Preservation Planning with Plato♣

© Hannes Kulovits, Andreas Rauber

Vienna University of Technology, Vienna, Austria
{kulovits, rauber}@ifs.tuwien.ac.at

Abstract

Due to the rapid technological changes in today's information landscape the preservation of digital information has turned into a pressing challenge affecting everybody from large cultural heritage institutions, via industry and small and medium enterprises, to private users wanting to preserve their holiday photos and email communications. Several examples of considerable data loss because of taking the longevity of digital objects for granted have drawn the public's attention to the problem that digital material does not last forever. Also a comprehensive survey among professional archivists conducted by SNIA's 100 Year Archive Task Force arrived at the result that 'Digital information is at risk of being lost' [15] and underlines the growing awareness of the urgency of digital preservation. The challenge of maintaining the physical readability is just one among multitude other problems one comes along when dealing with this issue.

A lot of different strategies, i.e. preservation actions, have been proposed to tackle this challenge. However, which strategy to choose, and subsequently which tools to select to implement it, is a crucial decision which poses significant challenges. The decision must be based on a well-documented and profound analysis of the requirements and performance of the tools taken into consideration.

This tutorial will introduce the major challenges that digital preservation activities face. It will then discuss the advantages and disadvantages of the various strategies that exist to tackle these problems. The OAIS reference model, an ISO standard describing the functional model as well as the information model of an archival information system will be presented, before focusing on a specific task in digital preservation, namely planning a sound strategy for preserving one's digital assets. To this end a solid workflow that leads to a profound, well-documented decision that one can be held accountable for, will be discussed. Additionally, we will present Plato, a software tool that is supporting this workflow, automating the core steps in this endeavour.

1 Introduction

Digital objects have become the dominant way that we create, shape, and exchange information. They increasingly contain essential parts of our cultural, intellectual and scientific heritage; they form a central part of our economy, and increasingly shape our private lives. The ever-growing heterogeneity and complexity of digital file formats together with rapid technological changes turn the preservation of digital information into a pressing challenge. The challenge is to keep electronic data accessible, viewable, and usable for the future, to ensure the survival of our digital artefacts when the original software or hardware to interpret them correctly becomes unavailable [16].

Digital preservation denotes the process of keeping electronic material accessible and usable for a certain period of time. The Digital Preservation Coalition defines it as *the series of managed activities necessary to ensure continued access to digital materials*, and moreover, *refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technical change* [7]. The preservation problem of digital data can be seen as twofold: (1) The physical media the data is recorded on as bit stream must be preserved. The lifespan of data written on today's prevalent media such as DVD or CD-ROM is considered only a couple of years, so reading from them might already be a serious problem in a decade. (2) As digital data is merely a series of binary codes the way to interpret this stream must be preserved. Digital objects require specific programs (and program versions) to open and render them, these in turn require a set of specific libraries and an operating system, which in turn runs on and supports a specific type of hardware components. If any of these is lost, a digital object cannot be rendered anymore and is left to be a useless concatenation of binary data. Even the serendipity of a storage medium being capable of retaining digital data for a millennium is worthless if the means of interpreting it is lost. A good overview of the challenges in Digital Preservation and of preservation strategies is provided in the accompanying document to the

Proceedings of the 10th All-Russian Scientific Conference "Digital Libraries: Advanced Methods and Technologies, Digital Collections" – RCDL'2008, Dubna, Russia, 2008.

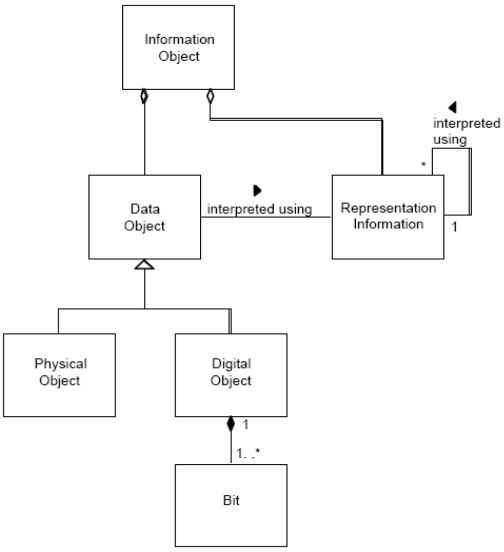

Figure 2 Information Object in the OAIS Model

UNESCO charter for the preservation of the digital heritage [16].

The remainder of this paper is organized as follows: The next section gives an overview of the OAIS reference model as a conceptual framework for an archival system. Section 3 provides an overview of digital preservation, its challenges, and different preservation strategies. Section 4 outlines the Planets Preservation Planning approach and describes the steps of the planning workflow. Section 5 describes the Planets preservation planning tool Plato and its application, and Section 6 summarizes this work.

2 The OAIS Reference Model

In digital preservation standards are very important not only because they promote openness but also for interoperability and sustainability reasons. A generic standard is the Open Archival Information System (OAIS) reference model which was published in 2002 by the Consultative Committee for Space Data Systems (CCSDS). It has proven to be a very useful high-level reference model, describing participants, roles and responsibilities as well as the exchange of information. It does not recommend or specify any particular implementation but identifies and defines certain terms and components that might be involved in an archival information system. Because of its growing acceptance in the community, the OAIS model is the most common framework for digital preservation systems. The reference model was adopted as a standard and registered under ISO 14721:2003 [6]. It defines an OAIS as *'… an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community'* over an indefinite period of time. Furthermore, the OAIS model *'… provides a framework for describing and comparing different long term preservation strategies and techniques.'*

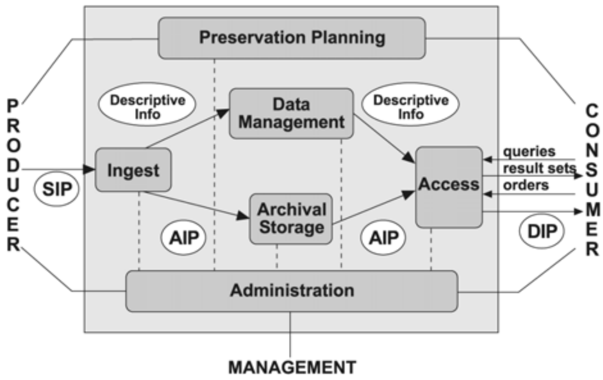

Figure 1 OAIS Reference Model

Figure 1 shows the main functional components and information packages of the OAIS model with the Producer, Consumer, and Management as the main stakeholders of the whole system. While the Producer provides the information to be preserved and the Consumer obtains particular preserved information of interest, the Management role sets overall OAIS policy within a broader policy domain.

The fundamental concept of the OAIS model is the concept of an Information Object consisting of Data and Representation Information whereas the Information Object can be either physical or digital. Representation Information is the information necessary to be capable of fully interpreting the data. Figure 2 illustrates the concept of the Information Object.

The OAIS model differentiates three information package variants: (1) Submission Information Package (SIP) which is negotiated between the Producer and the OAIS, and sent to the system by a Producer. (2) Archival Information Package (AIP) which is used for preservation and comprises a complete set of Preservation Description Information (PDI) for the Content Information. (3) Dissemination Information Package (DIP) which includes a part or all of one or more AIPs and is delivered to the user community by the OAIS.

When a producer submits a digital object into the system, it has to be packaged together with required metadata as a Submission Information Package (SIP). The *Ingest* module provides the services and functions to accept SIPs from Producers. It further performs quality assurance and generates the AIP complying to the archive's standards. Ingest also extracts descriptive information from the AIPs and coordinates updates to *Archival Storage* and *Data Management*. Archival Storage stores, maintains and retrieves AIPs, while Data Management populates, maintains and accesses descriptive information about archived objects as well as administrative data. Every action inside the archive that affects the object is added to the metadata of the AIP. The *Access* component is responsible for supporting consumers in finding, requesting and receiving information stored in the system. Access functions include access control, coordinating requests, generating responses as DIPs and delivering the responses to Consumers.

The *Preservation Planning* entity monitors the environment and provides recommendations to ensure the long-term accessibility of the stored information. This includes monitoring and evaluation of the archive and periodical recommendations on archival updates for migration. A central component is the development of preservation strategies and standards as well as packaging designs and plans.

3 Digital Preservation

3.1 Preservation Strategies

In recent years several different ways of preserving a digital collection have been evaluated, developed, and deployed. Most research on actual solutions is focused on two prevailing preservation strategies – migration and emulation.

Today, **migration** is the most common preservation strategy for reasons as the comparatively small initial effort necessary and price in terms of cost per object. Migration refers to updating a digital object by converting it from one hardware or software generation to another generation or representation. The main goal of migration is to transform to a more suitable or stable representation which is better suited for long-term access. When talking about file formats this might be a newer version of the same file format or a different format. For example, a text document stored in Microsoft's document format MS Word 97 can be migrated either to a newer more up-to-date version (e.g. MS Word 2007) or to PDF/A [5] which is considered more stable. Migration is subdivided into four types varying in the amount of changes applied to the object which in turn result in different levels of possible information loss [2]: (1) Refreshment, (2) Replication, (3) Repackaging, and (4) Transformation. The method involving the lowest risk of information loss is *Refreshment* where the physical medium is replaced by another physical medium with the same storage-mapping infrastructure. The bit stream stored on the medium is transferred without altering it and is used in order to prevent the data from being lost due to media decay. *Replication* is similar to Refreshment. It doesn't alter the Packaging Information but may require changes in the mapping infrastructure in Archival Storage. *Repackaging* causes some changes in the bits and bytes of the Packaging Information. *Transformation* poses the highest risk because changes to the actual content are made.

Basically there are three approaches to the migration of a collection of digital objects: Firstly, the whole collection can be converted to a file format that is standardised, with the goal of reducing the complexity of maintenance. The second option is to keep the original bit stream and migrate when the user accesses the object (migration on demand). Lastly, 'migration within the same format' performs migration every time the underlying format is at risk or changes significantly.

The main challenge when migrating from one format to another is to only transform the logical representation, ensure authenticity and consistency and

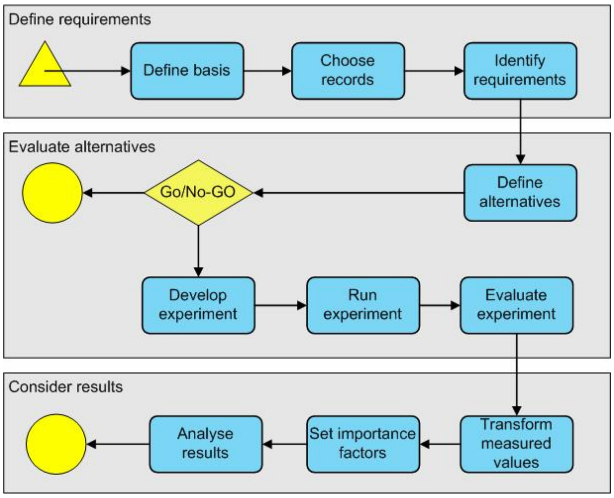

Figure 3 PLANETS Preservation Planning Approach

preserve all essential characteristics for the original object. The Council of Library and Information Resources investigated several migration projects and identified different kinds of risks [8]. As migration has to be performed continuously, i.e. at all times a file format is considered to be at risk, Jeff Rothenberg points out six problems of a migration approach [11]: (1) Labour intensive, (2) Time consuming, (3) Expensive, (4) Error prone, (5) Risky and (6) Non-scalable.

In contrast to migration, **emulation** operates on the environment of the object rather than the object itself. Emulation denotes the recreation of functionality of systems (software and/or hardware) which is needed to render, access, or edit a certain document to overcome technological obsolescence. In digital preservation this means mostly the emulation of a certain version of a software system needed to access a file in an obsolete version or format. Jeff Rothenberg [11] envisions a framework of an ideal preservation surrounding. The Universal Virtual Computer (UVC) concept [4] uses elements of both, migration and emulation, allowing digital objects to be reconstructed in their original appearance. The UVC is independent of any existing hardware or software; it simulates a basic architecture including memory, register and rules. An emerging approach of emulation is modular emulation. It imitates the hardware environment by emulating the components of the hardware architecture. Each hardware component can be seen as an emulator of the component and the components are assembled in order to create a full emulation process.

Emulation is considered the better means for preserving complex objects as the more complex digital objects get, the more loss can occur in the migration process, i.e. the more significant properties can get lost. While having the advantage of leaving the original file untouched without any modifications, writing an emulator is a very complex and time-consuming problem. Additionally there will probably be a point in future where users no longer know how to interact with today's applications such as a certain word processing software. Several projects have been working on the development of emulation approaches. The BBC Domesday project for instance implemented an emulation strategy to preserve access to the digitised version of a 900-year-old Britain book called the Domesday book [9].

Although not a desirable approach **data archaeology** also has been proposed as a preservation strategy especially because sometimes it might be essential to rescue neglected digital data of potentially vital importance. Also unexpected disasters may necessitate the application of special techniques to recover data already considered lost. An overview of different approaches to data archaeology can be found in [10].

3.2 Authenticity of Digital Objects

One of the most challenging problems in conjunction with digital data is that they can be altered easily, without leaving any trace or a clue what the object was like before. Compared to the analogue world a text written on a piece of paper and altered ex post, the original will in some cases be restorable or at least the change can be identified (that it took place) and thus becomes a part of the object. A user who wants to access a digital object at some point in the future must be assured that there have not been any unauthorised changes, neither accidental nor intentional by a malicious manipulation. This means that the digital object must be exactly what it purports to be, i.e. must be *authentic*, and assert confidence in the identity and integrity of it. Digital preservation therefore also has to deal with authentication of digital material in order to prevent it from being maliciously altered or corrupted without knowing it. Various threats to authenticity exist: (1) Decay of storage medium. (2) Different kinds of malicious attacks. (3) Unintentional damage to the object. (4) Natural disaster. (5) Business failure, e.g. company taking care of objects runs out of money. (6) Multiple versions of an electronic document may easily exist and circulate without being discovered.

4 Preservation Planning

4.1 Overview

As already discussed different strategies such as migration and emulation come into question for preserving digital objects such as electronic documents. However, the decision for a specific tool e.g. for format migration or an emulator, as well as appropriate parameter settings for these tools, is very complex. The process of evaluating potential solutions against specific requirements and building a plan for preserving a given set of objects is called preservation planning.

The Planets Preservation Planning approach allows the assessment of all kinds of preservation actions 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 preservation experiments. It is based on work performed in the DELOS Digital Preservation cluster introduced in [12] and described in more detail in [13].

4.2 PLANETS Preservation Planning Workflow

Figure 3 provides an overview of the workflow of the Planets Preservation Planning approach, which was described in [14]. *Define requirements* describes the scenario, the collection that is being considered as well as institutional policies and obligations. Then the requirements and goals for a preservation solution in a given application domain are defined. In the so-called objective tree, high-level goals and detailed requirements are collected and organised in a tree structure. *Evaluate alternatives* consists of the definition and evaluation of potential preservation alternatives. Alternatives are therefore identified, including technical settings and required resources for running the experiments. The Go/No-Go-Decision enforces a review of the work in the previous steps. In the experiments the preservation alternatives are applied to the sample records. The final 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 phase *Consider Results*, results of the experiments are aggregated to make them comparable, 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 accountable recommendation can be made for one of the alternatives.

Define Basis

In the first step, the preservation scenario is described in a semi-structured way including the collection to be considered. Information about the collection includes details about the objects, number of objects in the collection, and legal requirements for handling the records. Moreover, the environment is described in which the preservation process takes place including institutional policies for preservation.

Choose records

In this step, a small number of representative sample records from the collection is taken. The samples, usually between 5 and 10 objects, are used for evaluating the preservation alternatives.

Identify Requirements

The goal of this step is to define clearly the requirements and goals (objectives) for a preservation solution in a given domain. High-level goals are specified, collect detailed requirements, and organise them into a tree structure, referred to as the tree of objectives or shortly, ‘objective tree‘. While the resulting trees usually differ according to specific preservation context, some general principles can be observed. At the top level, the objectives can usually be organised into four main categories:

Object characteristics – describe the visual and contextual experience a user has when dealing with a digital object. Subdivisions may be ‘Appearance’, ‘Content’, ‘Structure’ and ‘Behaviour’, with lowest level characteristics being colour depth, image resolution, forms of interactivity, macro support, or embedded metadata.

Record characteristics – denote the technical foundations of a digital record, the context, interrelationships among records and metadata. A record can include one or more objects.

Process characteristics – refer to the preservation process. These include usability, complexity, or scalability.

Costs – have a significant influence on the choice of a preservation strategy and may usually be divided into technical and personnel costs or start-up and operational expenditures.

The objective tree is usually created in a workshop with experts from different domains contributing to the requirements gathering process. The tree is independent from the preservation actions that are considered. It models the requirements, not the actions to be taken. The tree documents the individual preservation requirements of an institution for a given collection of objects. Typical trees may contain from 50 up to several hundred objectives, usually organised in four to six hierarchy levels.

Objective trees were initially created with post-it notes on a flip chart. While this is convenient for certain environments, an alternative way has been introduced and the feedback on it has been very positive. This involves the use of mind-mapping software, usually projected onto a large screen to provide an overview, to allow multiple stakeholders working on the tree. A mind-mapping software that has been used several times and has proved very helpful is FreeMind[1] which is open source and can be downloaded and used freely.

Having defined the objectives, the next step is to assign measures to each of the objectives in the tree, which provides metrics to determine how successful a requirement is met. Wherever possible, the metrics should be objectively quantifiable (e.g. € per year, frames per second). In some cases, (semi-) subjective scales are necessary, for example degrees of openness and stability, support of a standard, degree of file format acceptance within different communities.

Define alternatives

Different preservation strategies, using for example migration tools or emulators, are selected. A detailed description of each preservation alternative is provided. The description includes the software environment and parameter settings of the tool, in order to ensure a clear understanding of the alternative and allow a later re-evaluation of the planning process. For each defined alternative, the amount of work, time, and money required for running experiments is estimated.

Go/No-Go

Some experiments need a considerable amount of effort and required resources to run the experiments, for example experiment with great number of alternatives or high cost of hardware and software to run the experiments. Feasibility of the proposed alternatives is determined in this step by considering the defined requirements, the selected preservation alternatives, and estimated resources. The result is a decision for continuing the evaluation process or a justification of the abandonment or postponement of certain alternatives.

Develop experiment

In the experiments, the preservation alternatives are applied to the previous defined sample records. The results of the experiments are later evaluated against the goals and requirements of the objective tree. In order to run repeatable tests, it is important to document all relevant experiment settings. This stage produces a specific development plan for each experiment, which includes the workflow, the software and hardware systems used for the experiments, and the mechanisms to capture the results. All items needed for the experiment will be developed and/or installed and tested, including copies of the sample objects, software packages and programs, and mechanisms for capturing the results.

Run experiment

Experiments are designed to test one or more aspects of a specific preservation alternative when applied to the previously defined sample records. Running an experiment produces results, for example converted computer files, revised metadata, and measured workload of the hardware. The results are evaluated in the next step.

Evaluate experiments

The results of the experiments are evaluated to determine the degree to which the requirements defined in the objective tree were met. Therefore, the leaf objectives defined in the objective tree are evaluated with the defined measurement unit. For each alternative, the outcomes of this stage are measured performance values for each leaf in the objective tree.

Transform measured values

The measurements taken in the experiments might all be measured on different scales. In order to make these comparable, they are transformed to a uniform scale using transformation functions. These transformation functions can define thresholds or injective mathematic functions to map the measured values to the uniform scale. The resulting scale ranges from zero to five. A value of zero denotes an unacceptable result and thus serves as a dropout criterion for the whole preservation alternative.

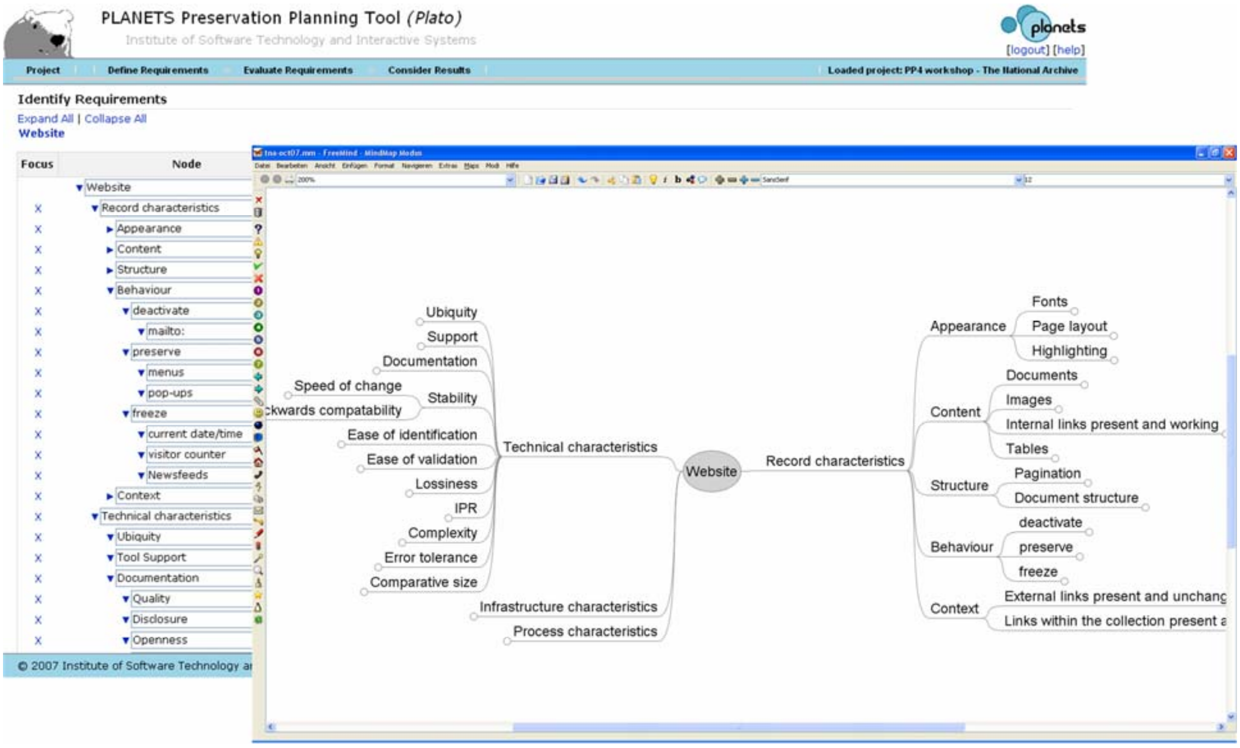

Figure 4 Requirements Definition in FreeMind and subsequent import into Plato

Set importance factors

Not all of the identified objectives are equally important and different degrees of conformance of a solution are accepted in different objectives. This step assigns importance factors to each objective depending on the specific preferences and requirements in the scenario.

Analyse results

In this step, the performance measures for the individual objectives are aggregated to one single comparable value for each alternative. The following two methods are mostly used:

Sum – The measured performance values, as transformed by the transformation functions, are multiplied by the weighting factor. These values are summed up to a single comparable value per alternative. Leaf values that score zero (measured performance under required minimum threshold) have no decisive effect on the final root value.

Multiplication – The first step here is to multiply the comparable value per leaf by the weight of that leaf. The results are then multiplied throughout the tree for the whole alternative. The multiplication method highlights alternatives with drop out values, as these alternatives with leaf values zero have a final root value of zero.

We thus obtain aggregated performance values for every part of the objective tree for each alternative, including an overall performance value at the root level. A first ranking of the alternatives can be done based on the final root values associated with each alternative. This ranking is based on the specific requirements of the preservation context. It forms the basis for a documented and accountable selection of a specific preservation alternative. Furthermore, an analysis of all parts of the objective tree can identify the strengths and weakness of an alternative.

5 Preservation Planning Tool (Plato)

The preservation planning tool Plato is a web-based software tool that implements the workflow depicted in Figure 3 with the aim to provide support to the user in the complex decision making process of finding the optimal preservation strategy. Plato enables the evaluation of potential preservation strategies against defined requirements by conducting experiments defined by the planner on representative sample objects. Plato includes additional external services to automate this process. Furthermore it extends the workflow with a fourth phase in which an executable preservation plan is created, based on the well-documented recommendation.

The software itself is a J2EE web application that relies on open frameworks such as Java Server Faces and AJAX for the presentation layer and Enterprise Java Beans for the backend. It is integrated in an interoperability framework that guarantees loose coupling of services and registries through standard interfaces and provides common services such as user management, security, and a common workspace.

Based on this technical foundation, the aim is to create an interactive and highly supportive software environment that advances the insight of preservation planners and enables proactive preservation planning. In principle, there are three aspects to consider:

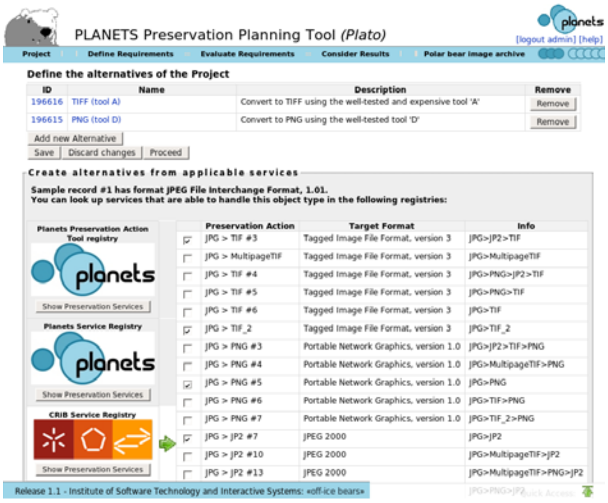

Figure 5 Discovery of Preservation Actions in Plato

(1) Integrating registries for information discovery; (2) Integrating services for preservation action and characterisation of objects; and (3) Proactively supporting the planning with a knowledge base that holds reusable patterns and templates for requirements recurring in different planning situations. The right choice of samples that are representative for the collection under consideration is essential, as any skewed representation might lead to wrong results. Collection profiling services based on characterisation services and format registries inform the selection process and ensure the right stratification of samples. Risk assessment services further assist by quantifying both the inherent risks of object formats and the salient risks present in the objects which are of particular relevance to a specific file format, such as the number of pages for some document formats or the presence of transparency layers in images. The specification of requirements in a tree structure is often done in a workshop setting. This is supported by both a flexible web interface as depicted in Figure 4 and a direct tree import from the openly available mind-mapping software FreeMind. The knowledge base provides recurring fragments and templates, such as process requirements for an archival institution or essential object characteristics for electronic documents in a library, to assist in the process of tree creation. Service discovery is the prime issue during the next step of defining alternatives to consider for evaluation. Starting from the sample objects and their formats, the system queries available registries of preservation actions and looks up applicable tools such as emulators of the original environment or migration tools that can handle the provided input format. Figure 5 shows potential atomic and chained migration services for migrating JPEG files to several different file formats. The Planets registry moreover holds information on benchmark evaluation results produced by experiments carried out in the Planets Testbed, which provides a controlled environment for preservation experiments [3]. Preservation action tools that are accessible through a web service are directly invoked during the execution of

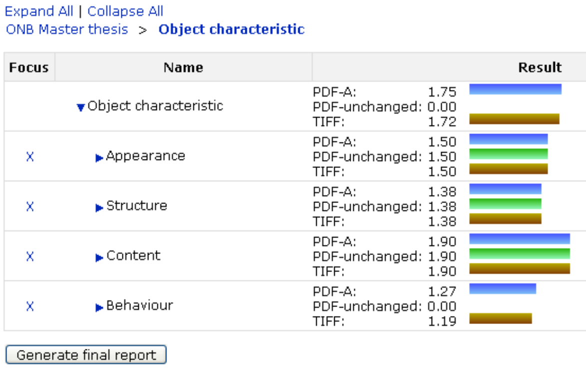

Figure 6 Visualisation of Evaluation Results

experiments on the sample objects; other tools such as emulators have to be executed externally. The evaluation of experiments is probably the most complex and, so far, least automated step in preservation planning. Until today, most of the judgement, e.g. if a migration tool accurately preserves the colour model of an image or the line breaks in a document, has to be carried out manually by looking at the rendered objects. However, characterisation services are available that can measure some of the essential characteristics of objects such as the dimensions of images. In contrast to characterisation tools like JHove, the extensible characterisation languages (XCL) [1] do not attempt to extract a set of characteristics from a file, but instead are able to express the complete informational content of a file in a format independent model. Comparison services specify measurable properties as well as property-specific metrics and their implementation as algorithms in order to identify degrees of equality between two objects. This is in principle independent of the applied strategy, i.e. migration or emulation. The compared objects can be both the original and a migrated object, or the original object in two different environments. To allow comparison and evaluation, a mapping is created between the requirements specified in the objective tree and the characteristics that can be measured and compared automatically by the available characterisation tools. This mapping partly stems from the knowledge base, but can be adapted by the user. Both XCL and other characterisation tools such as JHove are integrated in the evaluation of experiments. This also includes risk assessment services which compare the risk scores of objects resulting from the application of preservation actions against the scores of the original samples. The transformation of measured values to a uniform scale as needed for the aggregation of results and the importance weighting of requirements are supported by the knowledge base. Analysis of results is facilitated by a dynamic and flexible visualisation as depicted in Figure 6, where the planner can choose between different aggregation methods and dynamically configure the information content to analyse the strengths and weaknesses of the alternatives considered. Based on this analysis, a well-documented and solid recommendation for a solution can be made. This recommendation forms the basis for building a preservation plan in the fourth stage. A preservation plan contains a description of the context and the decision taken, including the complete evidence base. This evidence base comprises a thorough description of the planning context and environment, ranging from the institution's mission statement via user group characteristics and policies to the collection at hand (documented in a collection profile). Moreover it contains the chosen sample objects, the requirements and additional documentation as well as considered solutions and the evaluation results. The plan furthermore contains cost indications and triggers for re-iterating the planning, and as a core part it entails a preservation action plan. If the applied strategy and its deployment support it, this is an executable workflow accessing distributed services. During the fourth stage, the planner may select a subset of the criteria used for evaluating solutions to be applied automatically with each preservation action as a mechanism for quality assurance. The corresponding characterisation actions which are used for property extraction and validation are then included in the executable preservation plan.

The first version of Plato is available[2] and freely accessible to the public. It implements the described workflow [14] and provides partial service integration such as file format identification. Subsequent versions will include a wider set of services for preservation actions and characterization. The final version will include the creation of an executable preservation plan.

6 Conclusions

Starting from the challenges of digital preservation and the OAIS reference model this paper presented the Planets Preservation Planning tool Plato. It supports the planner in evaluating available solutions against the specific requirements of a particular situation. In order to arrive at a profound and well-documented decision for a specific preservation strategy Plato implements a solid workflow that has been validated in a series of case studies. The procedure can be applied to any class of strategy, be it migration, emulation or any other approaches.

References

[1] C. Becker, A. Rauber, V. Heydegger, J. Schnasse, and M. Thaller. A generic XML language for characterising objects to support digital preservation. In Proc. 23rd Annual ACM Symposium on Applied Computing (SAC'08) (Fortaleza, Brazil, March 16–20 2008), vol. 1, ACM, pp. 402–406.

[2] CCSDS. Reference Model for an Open Archival Information System (OAIS) – blue book, 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf. Accessed: July 1st, 2008.

[3] A. Farquhar, and H. Hockx-Yu. Planets: Integrated services for digital preservation. Int. Journal of Digital Curation 2, 2 (2007), 88–99.

[4] J. Hoeven, R. Van Der Diessen, and K. Van En Meer (2005). Development of a universal virtual computer (UVC) for long-term preservation of digital objects. Journal of Information Science, Vol. 31 (3), 196–208.

[5] ISO. Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A) ISO/CD 19005-1, 2004.

[6] ISO. Space data and information transfer systems – Open archival information system – Reference model (ISO 14721:2003), 2003.

[7] M. Jones and N. Beagrie. Preservation Management of Digital Materials: A Handbook. Digital Preservation Coalition, 2002. http://www.dpconline.org/graphics/handbook, accessed: July 1st, 2008.

[8] G. W. Lawrence, W. R. Kehoe, O. Y. Rieger, W. H. Walters, and A. R. Kenney. Risk management of digital information: A file format investigation. CLIR Report 93, CLIR, June 2000.

[9] P. Mellor. Camileon: Emulation and BBC domesday. RLG DigiNews, 7(2), April 2003.

[10] S. Ross, and A. Gow. Digital Archaeology: rescuing neglected and damaged data resources, London: South Bank University, Library Information Technology Centre, www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2.pdf, Accessed: July 10th, 2008.

[11] J. Rothenberg. Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Council on Library and Information Resources, January 1999. http://www.clir.org/pubs/reports/rothenberg/contents.html.

[12] C. Rauch, and A. Rauber. (2004). Preserving digital media: Towards a preservation solution evaluation metric. In Proceedings of the 7th international conference on asian digital libraries (ICADL 2004) (p. 203–212). Berlin, Heidelberg: Springer.

[13] S. Strodl, A. Rauber, C. Rauch, H. Hofman, FrancaDebole, and G. Amato. (2006). The DELOS testbed for choosing a digital preservation strategy. In Proceedings of the 9th International Conference on Asian Digital Libraries (ICADL 2006) (p. 323–332). Berlin, Heidelberg: Springer.

[14] S. Strodl, C. Becker, R. Neumayer, and A. Rauber. (2007). 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) (p. 29–38). New York, NY, USA: ACM.

[15] The 100 Year Archive Task Force. The 100 year archive requirements survey (2007). http://www.snia.org/forums/dmf/programs/ltacsi/100_year, accessed: July 4th, 2008.

[16] UNESCO charter on the preservation of digital heritage. (Adopted at the 32nd session of the General Conference of UNESCO).

♣ Work presented in this paper is partially supported by European Community under the Information Society Technologies (IST) Programme of the 6th FP for RTD – Project IST-033789.

[1] http://freemind.sourceforge.net

[2] http://olymp.ifs.tuwien.ac.at:8080/plato