed and implemented. The evolution plans are sent to the *Archival Information Update* function. These plans include change request, procedures and tools. The *Archival Information Update* function enables mechanism for updating of objects in the archive. The affected objects are requested from the *Access* and the new preservation strategies are applied on the objects. The new objects are submitted to the *Receive Submission* function. The archival policy defines whether updated versions of an object replace the old version or new version are added whilst the older versions are kept.

New or updated preservation strategies effect the *Ingest*, the *Administration*, the *Data Management*, and the *Preservation Planning* entity of the OAIS reference model. Furthermore the producer is affected by new submission requirements. The consumer view on the archived object can also change by new preservation strategies. These far-ranging consequences have to be considered on the implementation of a new preservation strategy in an archive.

4 Conclusion

This paper provides a more detailed specification of the OAIS Preservation Planning functionality. Specifically, it reviews the Preservation Planning approach adapted within the projects of DELOS and Planets for its conformity with and ability to integrate into the OAIS standard. It further identifies alerts that trigger preservation activities. Moreover, the far-ranging consequences of new preservation strategies in an archive are shown.

Acknowledgements

Part of this work was supported by the European Union in the 6th Framework Program, IST, through the PLANETS project, contract 033789.

References

- [1] BASS, M. J., STUVE, D., TANSLEY, R., BRANSCHOFKY, M., BRETON, P., CARMICHAEL, P., CATTEY, B., CHUDNOV, D., AND NG, J. Dspace – internal reference specification -functionality. Internal reference specification, Hewlett-Packard Company and Massachusetts Institute of Technology, 2002.
- [2] BASS, M. J., STUVE, D., TANSLEY, R., BRANSCHOFKY, M., BRETON, P., CARMICHAEL, P., CATTEY, B., CHUDNOV, D., AND NG, J. Dspace – internal reference specification -technology & architecture. Internal reference specification, Hewlett-Packard Company and Massachusetts Institute of Technology, 2002.
- [3] BECKER, C., STRODL, S., NEUMAYER, R., RAUBER, A., BETTELLI, E. N., AND KAISER, M. Long-term preservation of electronic theses and dissertations: A case study in preservation planning. In *Proceedings of the 9th Russian National Research Conference on Digital Libraries (RCDL'07)* (Pereslavl, Russia, October 15-18 2007).
- [4] CCSDS. Reference model for an open archival information system (OAIS). Website, 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- [5] DSPACE ARCHITECTURE REVIEW GROUP. Toward the next generation: Recommendations for the next dspace architecture. <http://www.dspace.org/images/stories/PDFs/dspace-recs.pdf>, 2007.
- [6] ISO. *Space data and information transfer systems – Open archival information system – Reference model (ISO 14721:2003)*, 2003.
- [7] STRODL, S., BECKER, C., NEUMAYER, R., AND RAUBER, A. How to choose a digital preservation strategy: Evaluating a preservation planning procedure. In *Proceedings of the 7th ACM IEEE Joint Conference on Digital Libraries (JCDL'07)* (Vancouver, British Columbia, Canada, June 18-23 2007), IEEE Computer Society, pp. 29–38.
- [8] THE NATIONAL LIBRARY OF AUSTRALIA. Automatic obsolescence notification system (AONS). http://pilot.apsr.edu.au/wiki/index.php/AONS_II.