Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans

Article *in* International Journal on Digital Libraries · December 2009

DOI: 10.1007/s00799-009-0057-1 · Source: DBLP

CITATIONS
81

READS
478

6 authors, including:

Christoph Becker
University of Toronto
92 PUBLICATIONS **1,059** CITATIONS

SEE PROFILE

Mark Guttenbrunner
TU Wien
15 PUBLICATIONS **203** CITATIONS

SEE PROFILE

Hans Hofman
18 PUBLICATIONS **224** CITATIONS

SEE PROFILE

Int J Digit Libr
DOI 10.1007/s00799-009-0057-1

Systematic planning for digital preservation: evaluating potential strategies and building preservation plans

Christoph Becker · Hannes Kulovits · Mark Guttenbrunner · Stephan Strodl · Andreas Rauber · Hans Hofman

Abstract A number of approaches have been proposed for the problem of digital preservation, and the number of tools offering solutions is steadily increasing. However, the decision making procedures are still largely ad-hoc actions. Especially, the process of selecting the most suitable preservation action tool as one of the key issues in preservation planning has not been sufficiently standardised in practice. The Open Archival Information Systems (OAIS) model and corresponding criteria catalogues for trustworthy repositories specify requirements that such a process should fulfill, but do not provide concrete guidance. This article describes a systematic approach for evaluating potential alternatives for preservation actions and building thoroughly defined, accountable preservation plans for keeping digital content alive over time. In this approach, preservation planners empirically evaluate potential action components in a controlled environment and select the most suitable one with respect to the particular requirements of a given setting. The method follows a variation of utility analysis to support multi-criteria decision making procedures in digital preservation planning. The selection procedure leads to well-documented, well-argued and transparent decisions that can be reproduced and revisited at a later point of time. We describe the context and foundation of the approach, discuss the definition of a preservation plan and describe the components that we consider necessary to constitute a solid and complete preservation plan. We then describe a repeatable workflow for accountable decision making in preservation planning. We analyse and discuss experiences in applying this workflow in case studies. We further set the approach in relation to the OAIS model and show how it supports criteria for trustworthy repositories. Finally, we present a planning tool supporting the workflow and point out directions for future research.

Keywords Digital preservation · Preservation planning · OAIS model · Decision making · Evaluation · Trusted repositories · Compliance

C. Becker (✉) · H. Kulovits · M. Guttenbrunner · S. Strodl · A. Rauber
Vienna University of Technology, Vienna, Austria
e-mail: becker@ifs.tuwien.ac.at

H. Kulovits
e-mail: kulovits@ifs.tuwien.ac.at

M. Guttenbrunner
e-mail: guttenbrunner@ifs.tuwien.ac.at

S. Strodl
e-mail: strodl@ifs.tuwien.ac.at

A. Rauber
e-mail: rauber@ifs.tuwien.ac.at

H. Hofman
Nationaal Archief, The Hague, The Netherlands
e-mail: hans.hofman@nationaalarchief.nl

1 Introduction

The longevity of digital objects used to be something taken for granted by many, until in the last decade, several instances of spectacular data loss drew the public's attention to the fact that digital objects do not last forever. One of the best known case studies in digital preservation, the rescue of BBC Domesday [37], is a prominent example of almost irrecoverable data loss due to obsolescence of hardware and software capable of reading and interpreting the content. Large amounts of money and effort were required to make the data accessible again and adequately preserve them for the future. Recently, a survey among professional archivists underlined the growing awareness of the urgency of digital preservation

Published online: 19 December 2009

[52]. This awareness has led to the development of various approaches that deal with the question of preserving digital objects over long periods of time. Thorough discussions are presented in [39,54].

An important part of ongoing efforts in many large international projects is the outreach to vendors for advocating document engineering technologies for sustainable documents. The effects can be seen in standards such as PDF/A [25], the Open Document Format (ODF) [27], or MPEG-7 [23]. However, many objects exist and many more are created every day that face the threats of obsolescence. Hence, ex-post actions for preserving access to content are necessary. Preserving authentic records also means being able to prove authenticity [17,48], but creating new manifestations of digital files in different representation formats always incurs the risk that parts of the content are not converted correctly. Hence, when migrating digital files, keeping the original bitstreams as a fallback strategy is common practice. However, having access to the original bitstreams does not guarantee that they are still legible in the future.

The primary reason why digital objects become inaccessible lies within their very nature. In contrast to traditional non-electronic objects such as books or photographs which immediately *are* the content, a digital object always needs an environment to render, or *perform*, it. These environments keep evolving and changing at a rapid pace, which brings about the problem of digital continuity. The prevailing approaches today can be divided along this line: While migration transforms the objects to more widely accessible representations, emulation creates a technical environment where the objects can be rendered.

Various migration tools are available for converting objects in standard file formats such as office documents to representations that are considered more stable. The picture is less positive for more exotic and complex compound objects. However, even within migration tools for office documents, variation regarding the quality of conversion is very high. Some tools fail to preserve the proper layout of tables contained in a document; others miss footnotes or hyperlinks. The task of finding out which information has been lost during a conversion, and whether this loss threatens the value of the object for a given purpose, is a very time-consuming one. Some losses might be acceptable, while others threaten the authenticity of documents. For example, if migrating the collection of Word documents mentioned above results in a loss of page breaks, then this might be irrelevant if the textual content is the only thing of interest. However, if there are page references in the text, then this loss might be unacceptable.

While migration operates on the objects and transforms them to more stable or more widely adopted representations, emulation operates on the environment of an object, trying to simulate the original environment that the object needs, e.g. a certain processor or a certain operating system. This has the advantage of not changing the original objects, and of providing authentic access in much the same way as before. However, emulation is technically complex to achieve and hard to scale up to large amounts of data. Furthermore, users may have difficulties in using old software environments, and some functionality of newer systems, such as the copy-and-paste which is ubiquitous today, might not be available when relying on the original environment of an object. Moreover, as with migration, specific characteristics of an object may be lost due to incomplete or faulty emulation, or due to the impossibility of emulating certain aspects.

The number of file viewers and file conversion tools for standard types of objects such as images or electronic documents is steadily increasing. However, choosing the right treatment for a given set of objects is a crucial decision that needs to be taken based on a profound and well-documented analysis of the requirements and the performance of the tools considered. The complex situations and requirements that need to be considered when deciding which solution is best suited for a given collection of objects mean that this decision is a complex task.

This task can, on the one hand, be seen as a domain-specific instance of the general component selection problem which has a long history in the areas of Software Engineering and Information Systems Design. On the other hand, it is one of the key responsibilities of the *preservation planning* function which is at the heart of the Open Archival Information Systems model (OAIS) [24].

Until now, the selection procedure is mostly an ad-hoc procedure with little tool support and poor documentation. This also implies that decisions that have been and are made are not transparent, hardly reproducible and often not well documented. However, accountability is widely seen as a major requirement for a trustworthy repository; and trustworthiness is probably the most fundamental requirement that a digital repository preserving content over the long term has to meet.

This article describes a solid and well-documented method and workflow for creating preservation plans for sets of digital objects. The method follows a variation of the utility analysis approach for supporting multi-criteria decision making procedures in digital preservation planning. Preservation planners empirically evaluate potential action components in a controlled setting and select the most suitable one with respect to the particular requirements of a given setting.

This article is structured as follows. Section 2 lines out related study in the areas of digital preservation, trusted digital repositories, and component selection. Section 3 introduces the concept of a *preservation plan* and discusses the main components that are considered necessary. Section 4 describes the Planets Preservation Planning methodology for defining plans that fulfill the requirements outlined in the previous section. It further reports on experiences gained in applying the approach in real-world case studies, and

discusses best-practice recommendations on its usage. Section 5 sets the described method in context to existing models and requirements. It discusses compliance with the OAIS model and criteria catalogues for trusted digital repositories. Section 6 describes the planning tool Plato which supports and automates the workflow as a reference implementation. Section 7 provides a detailed discussion of recent case studies lessons learned, while the final Section 8 summarises discussions and provides conclusions as well as an outlook on future work.

2 Related work

2.1 Digital preservation and preservation planning

A number of research initiatives have emerged in the last decade as memory institutions realised the urge of the digital preservation problem [52].

Many repositories follow the Reference Model for an OAIS described in [24]. The OAIS model was published in 2002 by the Consultative Committee for Space Data Systems (CCSDS) and adopted as ISO standard ISO 14721:2003. It has proven to be a very useful high-level reference model, describing functional entities and the exchange of information between them.

Because of its growing acceptance in the community, the OAIS model is the most common framework for digital preservation systems. One of its key functions is preservation planning which lies at the heart of any preservation endeavour.

Preservation planning is also one of the core issues addressed by the project Planets,[1] which is creating a distributed service oriented architecture for digital preservation [29]. Farquhar presents an overview of the distributed service infrastructure and the main components that form the Planets system [15]. Strodl et al.[50] present the PLANETS preservation planning methodology that aids in reaching well-founded decisions. The method has since then been evaluated in a series of case studies [4,5,18]. An OAIS-based analysis of the approach is described in [51]. The approach described here is an extension of this evaluation methodology.

Migration requires the repeated copying or conversion of digital objects from one technology to a more stable or current, be it hardware or software. Each migration incurs certain risks and preserves only a certain fraction of the characteristics of a digital object. The Council of Library and Information Resources (CLIR) described experiences with migration in [35], where different kinds of risks for a migration project are discussed.

Emulation as the second important strategy strives to reproduce all essential characteristics of the performance of a system, allowing programs and media designed for a particular environment to operate in a different, newer setting. Jeff Rothenberg [47] envisions a framework of an ideal preservation surrounding. The Universal Virtual Computer (UVC) concept [20] uses elements of both migration and emulation. It simulates a basic architecture including memory, register and rules. In the future, only a single emulation layer between the UVC and the computer is necessary to reconstruct a digital object in its original appearance. Recently, Van der Hoeven presented an emerging approach to emulation called *Modular emulation* in [55].

The evaluation and plan definition described in this article is entirely independent of the type of preservation action applied; migration and emulation can be evaluated and specified within the same workflow [18].

While migration and emulation perform the primary action functionality of rendering or converting objects, *characterisation* tools are needed to analyse and describe the content. Tools such as the Digital Repository Object Identification tool (DROID)[2] and JHove[3] perform file format identification, validation and characterisation of digital objects. The extensible characterisation languages [8] perform in-depth characterisation and extract the complete informational content of digital objects.

2.2 Trustworthiness in digital repositories

Trustworthiness as a fundamental issue has received considerable attention [13,44,46]. Establishing a trusted and reliable digital archive should increase the confidence of producers and consumers. Producers need to be able to trust in the long-term preservation and accessability of their digital resources held in the repository. On the other side, users need to have confidence in the reliability of the repository and the authenticity of its content.

Institutions have started to declare their repositories as 'trusted digital repositories' or as 'OAIS-compliant'. These claims of trustworthiness or compliance are made quickly. However, verifying them objectively is much more complex.

In 2003, RLG and the National Archives and Records Administration founded a joint task force to address digital repository certification. The task force developed criteria for long-term reliable digital repositories. In 2007, the Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) report was published [53]. The criteria checklist was created based on current best practices. It is currently undergoing an ISO certification process via ISO TC20/SC13.[4] It

[1] http://www.planets-project.eu.

[2] http://droid.sourceforge.net.

[3] http://hul.harvard.edu/jhove.

[4] http://www.dcc.ac.uk/tools/birds-of-a-feather/.

deals with the organisational and technical infrastructure for trustworthy repositories and covers capabilities of certification for repositories.

Among others, it defines criteria in several aspects that are of specific interest for preservation planning. These include

- Procedures, policies and their evolvement
- Review and assessment
- Documented history of changes
- Transparency and accountability
- Monitoring and notification

In Europe, the Catalogue of Criteria for Trusted Digital Repositories [13] published by the certification working group of NESTOR identifies criteria which facilitate the evaluation of digital repository trustworthiness. Of particular relevance are aspects such as long-term planning, change mechanisms and the definition of the significant properties of the digital objects that shall be preserved.

In contrast to these prescriptive criteria catalogues, the Digital Repository Audit Method Based on Risk Assessment[5] (DRAMBORA) supports the self-assessment of a digital repository by identifying assets, activities and associated risks in a structured way. It adapts standard risk assessment principles and tailors them to digital repository assessment.

The relation of the planning approach described here to the criteria catalogues, and how the contained prescriptive criteria are supported, is discussed in Sect. 5. DRAMBORA, on the other hand, can be applied to analyse and verify the risks that apply to preservation planning activities within an organisation and can thus support the ongoing improvement and implementation within an organisation.

2.3 Component evaluation and selection

In principle, the selection problem in digital preservation can be seen as a domain-specific instance of the general problem of component selection [45]. The field of component selection has received considerable attention in the area of Software Engineering, and a number of approaches have been proposed. A comprehensive overview and comparison of methods is provided in [28,34,38].

The Off-the-Shelf-Option (OTSO) [31,32] was one of the first methods proposed. It defines a repeatable process for evaluating, selecting and implementing reusable software components. OTSO relies on the Analytic Hierarchy Process (AHP) [49] to facilitate evaluation against hierarchically defined criteria through series of pairwise comparisons. Other selection methods include 'COTS-Based Requirements Engineering' [1] and 'Procurement-Oriented Requirements Engineering' (PORE) [41]. Most selection methods follow a goal-oriented approach [56]. The process they are following can most often be abstracted to what Mohamed calls a 'General COTS selection process' [38], a procedure with the steps *Define criteria, Search for products, Create shortlist, Evaluate candidates, Analyze data and select product*.

Ncube discusses limitations of multi-criteria decision making techniques such as Weighted Scoring Methods (WSM) or the AHP, which are often used in COTS selection [40]. WSM has earned criticism for the need to determine criteria weights in advance, before the evaluation values of alternative solutions are known. AHP is considered problematic because of the considerable complexity and effort that is needed for the pairwise comparison of large numbers of criteria [36,40,43].

Franch describes quality models for COTS selection based on the ISO/IEC 9126 quality model [22] in [16]. Carvallo discusses experiences with applying these models [11]. He further proposes a method called RECSS which combines the quality models with an evaluation process [12].

These general component selection approaches are designed to be applied across a wide range of different domains. They are intended to be usable for selecting a large variety of tools that may differ significantly in terms of functionality offered, and the primary situation of applying them is when *building* component-based systems. The natural assumption is that decision makers using them have considerable knowledge and experience in software and requirements engineering.

The generality of these methods, combined with the complexity of applying them, implies that they are not truly supportive in the particular decision making processes in digital preservation.

On the one hand, decision makers in digital preservation are in general not experts in Software Engineering. The component selection in digital preservation is a recurring process that often takes place *after* the surrounding system has been put to work.

On the other hand, the functionality of the tools is very constrained; often, the only function of importance is *render* (*object*) or *convert* (*object*). Thus, the homogeneity of the evaluation problem in DP provides opportunities to leverage economies of scale in component selection procedures. Specifically, it makes it possible to rely on controlled experimentation and automated measurements to support the empirical evaluation of candidates. The proposed approach for preservation planning thus contains a domain-specific COTS component selection procedure specifically tailored towards the needs of digital preservation.

We will discuss the principal issue of what a preservation plan actually entails in the next section, before outlining the Planets preservation planning approach in Sect. 4.

[5] http://www.repositoryaudit.eu/.

3 What is a preservation plan?

3.1 A pragmatic definition

An important distinction has to be made between concrete preservation *plans* and high-level *policies* which are generally made at an institutional level and regulate fundamental constraints and strategies.

There is a number of documents available which lay out policies for digital preservation. The erpanet policy tool supports policy definition on an institutional level [14]. The recently published JISC funded study on digital preservation policies outlines a model for digital preservation policies with the aim of helping institutions develop appropriate digital preservation policies [3].

The 'ICPSR Digital Preservation Policy Framework'[6] defines high-level factors and makes the institution's commitment explicit. The British Library's Digital Object Management team has defined a preservation plan for the Microsoft Live Book data, laying out the preservation policies for digitised books.[7] It defines high-level responsibilities and certain formats which are subject to continuous monitoring, but does not specify actionable steps. The self-assessment tool developed at the Northeast Document Conservation Center[8] aids in preservation planning, however at a similarly high conceptual level.

These documents define abstract, high-level policy concerns. While they provide very useful and important guidance, they are more setting a framework for concrete planning than actually providing actionable steps for ensuring long-term access.

Examples of policy elements that are covered include 'Preservation action must be open source' and 'Cost of preservation action must not exceed estimated value of digital object'.

A preservation plan, on the contrary, is seen on a more specific and concrete level as specifying an *action plan* for preserving a specific set of objects for a given purpose. For reasons of traceability and accountability, this also needs to include the reasons underlying the decisions taken. We thus rely on the following definition, which has been adopted by the Planets project [21].

> A preservation plan defines a series of preservation actions to be taken by a responsible institution due to an identified risk for a given set of digital objects or records (called collection). The Preservation Plan takes into account the preservation policies, legal obligations, organisational and technical constraints, user requirements and preservation goals and describes the preservation context, the evaluated preservation strategies and the resulting decision for one strategy, including the reasoning for the decision. It also specifies a series of steps or actions (called *preservation action plan*) along with responsibilities and rules and conditions for execution on the collection. Provided that the actions and their deployment as well as the technical environment allow it, this action plan is an executable workflow definition.

3.2 Elements of a preservation plan

A preservation plan thus should contain the following elements:

- Identification,
- Status and triggers,
- Description of the institutional setting,
- Description of the collection,
- Requirements for preservation,
- Evidence of decision for a preservation strategy,
- Costs,
- Roles and responsibilities, and
- Preservation action plan.

We will discuss these elements in detail in the following.

3.2.1 Identification

A preservation plan should be uniquely identified so that it can easily be referred to and retrieved.

3.2.2 Status and triggers

The status of a plan includes both the planning progress—whether a plan is currently being defined, awaiting approval, or already has been deployed and is active—and the triggers which have led to the definition or refinement of the plan.

Specifically, the following events may trigger a planning activity and should thus be included in the documentation of the plan.

- *New collection*. This is the most common event, where a preservation plan is created from scratch for a new collection, for which no plan was previously defined.
- *Changed collection profile*. Changes in the collection profile of an existing collection may require a revision of an established preservation plan. Examples for changes in the collection profile are newly accepted

[6] http://www.icpsr.umich.edu/DP/policies/dpp-framework.html.

[7] http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresmicro.pdf.

[8] http://www.nedcc.org/resources/digital/downloads/DigitalPreservationSelfAssessmentfinal.pdf.

Table 1 Alerts, triggers and events

Alert	Triggered by OAIS functional entity	Event (examples)
New collection	Administration	Agreement for a new collection
	Monitor Designated Community	New object type in use
		Frequent submissions of unanticipated formats
Changed collection profile	Monitor Designated Community	Use of a new version of an object format in the designated community
		Frequent submission of unanticipated formats or new versions of an object format, or objects with new functionality/characteristics
	Manage System Configuration (in Administration)	Collection grows faster than initially foreseen and specified in the existing preservation plan
Changed environment	Monitor Technology	Change in the results of the evaluation of objectives of an existing preservation plan, for example price changes or changed risk assessment
		New available preservation strategies, for example new versions of tools and services
		Impending obsolescence of used technology, for example when a target format used in a migration-based preservation plan is becoming obsolete
	Monitor Designated Community	Change of software available at user sites (e.g. indicated by reports about problems with DIPs)
Changed objective	Monitor Technology	New standards that have to be adopted
	Monitor Designated Community	Change in computer platform or communication technologies used
	Manage System Configuration (in Administration)	Change in designated community of consumers or producer community
		Change of institutional policies
Periodic review	Develop Packaging Design and Migration Plans	Raised on a scheduled basis defined in the institutional policy or in the preservation plan

object formats or significant changes in the collection size.

It is the responsibility of technology watch functions to ensure that these triggers are actually fired; the corresponding events should then be recorded in the planning documentation.

– *Changed environment*. The environment of a preservation plan consists of the technical environment, the designated communities and the host institution. Changes in the environment can lead to a change in preferences, for example with respect to the system context in which a preservation action needs to operate. They might also imply a change in factors which influence existing preservation plans, for example changed prices for hardware or software. Other relevant changes are the availability of new preservation strategies or impending obsolescence of object formats which are used in an existing plan. Changes in the environment require a revision of existing preservation plans, while the objectives for the evaluation usually will remain unchanged.
– *Changed objective*. Changes and developments in the environment can change the objectives for preservation evaluation over time. In this case, it is necessary to evaluate existing preservation plans against changed objectives. Examples are changes in high-level policies or legal obligations that have an impact on preferences and objectives. Changes in the designated community, such as the type of software available to the users or new ways of using the objects of interest, may also affect the goals and objectives.
– *Periodic review*. Periodic reviews of existing preservation plans are needed to verify the appropriateness of plans, and to improve and further develop existing plans. A periodic review, e.g. every 3–5 years, should re-iterate the planning activity taking into account new developed preservation strategies, and seek to verify and potentially improve established plans.

Examples for these triggers, as well as the OAIS functional entities raising them, are provided in Table 1. A more detailed discussion of the interaction between the planning process and various OAIS functional entities is provided in Sect. 5.

Complementary to this documentation of recorded events that triggered an activity, the completed preservation plan also contains a specific definition of events that should trigger

a revision of the preservation plan. This enacts a monitoring of those aspects of the environment that are considered to be of particular relevance or particularly prone to change.

Examples of such aspects of interest include new versions of object formats that are included in the plan or a change in their risk assessment; changes in the support of technical environments that are used; changes in prices of software tools or services that are used; or a changed availability of tools for preservation action or characterisation. The above events should be continually monitored after the plan has been specified, and might lead to a re-evaluation of potential actions and a potential update of the preservation plan prior to the next periodic review, which should also be scheduled.

This section of the preservation plan further contains several key dates and relations to other plans, which normally are referring to the events discussed above. This includes

- *Valid from* defines the date on which the plan becomes active.
- *Based on* identifies a preservation plan on which the plan is based. This could for example be a plan that was overridden because of a changed objective.
- *Replaced by*, *Replaced on date* and *Invalidated on date* are the corresponding counterparts which create a bi-directional reference between related preservation plans.
- *Approved by* and *Approved on* document the responsible approval of the plan.

3.2.3 Description of the institutional setting

This part documents the reference frame of the preservation plan, the main context in which the planning activity takes place and in which the plan needs to be functional. Thus, it needs to cover a broad range of high-level influence factors that have an impact on the decisions taken in defining the plan.

Prime examples of aspects that are considered essential in this context include

- the *mandate* of the repository, e.g. the mission statement of the organisation;
- a description of the *designated community* for the considered collection; and
- references to applying legal, operational, and preservation policies.

Further aspects of interest are, for example

- a description of relevant organisational procedures and workflows;
- references to contracts and agreements specifying preservation rights; or
- references to agreements of maintenance and access.

The foundation for a thorough description of the institutional setting is a clear understanding of the institution's designated user community and policies, as both are important parameters for decisions throughout the preservation planning process. A detailed usage model which describes how users work with their collection and which priorities they have supports the specification of requirements and brings to light the users' priorities. Policies describe how the institution is carrying out its mandate and define organisational characteristics and goals of the repository. Particular policies may also constrain the range of potential preservation actions to be considered.

3.2.4 Description of the collection

The collection is the set of digital objects or records for which a preservation plan is created. It can be technically homogeneous (e.g. one file format), but might also consist of different types of objects or file formats. It can also be based on a *genre* in the sense of 'all emails in my repository'.

Technically speaking, it refers to all the objects that shall be treated with the same tool with identical parameter settings during the application of preservation actions.

This includes

- an identification of the objects that shall be preserved, such as IDs pointing to a repository or a unique name identifying the set of objects;
- a description of the *type of objects* mentioning general characteristics such as the contained class of objects and the file format(s); and
- *sample objects* that are representative for the collection and, thus, can be used for the evaluation process. This should include the actual objects and a description of their well-understood properties as well as their original technical environment.

3.2.5 Requirements for preservation

This section shall describe as detailed as possible the requirements that are underlying all preservation planning decisions.

Relevant requirements include a specification of the significant properties of the objects under consideration, to ensure that the potential effects of applying preservation actions are evaluated against the clearly specified aspects of objects and potential impacts are considered during the decision process.

They will usually also cover aspects such as desired process characteristics, cost limits that need to be taken into account, or technical constraints that have to be considered. Potential requirements and a specific approach of defining these in a hierarchical form are discussed in detail in Sect. 4.2.1.

3.2.6 Evidence of decision for preservation strategy

Evidence plays an essential role in establishing trust in digital repositories; evidence-based decisions and proper documentation foster transparency and support the building of trust [46,53]. This section is thus considered vital to guarantee and document that an accountable decision has been made.

The following elements are considered necessary to establish a chain of evidence that enables accountability and the tracing of decisions to link them to influence factors and assess the impact of changes further on.

- A *list of alternative actions* that have been closely considered for preservation. This should include the selection criteria that were used for narrowing the list of alternatives down from the total set of available approaches to a 'shortlist'.
- *Evaluation results* that take into account how the considered alternatives fulfill the specified requirements and document the degree of fulfillment as objectively as possible.
- A documented *decision* on what preservation strategy will be used, including the reasons underlying this decision.
- A documentation of the *effect* of applying this specific action on the collection, explicitly describing potential information loss.

3.2.7 Costs

This section specifies the estimated costs arising from the application of this preservation plan. A quantitative assessment relying on an accepted cost model such as LIFE2 [2] is desirable.

3.2.8 Roles and responsibilities

This section specifies the responsible persons and roles carrying out, monitoring and potentially re-evaluating the plan.

3.2.9 Preservation action plan

The preservation action plan specifies the concrete actions to be undertaken to keep the collection of digital objects alive and accessible over time.

A preservation action might be just the application of a single tool to a set of objects, but can also be a composite workflow consisting of multiple characterisation and action services. In this sense, the preservation action plan specifies two main aspects: the *When* and the *What*.

- Triggers and conditions specify when the plan shall be executed, as well as specific hardware and software requirements and other dependencies.
- The *executable preservation plan* specifies the actions that will be applied to the digital objects and should also include automated mechanisms for validating the results of the actions, i.e. automated quality assurance, wherever possible. The concrete elements of this part depend on the system architecture of the target environment where it shall be deployed. It can, for example be an executable web service workflow deployable in the Planets environment [29].
- Other actions needed might include reporting and documentation of the steps performed.

This section described the main components of a preservation plan as currently defined in the Planets project. The next section describes a systematic method of defining preservation plans that conform to this structure through a repeatable and transparent workflow that supports the automated documentation of decisions.

4 Systematic preservation planning

4.1 Introduction

We have discussed which aspects should be covered by preservation plans as opposed to general policies, and described the desirable components of a preservation plan. What is clearly needed is a method of specifying, monitoring and updating these preservation plans in a transparent, accountable and well-documented way. Such a method should support and streamline the information-gathering process that is necessary for informed decision making, and provide an understandable and repeatable form of reasoning, based on true evidence. It should not neglect existing work done in similar areas such as COTS component selection, but take into account the specific peculiarities applicable in the described context.

This section proposes such a method which has been developed in the course of the Planets project. It is based on earlier work done in the DELOS project[9] which has been revised and extended. The next section describes the primary planning workflow for evaluating potential actions and specifying concrete preservation plans.

4.2 Preservation planning workflow

In the previous section, we identified two key issues to be addressed by a preservation planning workflow: Evaluating potential actions and specifying concrete steps to be taken.

[9] http://www.dpc.delos.info/.

The selection of the most suitable component in a situation of complex constraints and multiple objectives is a multi-criteria decision making problem which has been discussed at lengths in the literature of software and requirements engineering. While the existing approaches cannot be beneficially applied directly to the problem at hand, several analogies and observations are useful, and they support the definition of a selection procedure in digital preservation.

The evaluation procedure described below follows a goal-oriented approach [56] and conforms to the 'General COTS selection process (GCS)' [38], an abstract procedure with the steps: *Define criteria, Search for products, Create shortlist, Evaluate candidates, Analyze data and select product*.

Based on the product selection, a concrete plan is defined, which corresponds to the definition discussed in the previous section.

The resulting workflow thus consists of four phases:

1. Define requirements
2. Evaluate alternatives
3. Analyse results
4. Build preservation plan.

Figure 1 illustrates the preservation planning environment, putting the high-level workflow in the context of the main environment factors to which it relates. The four phases result in a working preservation plan that can be continually executed. An ongoing monitor function is necessary to ensure the ability to adapt to detected changes in either the environment, the technologies used in operations or changing objectives. This results in a continuous circle of revisions to preservation plans and enables the repository to react accordingly to the inevitable changes to be expected. Figure 2 shows the concrete steps within this high-level workflow, which the next sections will discuss in detail.

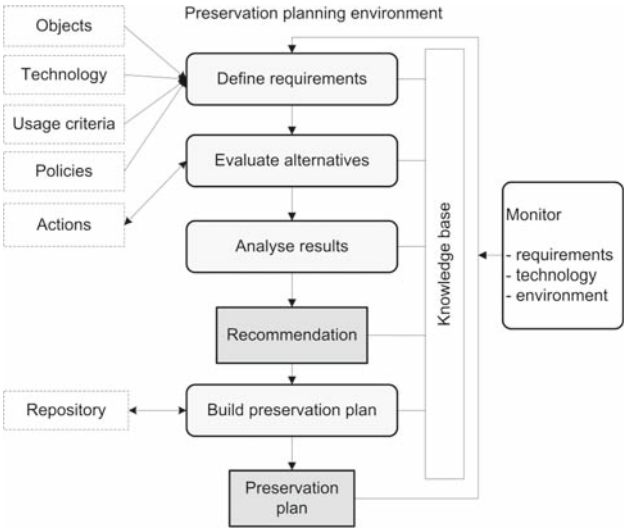

Fig. 1 Preservation planning environment

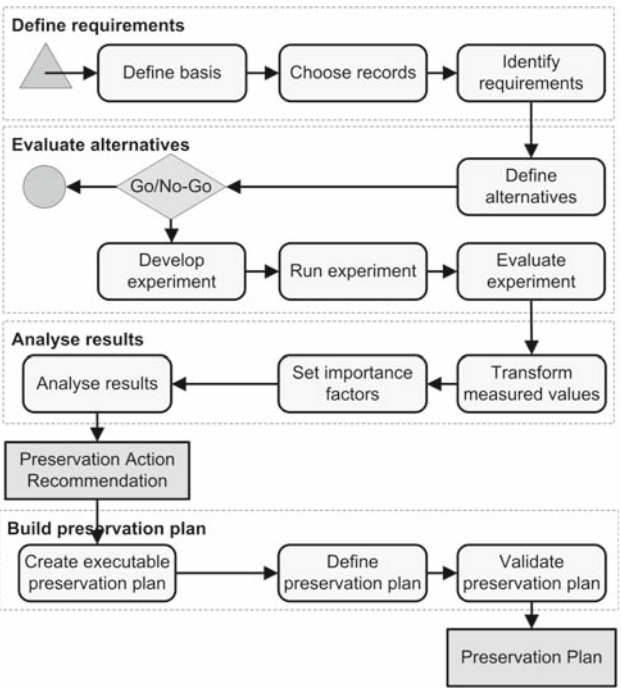

Fig. 2 Preservation planning workflow

4.2.1 Define requirements

The first phase of the workflow lays out the cornerstones of the planning endeavour. It starts with collecting and documenting the influence factors and constraints on possible actions and procedures, then describes the set of objects under consideration and finally defines the complete set of requirements to be taken into account.

Define basis: The first step of the procedure documents the main elements underlying the planning activity. It collects and documents the primary influence factors constraining the decision space and, thus, lays the foundation for a thorough documentation and makes sure that all relevant aspects are established and considered. This covers the Sects. 3.2.1–3.2.4 of the preservation plan as described in Sect. 3.

Experience has shown that a comprehensive definition of influence factors is an important prerequisite for successful planning. The documentation of constraints that might limit the choice of applicable options in this stage simplifies and streamlines the selection procedure and ensures that the outcome is indeed in line with the needs of the institution.

In this step, the preservation planner documents applying institutional policies, legal regulations, and usage criteria that might affect planning decisions for preservation. This may happen in an unstructured form, but preferably these factors are captured in a more formal way making it easier to derive decisions in the respective workflow steps. Examples include policies defining permitted file formats for

ingest, and policies related to intellectual property rights and legal access regulations. Further important policy elements pertain to characteristics of the preservation action, whether preservation actions that are open source shall be preferred or whether just a specific class of preservation action may be applied, such as emulation. The latter can occur in cases where the institution does not have the copyright and thus any modifications of the digital objects are prohibited.

Furthermore, the event that led to the planning procedure is documented. As described in Sect. 3, planning can be triggered by a new object type that is accepted, or a change in collection profiles, objectives or the environment.

Choose records: The second step describes the set of objects that forms the scope of the current plan, and selects a subset of representative objects for experimentation, as required in Sect. 4 of the preservation plan.

A general description of the characteristics of the set of objects, called *collection*, includes basic properties such as the size of the collection, the class of objects and the object formats they are currently represented in. While this can be done in a manual descriptive way, a formal representation is desirable. Collection profiling tools can provide automated descriptions of the technical characteristics of objects. An example of such a profiling service is described in [10].

Characteristics of interest include not only object formats, file sizes and their variation within the collection, but also aspects such as an assessment of the risks of each object type and each object, thus leading to a risk profile of the collection.

As a complete evaluation of the quality of preservation action tools is infeasible on the potentially very large collection of objects, the planner selects representative sample objects that should cover the range of essential characteristics present in the collection at hand.

In order to reduce effort to a minimum, this subset should be as small as possible. However, the sample objects are used as a representative set for testing the effects of applying preservation actions to the whole set of objects. A complete and thorough evaluation of the quality of preservation actions relies heavily on the completeness of features present within the test set. Thus it needs to be as large as needed to cover the variety of essential characteristics on both a technical and an intellectual level.

Depending on the degree of variance within the collection, typically, between 3 and 10 sample objects are selected. For these samples, an in-depth characterisation is performed, describing the significant properties and their technical characteristics such as their name and provenance, the file format and specific risk factors.

Identify requirements: Requirements definition is the heart of preservation planning. It is the basis for the decisions to be taken, and documents the priorities and preferences of the institution. This step enlists all requirements that the optimal digital preservation solution needs to fulfill, as in Heading 5

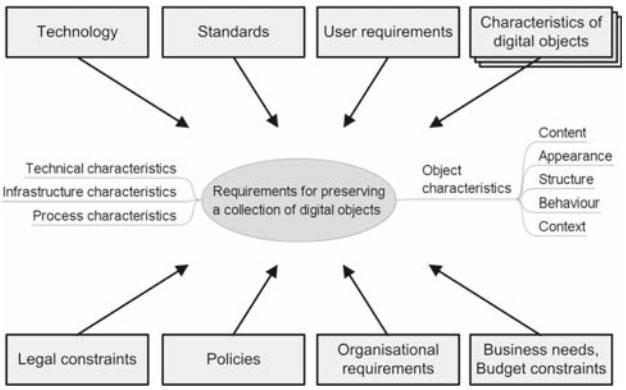

Fig. 3 Influence factors

of the preservation plan (cf. Sect. 3.2.5). Requirements are collected from the wide range of stakeholders and influence factors that have to be considered for a given institutional setting. This may include the involvement of curators and domain experts as well as IT administrators and consumers. The requirements are specified in a quantifiable way, starting at high-level objectives and breaking them down into measurable criteria, thus creating an *objective tree* which forms the basis of the evaluation of alternative strategies.

Figure 3 shows the root levels of such a tree, together with the factors that are influencing the requirements definition. Some of these high-level factors have been documented in the first two steps; in this step, they are informing the concrete specification of the objective tree.

Requirements definition has proven to be the most critical and complicated stage of the planning procedure. An incomplete requirement specification leads to a skewed evaluation and potentially wrong decisions. On the other hand, curators tend to exhibit a reluctancy to quantify their preferences, and especially try to avoid questions such as *What is the loss I am willing to accept?* which are of central importance.

The complexity involved in specifying goals and breaking them down to concrete, quantifiable criteria is a considerable challenge. However, through iterative refinement of abstract goals, such as *I want to preserve these objects exactly as they are*, towards more concrete requirements (*The size needs to remain unchanged*), we ultimately arrive at measurable criteria such as *The image width, measured in pixel, needs to remain unchanged*. This procedure benefits from a broad involvement of stakeholders to elicit all necessary pieces of information, to correctly document institutional policies and priorities and to establish constraints. A common approach is to elicit the requirements in a workshop setting where as many stakeholders as feasible are involved, moderated by an experienced preservation expert. This involvement has to avoid skewed decision priorities incurred by dominant stakeholders and needs to be managed carefully in the beginning by an expert responsible for modelling the

requirements in the objective tree. As an organisation is successively repeating the planning procedure for different types of objects, it is gaining expertise and experience and accumulating known constraints. These are documented in its knowledge base, and the need for constant stakeholder involvement gradually declines.

It is, of course, also possible to perform the elicitation of requirements in a sequential order, having all individual stakeholders list their specific requirements individually, and then integrate them in to a single objective tree. However, we have so far found the joint elicitation very helpful in current case studies, as different aspects raised by some stakeholders sometimes lead to a better understanding of the various characteristics of the objects as well as the preservation process and forms of usage. Note, that—contrary to conventional requirements elicitation, where trade-offs between requirements are defined in such a workshop setting—this is not the case here, as the focus is on complementary requirements and views on the preservation process and the objectives it shall meet. Trade-offs and weightings are performed in the third stage of the process (cf. Sect. 4.2.3).

On a practical level, two tools have been very useful for the requirements elicitation process: sticky notes and mind-mapping software. While sticky notes and flip charts as classical tool to support brainstorming activities have the benefits of allowing everyone to act at the same time, mind maps provide the better overview of the current state of requirements for all participants and allow a moderator to channel the discussion process. Often, a combination of both tools is the most productive approach. Using these tools, the requirements are structured in a hierarchical way, starting from general objectives and refining them via more specific requirements to arrive at measurable criteria that a successful digital preservation solution has to meet. This structure is further referred to as the 'objective tree', i.e. a tree capturing the objectives to be met.

While the resulting objective trees usually differ through changing preservation settings, some general principles can be observed. At the top level, the objectives can usually be organised into four main categories—characteristics of the objects, the records, and the process, and requirements on costs.

- *Object characteristics* describe the visual and contextual experience a user has when dealing with a digital object. These characteristics are often referred to as *significant properties*. A common way of describing them is to consider the five aspects 'Content,' 'Context,' 'Structure,' 'Appearance' and 'Behaviour' [48].
  Figure 4 highlights an example of specifying desirable *transformation of behaviour* when preserving a web archive on a national scale. The tree contains the requirements for preserving a collection of static web pages containing documents and images. The branch *Behaviour* is divided into three different groups of criteria: *deactivate, preserve* and *freeze*. This reflects the preferences of the archive that some functionality, such as menu navigation, is needed for properly accessing the web pages, while most active content shall be disabled or frozen. For example, visitor counters shall be preserved in the state they had at the moment of ingest, rather than preserving their activity and to continue counting within the archive. (This scenario may well be of interest for a different designated community of, e.g. internet historians who want to analyse the technical principles of how counters were implemented in earlier days.)
  Recently, several projects such as INSPECT.[10] have presented detailed analysis of the significant properties of different categories of objects including vector images, moving images, e-Learning objects, and software [19].[11] These can provide a very valuable input to this aspect of requirements specification. The automated characterisation of the sample objects defined in the previous step further supports the analysis of their significant technical properties.
- *Record characteristics* describe the foundations of a digital record, the context, interrelationships and metadata. This may include simple, but often overlooked, linking requirements, such as the fact that filenames need to remain unchanged or consistently renamed across subcollections if they form the basis for cross-referencing or inclusion.
- *Process characteristics* describe the preservation process itself, for example the procedure of migrating objects. These characteristics include not just the complexity of applying preservation action tools or their performance, scalability, and usability, but should equally cover aspects such as documentation or the degree of validation.
  The definition of process characteristics is particularly dependent on the specific context in which the preservation process is taking place. The technical environment may effectuate specific requirements on the interoperability of tools, while institutional policies or legal regulations may enforce specific licensing requirements or require a particular degree of automated documentation. Thus, the institutional and technical context and constraints posed by it have to be considered carefully.
- *Costs* have a significant influence on the choice of a preservation solution, but are inherently hard to quantify. Ultimately, the Total Cost of Ownership (TCO) is the guiding figure for deciding whether or not a preservation strategy

[10] http://www.significantproperties.org.uk/about.html.

[11] http://www.jisc.ac.uk/whatwedo/pro\discretionary-grammes/preservation/2008sig\discretionary-props.aspx.

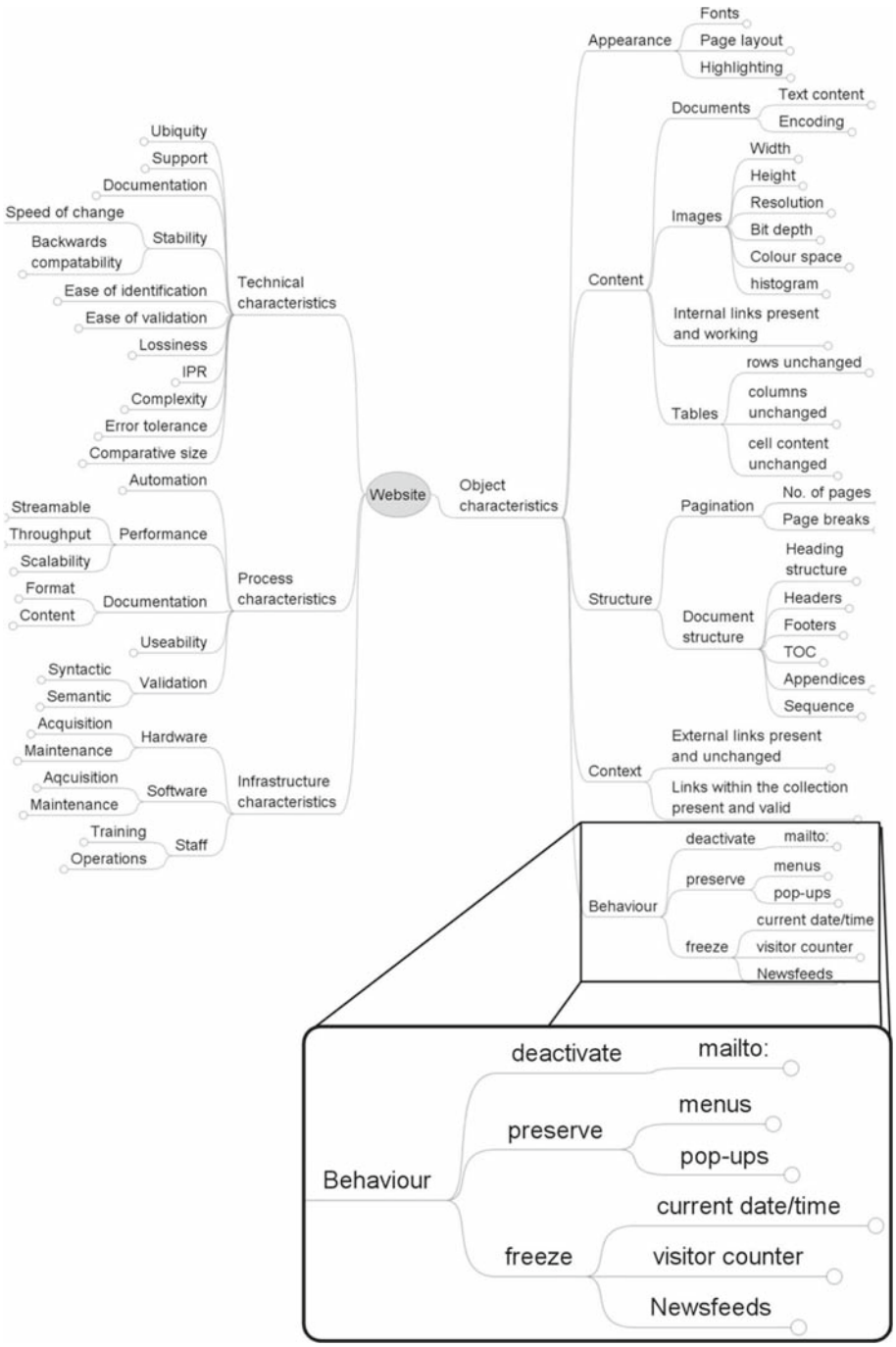

Fig. 4 Requirements specified in an objective tree

meets the needs of an institution within the constraints of its budget. Instead of providing a single numeric criterion which is extremely complex to quantify, costs might also be defined as *infrastructure characteristics*, putting an emphasis on cost factors instead of the resulting figures for cost estimates. These cost factors can, then, be further broken down to cover hardware, software and staff costs, as shown in Fig. 4.

An essential step of requirements definition is the assignment of measurable effects to the criteria at the leaf level of the objective tree. Wherever possible, these effects should be objectively measurable (e.g. € per year, frames per second, bits per sample) and thus comparable. However, in some cases, (semi-) subjective scales need to be employed. For example, the quality of documentation that is available for a file format or a tool should not be judged by the number

of pages alone; instead, a subjective scale such as *excellent, good, average, poor, very poor* could be used. Similarly, the *openness* of documentation of a file format could be one of *fully standardised*; *openly published*, but not standardised by a recognised body; and *proprietary*. Along the same lines, the *stability* of a format can be measured in revision time intervals and backwards compatibility.

The assignment of measurable effects to criteria can also align them with characteristics that can be automatically extracted from objects to automate the evaluation procedure. Existing software tools such as JHove[12] allow automated extraction of some of the basic properties of common object formats; the eXtensible Characterisation Languages provide an in-depth description of the complete informational content of an object in an abstract representation [8]. These descriptions can be used to derive properties to be measured, and support the automated comparison of these properties when migrating the objects to different formats.

Related to the categorisation of requirements presented above, a distinction can be made between binary criteria which must be fulfilled and gradual factors that need to be balanced against each other. Significant properties of digital objects are most frequently seen as binary criteria that are either preserved or not, and where usually no loss can be tolerated. On the other hand, two preservation actions might both keep all essential characteristics and thus be acceptable. The decision then can take into account gradual factors such as the total costs incurred by each alternative action, processing time or the assessment of risks that are associated with each alternative. These factors cannot be measured on binary scales. Our approach of tailored utility analysis unifies both kinds of criteria by allowing different scales to be used for the actual measurements of the respective criteria. In the third phase, these measurements are transformed and, thus, made comparable through the definition of transformation rules, which calculate unified utility values based on the knowledge gained in the experiments. In the final step, binary criteria can be used to filter alternatives, while the weighted overall performance across all criteria is then used for the final selection of the best action.

The objective tree, thus, documents the individual preservation requirements of an institution for a given partially homogeneous collection of objects. The tree as such is entirely independent of the strategy employed, be it migration, emulation or another [18]. It is of vital importance that it is concerned solely with the *problem space*, and does not specify solutions such as *We want to migrate to PDF/A*, unless these decisions have been made already on a higher level, e.g. an institutional policy.

While such specifications are sometimes brought forward in the requirements workshops, they commonly can be traced back to the reasons underlying them, such as preferences for transforming objects to standardised, widely supported file formats and deactivation of active content. The decision to migrate to PDF/A using a specific tool might be the right one; however, without proper documentation of the reasons and the evaluation leading to it, the recommendation cannot be considered trustworthy.

The tree shown in Fig. 4 contains a branch named *technical characteristics*. In this specific case, the institutional policy constrained the class of preservation action to be considered to migration; emulation was not an option. Thus, the requirements describe in a very specific form the desired characteristics of the target format in which the objects should be kept. These characteristics together form a *risk assessment* of the format and become a central part of evaluating applicable tools and strategies.

A series of case studies have been conducted where objective trees were created for different settings. Examples include electronic publications in a national library [5]; web archives and electronic documents with national archives [50]; interactive multimedia in an electronic art museum [4]; and computer video games [18].

Ongoing case studies revise and extend the previously conducted evaluation studies, build concrete preservation plans for specific collections of objects and cover new scenarios that have not been evaluated yet, such as database archiving, in a variety of institutional settings.

The experience which is accumulated through carrying out planning activities and requirements definition can be easily shared between institutions through the supporting software, which contains a knowledge base of recurring fragments of objective trees and templates that can be used as a starting point, as described in Sect. 6.

The outcome of the first phase is a complete documentation of the planning context, the collection of objects at question and the specific requirements that form the basis for the evaluation of alternative action paths.

4.2.2 Evaluate alternatives

The second phase of the planning workflow relies on controlled experimentation. It evaluates potential actions in a quantitative way by applying them to the previously defined sample content and analysing the outcomes with respect to the requirements specified in the objective tree. This empirical evaluation procedure results in an evidence base that underlies the decisions to be taken in the successive phases. It basically provides all information for Sect. 6 of the preservation plan (cf. Sect. 3.2.6).

Define alternatives: The natural first step of evaluation is to define the possible courses of actions to be taken into consideration. A variety of different strategies might be applicable; for each alternative action, a complete

[12] http://hul.harvard.edu/jhove/.

specification of the entailed steps and the configuration of the software tool employed is desired. The discovery of potential actions that are applicable varies in complexity according to the type of content. Often, this implies an extensive search phase, investigating which tools are available to treat the type of objects at hand. Registries holding applicable preservation action tools can be consulted for reference, and are potentially very beneficial to support the search.

The outcome is a *shortlist* of potential candidates for performing preservation actions, which will be evaluated empirically during the next steps. The description of an alternative includes the tool name and version used, the operating system on which it shall run and the technical environment specification such as installed libraries and fonts.

Go/No–Go decision: Before continuing with the experimentation procedure, this step reconsiders the situation at hand and evaluates whether it is feasible and cost effective to continue the planning procedure. In cases where the evaluation is considered infeasible or too expensive, a reduction of candidate tools might be necessary. The evaluation of some tools may also be postponed due to unavailability or cost issues, or because of known bad performance. This is individually described and documented.

Develop experiment: This step sets up and documents the configuration of the tools on which experiments are carried out, and, thus, builds the basis for experiment execution in the next step. This includes setup procedures, a documentation of the hard- and software environment, and additional steps needed to carry out the evaluation of experiments, such as setup of time measurement and logging facilities.

Run experiment: In this step, all the considered candidate tools are applied to the set of sample objects that have been defined in the first phase. This produces a series of experiment results that can be analysed, and are stored as evidence. In the case of object conversion, this means that the resulting output files shall be stored for further reference. When evaluating emulators, a documentation detailing the experience of rendering the object is needed. Furthermore, any errors or logging messages occurring are documented.

Evaluate experiment: The evaluation of experiments is based on the requirements specified in the objective tree. All the criteria on the leaf level of the objective tree are evaluated, taking into account the empirical evidence resulting from the experiments conducted.

Figure 5 shows a simplified abstraction of the core elements of the requirements and evaluation model. Each *preservation action tool* is evaluated through applying it on *sample objects* in a controlled experiment. This creates an *experiment result* that constitutes part of the evidence base. A *criterion* is a measurable *requirement*. It can be associated with a tool (*tool criterion*) or varying with every object a tool is applied to (*object criterion*). In the latter case, it can be mapped to an *object property*. These properties are measured of the original *sample object* and the *experiment result*, and the obtained values are compared through a comparison metric. Tool criteria, on the contrary, are associated with a *tool property* and evaluated in a *tool evaluation*.

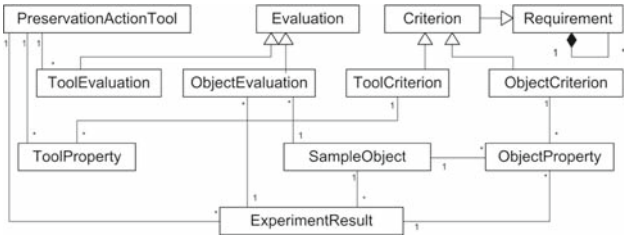

Fig. 5 Core model of requirements and evaluation

Thus, the performance of each leaf criterion is measured for each alternative and collected in the objective tree. For some objectives, this has to be done manually, while for others it can be performed automatically using characterisation tools. For example, the previously mentioned criterion *image width unchanged* is an *object criterion* which can be measured by characterisation tools such as JHove or XCL and compared automatically for each result of an experiment. Similarly, the relative file size of objects can be measured automatically per object. The relative file size averaged over the sample objects would then be used as evaluation value for the corresponding criterion.

In other cases, information might be obtained from registries or inserted manually. For example, the judgement of quality of documentation, or the degree of adoption of a file format, can be queried in registries such as PRONOM, or judged by the preservation planner. Some criteria that are tool specific rather than object specific only need to be measured once, e.g. the cost of a tool.

Documenting the evaluation of experiment results completes the empirical evidence base for decision making and concludes the second phase of the preservation planning workflow.

4.2.3 Analyse results

In the third phase, the experiment results are analysed, aggregated and consolidated in three steps.

Transform measured values: The result of the previous phase is an objective tree fully populated with evaluation values for all the criteria. However, the measurements in this tree are of varying scales and, thus, cannot be aggregated and compared directly. Thus, transformation rules are defined, which result in a mapping from all possible measurement scales to a uniform target scale, thus computing a *utility function* for each value. This scale usually consists of real numbers ranging from 0 to 5. The lowest value 0 denotes an unacceptable result, while 5 is the best possible

evaluation value. (Although other scales such as 0 to 100 may be employed, experience has shown that this scheme is very useful and comfortable to use.)

Corresponding to the scales employed, we can distinguish two types of transformation settings: numerical thresholds and ordinal mappings.

- For ordinal values, a mapping is defined for each possible category, resulting in a value between 0 and 5 to be assigned. For a boolean scale, *Yes* might be mapped to 5, whereas *No* will often be mapped to a low value. In this case, a decision has to be made whether the negative result *No* should be acceptable or not, i.e. mapped to 1 or to 0.
- For numeric values, thresholds are defined for each integer number from 1 to 5. All the numbers below the lowest threshold (or above the highest, in case of descending order) will then be transformed to 0. The calculation of values between the threshold is usually done using linear interpolation. In some cases such as costs, storage or processing time, individual thresholds or logarithmic transformation is used.

In both cases, the definition of *acceptance criteria* is an essential step, where decision makers have to clearly specify the constraints they are willing to accept. This further provides a gap analysis which clearly points out both strengths and limitations of the candidate tools under evaluation.

Set importance factors: This step takes into account the fact that not all requirements are of equal importance, and assigns weight factors to the nodes in the objective tree. There has been considerable discussion on the question of importance weighting in component selection methods. Several methods using Weighted Scoring Methods have earned criticism for the fact that weight settings need to be specified upfront, in a situation where little or nothing is known about the actual performance and differences of candidate components. Furthermore, the reduction to a single number and the corresponding ranking is considered too simplistic by many. The Analytic Hierarchy Process, on the contrary, which is used in a number of component selection approaches, is often considered too effort intensive and complex, with the number of pairwise comparison exploding with the size of the requirements tree.

In the presented approach, relative importance factors are balanced on each level of the objective tree, at a time when evaluation values for all the candidates are already known. This deviates from the standard Utility Analysis workflow, but has proven more useful in the component selection scenario represented by preservation planning in numerous case studies.

The weighting of the top-level branches of the requirements trees often depends on institutional policies and may have significant impact on the final evaluation result. In particular, preferences might have to be negotiated between the quality of preservation actions and the costs needed to setup the necessary migration or emulation software, or within the different aspects of significant properties of objects. For example, the 'behaviour' branch of an objective tree for preserving static documents will have a much lower importance weighting than in the context of multimedia art objects, where interactivity is a central aspect.

As a general rule, the acceptance criteria defined in the transformation rules should be used to model the actual evaluation values, while importance weighting is meant to reflect the overall priorities of an institution.

Analyse results: The final step of the evaluation phase considers the complete evidence base of information produced during the previous phases of the workflow. It analyses the performance of the candidate components in the experiment evaluation to arrive at a conclusion and recommendation for the best tool to be employed, and the corresponding configuration.

The measurements described above are transformed and multiplied with the weights of the corresponding requirements. This results in an evaluated objective tree where the leaf criteria have been populated. Aggregating these values leads to a performance value of each alternative action on all levels of the tree hierarchy, which is directly comparable.

Alternatives are then ranked by their root evaluation values, which are aggregated over the tree hierarchy using two different methods.

- Weighted multiplication is used to filter alternatives which exhibit unacceptable evaluation values at the criterion level, as these have been mapped to a target value of 0 during transformation and thus result in a total performance of 0.
- On the remaining alternatives, i.e. those which do not show unacceptable performance values, weighted addition is used to directly compare the performance of tools on all the levels of the tree hierarchy.

For example, a tool might preserve all the object characteristics perfectly well and be reasonably affordable in operation, but not perform fast enough or not be scalable enough for a given technical environment, i.e. its measured performance lies clearly outside the worst acceptable range of values. If reconsidering the transformation thresholds set before is not an option, then this tool cannot be considered for operation. Thus, a second tool that has some minor weaknesses, but shows no issues that are unresolvable, can be the preferred choice.

This analysis and comparison of the alternatives considered can be guided significantly by a graphical visualisation as provided by the planning tool described in Sect. 6. In order to safeguard against potentially negative effects of minor variations in the weighting, a sensitivity analysis is performed. In this step, several hundred iterations are automatically computed, where weights are randomly varied in a certain percentage margin around the given value to identify potential changes in the ranking of alternatives.

As a result of the evaluation, the preservation planner makes a decision and recommendation for a tool to be selected. The method allows for the selection of multiple components that are considered to be complementary. For example, many conversion tools for electronic documents have problems with entirely preserving the layout as it was displayed in the original environment, whereas migrating a document to an image loses the future potential for full-text search access. In some cases, it might be desirable to combine both approaches and, thus, select multiple tools for the incorporation into a preservation system.

As an essential element of the recommendation, the reasons underlying it are documented, together with the expected effects of applying this strategy on the set of objects at hand. For example, it may be known that the easy *editability* of objects will be lost as a direct cause of converting them to a format such as PDF/A. As this might not be a requirement, or not be assigned significant weight, it might not influence the decision in a significant way. However, this reasoning needs to be documented as part of the decision making procedure.

4.2.4 Define preservation plan

In the fourth and final phase of the planning workflow, a preservation plan is created, based on the decision for a preservation action. In OAIS terminology this corresponds to the *Develop Packaging Designs and Migration Plans* functionality. It specifies a series of concrete steps or actions, along with organisational responsibilities, rules and conditions for executing the preservation action on the collection. This completes the information necessary for the preservation plan as described in Sects. 3.2.7–3.2.9.

Create executable preservation plan: This step of the workflow defines the triggers for the execution and the conditions under which the preservation action will be carried out. Hard- and software requirements as well as dependencies on other systems are documented. In order to enable the execution of the preservation action, tool settings and details about the location of the collection on which the action is to be performed are defined, thus resulting in a *preservation action plan*.

In order to perform quality assurance of the executed actions, a subset of the criteria used for evaluating solutions can be selected. These criteria should then be evaluated automatically during the execution of the action plan to validate that the defined thresholds of these criteria are met. The necessary documentation that has to be recorded when performing the action is also defined in this step.

Define preservation plan: While many parts of the preservation planning workflow take care of the technical aspects of the preservation plan, this step mainly defines organisational procedures and responsibilities.

Cost factors influence the decision on a specific alternative. In this step, a more detailed calculation of costs using an approved cost model is performed. Cost models that can be used are, for example Life2 [2] or the Total Cost of Ownership (TCO)[13] model. While an estimate of the costs may be fine for evaluating the alternatives, the costs for adopting the solution have to be determined as accurately as possible in this step.

The assignment of responsibilities is also documented in this step. Monitoring the process of applying the preservation actions has to be done by a different role than executing the preservation plan. It also has to be monitored whether an event occurs that makes it necessary to re-evaluate the plan. Possible triggers for this are either a scheduled periodic review, changes in the environment such as new available tools detected through technology watch, changed objectives (such as changed target community requirements) or a changed collection profile, when objects show new characteristics diverging from the specified profile. An indication for a changed collection profile is also that values measured during the quality assurance deviate significantly from the values measured for the sample objects during the evaluation.

Validate preservation plan: In the final stage, the whole documentation on the preservation plan and the decisions taken during the planning process are reviewed. Tests on an extended set of sample objects may be performed in this step to check the validity of the preservation action plan.

Finally, the validated plan has to be approved by the responsible person. After the plan has been approved, no more changes to the plan should be done without formally revising the whole plan.

4.2.5 Summary

This section described the four-phase workflow for creating a preservation plan in detail. The final outcome is a completely specified, validated and formally approved preservation plan defining concrete steps and responsibilities for keeping a certain set of objects alive. The plan includes the complete evidence base of decision making and conforms to the plan definition discussed in Sect. 3.

[13] http://amt.gartner.com/TCO/MoreAboutTCO.htm.

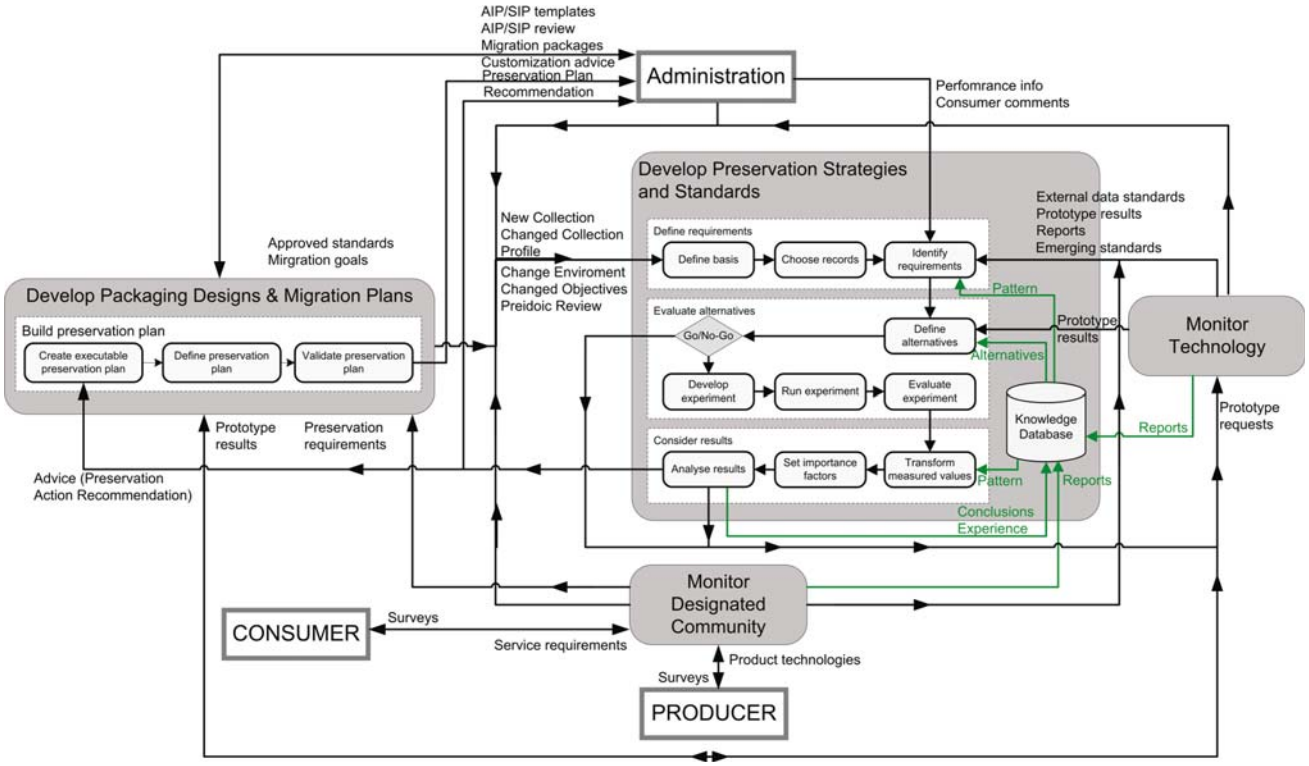

Fig. 6 Planets Preservation Planning approach within the OAIS

5 Compliance to existing models and criteria

While the previous section described the planning workflow in detail, this section puts it in context with the OAIS model and explores the relation to the functional entities that the OAIS model has described for preservation planning. We further discuss how the method presented here supports criteria for trustworthy repositories as defined by TRAC and nestor.

5.1 The OAIS model

The Reference Model for an OAIS was published in May 1999 by the Consultative Committee for Space Data Systems (CCSDS). In 2003, the OAIS Model was adopted as ISO 14721:2003 [24]. In the community of digital preservation, the OAIS model has been widely accepted as a key standard reference model for archival systems. The primary goal of an OAIS is to preserve information for a designated community over a long period of time. Figure 6 shows the integration of the planning approach within the OAIS model and the main information flows. A detailed analysis of the information flows and the planning activities are presented in [51]. The Planets Planning method implements the *Develop Preservation Strategies and Standards* and the *Develop Packaging Designs and Migration Plans* functions of the OAIS model.

The *Develop Preservation Strategies and Standards* function is responsible for developing and recommending strategies and standards to preserve the current holdings and new submissions for the future. The first three phases of the planning method evaluate different preservation strategies for a collection or new submissions as described in Sect. 4.2. The outcome is a preservation action recommendation which identifies the most suitable preservation action for a collection in a specific context. The recommendation is provided to the *Develop Packaging Designs and Migration Plans* function as advice to create a detailed migration plan in Phase 4 of the presented workflow, and to the Administration entity for system evolution.

The functional entities of the OAIS model provide possible constraints and requirements for the steps within the planning approach [51]. They can further trigger new planning activities corresponding to the events defined in Sect. 3.2.2.

- The functions *Monitor Designated Community* and *Monitor Technology* perform a watch that provides reports about developments and changes in the designated community and relevant technologies.
- The *Manage System Configuration* and the *Consumer Service* functions of the *Administration* entity report performance information of the archiving system and consumer comments to the *Develop Preservation*

Table 2 Supported criteria for trustworthy repositories

Aspect	Criterion	Artefacts and actions
Procedures and policies	TRAC A3.2 Repository has procedures and policies in place, and mechanisms for their review, update and development as the repository grows and as technology and community practice evolve	The preservation plan specifies monitoring conditions and triggers. Periodic reviews following the specified workflow lead to revisions of the plan
	Nestor 4.4 The digital repository engages in long-term planning	
	TRAC A3.4 Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements	Environment conditions to monitor are specified; periodic reviews following the planning workflow are conducted
Transparency and documentation	TRAC A3.6 Repository has a documented history of the changes to its operations, procedures, software and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content	The preservation plan contains a change history, and the evidence from controlled experiments describes potential effects on objects
	TRAC A3.7 Repository commits to transparency and accountability in all actions supporting the operation and management of the repository, especially those that affect the preservation of digital content over time	The preservation plan is fully documented and traceable. All evidence from the experiments is kept as inherent component of the plan
	TRAC B3.1 Repository has documented preservation strategies	
Monitoring	TRAC B3.2 Repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable	As part of the preservation plan, appropriate monitoring conditions are specified
	Nestor 5.3 The digital repository reacts to substantial changes	
Significant properties	TRAC B1.1 Repository identifies properties that it will preserve for digital objects	The objective tree provides a full specification of all properties considered to be significant
	Nestor 9.2 The digital repository identifies which characteristics of the digital objects are significant for information preservation	

Strategies and Standards function. These comments can imply requirements regarding access, behaviour and usage of the digital objects in the system. The performance information can, thus, raise requirements that have to be fulfilled by potential preservation strategies.

– The function *Monitor Technology* offers the functionality to evaluate emerging technologies by prototype requests. The results are first indications for closer consideration of new and untested tools and services in the step *Define Alternatives* of the planning method. The outcome of the first three phases is a recommendation for a preservation action.

These aspects are, thus, not only important during the planning workflow described above, but also form the basis of an ongoing monitoring process that is essential for successful continuous preservation management.

The *Develop Packaging Designs and Migration Plans* function is responsible for providing detailed migration plans. It uses the recommendation from the function *Develop Preservation Strategies and Standards* as a basis for building a preservation plan, incorporating organisational aspects such as the responsible roles and persons to carry out the plan. It further creates an executable preservation plan that includes mechanisms for quality assurance and capturing metadata.

5.2 Criteria for trustworthy repositories

Trustworthiness is one of the fundamental goals of every repository. This section analyses the Planets Preservation Planning approach in relation to the TRAC checklist [53] and the Nestor criteria catalogue [13,42]. Both include, among others, several criteria covering the following aspects:

1. Procedures, policies, transparent documentation;
2. Monitoring, evolvement, and history of changes; and
3. Significant properties.

The next paragraphs discuss each of these aspects, while Table 2 contains a list of specific criteria relevant to each

aspect and summarises which artefacts and actions of the planning approach contribute to the fulfillment of each criterion.

5.2.1 Procedures, policies, and transparent documentation

Well-defined policies and transparent documentation are considered essential by both TRAC and nestor. The TRAC report states that *transparency is the best assurance that the repository operates in accordance with accepted standards and practice. Transparency is essential to accountability, and both are achieved through active, ongoing documentation* [53].

The Preservation Planning approach evaluates preservation strategies in a consistent manner, enabling informed and well-documented decisions. It enforces the explicit definition of preservation requirements in the form of specific objectives. The definition of the objectives and their measurement units as well as the evaluation itself have to be as objective and traceable as possible. This complete specification of underlying policies and constraints, the collection profile, the requirements and evaluation results as well as the resulting preservation plan results in a comprehensive documentation and a reliable, accountable and transparent decision on which strategies and PA tools to deploy.

Furthermore, the software tool *Plato* which implements the planning approach supports automated documentation of the planning activities. The potential effects of preservation strategies on digital objects are described, and the history of preservation plans created, reviewed and updated with the planning method documents the operations, procedures, software and hardware used in the context of DP actions. Additional documentation, of course, needs to be provided for the general system hardware and procedures outside the preservation planning setting.

5.2.2 Monitoring and change management

The institutional policies need to define watch services for the collection and its development, and for changes in technology and the designated communities. These watch services trigger the according alerts as defined in Sect. 3.2.2 and Table 1. While reviewing the affected plans using the planning workflow, the repository is able to assess the impact of changes and react accordingly. The review verifies an implemented preservation plan, considering changes in requirements and practice or changes in the collection, and might result in an update of the preservation plan, replacing the existing plan. It supports impact assessment and reaction as environments and technology change. The accumulated history of changes and updates to preservation plans is fully documented, and provides a traceable chain of evidence.

5.2.3 Significant properties

The objective tree specifies requirements and goals for preservation solutions. The core part of it is formed by the specification of the significant properties of objects that shall be preserved. These requirements document the properties of objects that have to be preserved, and align them with automated measurement tools, if these are available. The supporting tool described in Sect. 6 provides templates and fragment trees to facilitate the tree creation. These templates and fragments are based on experiences from case studies, and enable the exchange of best practice.

6 Tool support: the planning tool Plato

As part of the Planets project, a planning tool is being developed which supports and automates the described planning workflow [7]. This software tool called *Plato* is publicly available since March 2008; the latest version 2.1 has been published in November 2009.[14] Current work is focused on providing increased automation and proactive planning support.

The software supports and guides the planner through the workflow and integrates a knowledge base for supporting the creation of objective trees by providing recurring templates and fragments that are applicable across different planning situations. It furthermore integrates a range of services to provide an automated planning environment [6,9]. This includes semi-automated service discovery for informing users about potential preservation actions that are applicable to their set of sample objects; automated identification and description of sample content using characterisation tools; and automated measurements and comparison of original and transformed objects using the XCL languages [8]. Figure 7 shows requirements specification and visual analysis of evaluation results in Plato.

The outcome of applying the planning tool is a complete preservation plan conforming to the structure specified in Sect. 3, both as PDF document and in XML. The latter contains the entire documentation of the decision making process, including the sample content as well as the complete evidence created in the experiments as evidence base. Based on the recommendation and the decision taken by the preservation planner, the planning tool automatically generates an XML based executable preservation plan, which contains a workflow that can be executed by the Planets Workflow Execution Engine [29]. The plan contains references to preservation action and characterisation services available through web service registries as described in [6], and specifies the actions to be taken on each object, such as

[14] http://www.ifs.tuwien.ac.at/dp/plato.

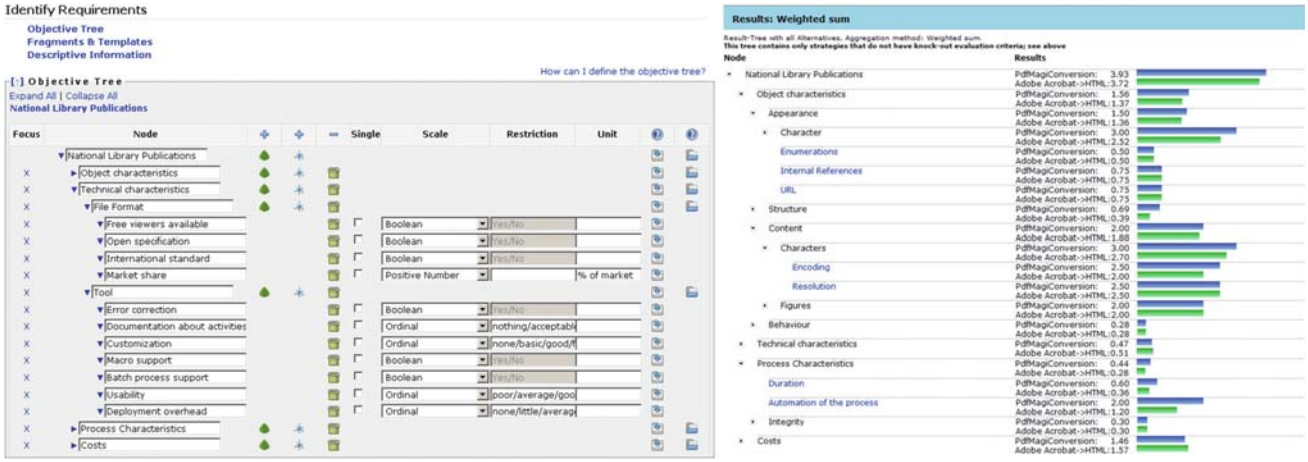

Fig. 7 Requirements specification and visual analysis of potential alternatives in Plato

1. Identify format,
2. Extract properties,
3. Migrate,
4. Identify format of converted object,
5. Extract properties from converted object,
6. Compare original and migrated object,
7. Generate report and store results.

The software is being used by several national libraries and archives around the world. It has been used in training professionals in a number of summer schools, tutorials and university curricula. Case studies have been conducted that deal with content ranging from electronic publications and web archives to interactive multimedia art, computer games, and database archiving. The next section discusses some newer case studies on large image collections.

7 Evaluation

The planning approach and the corresponding tool Plato have been applied by a number of institutions across Europe. A first round of case studies [4,5,18] led to refinements and extensions of the planning procedure and the definition of the preservation plan as it is described here. In this section, we shortly present three related new case studies and their results and discuss lessons learned over the previous years in applying the presented approach.

7.1 Evaluating preservation actions for scanned images

This section discusses three related exemplary case studies, each seeking the optimal preservation solution for large collections of scanned images in a different national institution in Europe. The first case study was carried out with the British Library[15] and focused on a collection of 2 million images in TIFF-5 format with a size of about 40 MB per image. The images were scanned from old newspaper pages.

Concerns were raised about the suitability of the LTO tapes on which the content was held, and the images were transferred to hard disk storage and reviewed. This move highlighted difficulties accessing some of the tapes, and a decision was taken to transfer the material into the main digital library system. Before the ingest, it was decided to review the format of the master files to see whether the current format was the most suitable or whether a migration should be performed as part of the media replacement.

Some of the high-level policies that affect the decision making in terms of file formats include

1. Open target formats are highly preferred.
2. Compression must be lossless.
3. Original copies may be deleted.

The requirements tree as shown in Fig. 8 is quite compact, as significant properties of images are not overly complex. A variety of options, including not changing the format of the images, were evaluated in a series of controlled experiments. The costs were calculated using the LIFE models,[16] and the evaluation was partly automated. Table 3 shows the evaluated preservation actions and their aggregated scores. Conversion to BMP was ruled out due to large file sizes and lack of compression, while GIF was discarded because of the palette size limitations.

The results show that migration to JPEG2000 achieves a slightly higher root score than that achieved leaving the master files untouched, as indicated by the bold numbers in Table 3. The reasons are that the long-term storage costs, and

[15] http://www.bl.uk.

[16] http://www.life.ac.uk.

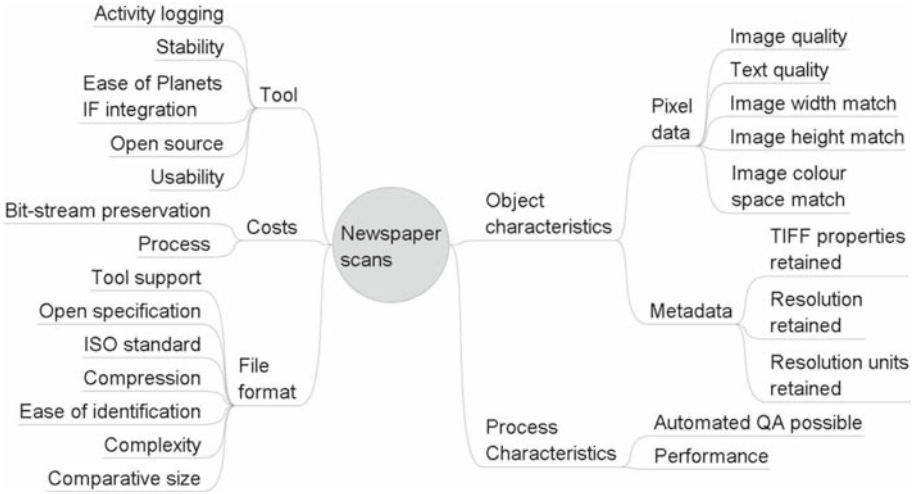

Fig. 8 Scanned newspaper requirements tree

Table 3 Evaluation results for preservation actions

Name	Weighted multiplication	Weighted sum
Leave in TIFF v5	3.01	3.46
Convert TIFF to PNG	2.72	3.27
Convert TIFF to BMP	–	–
Convert TIFF to GIF	–	–
Convert TIFF to JPEG	0.00	–
Convert TIFF to JP2	**3.44**	**3.69**
Convert TIFF to JP2 95	0.00	–
Convert TIFF to JP2 90	0.00	–
Convert TIFF to JP2 80	0.00	–

the fact that JP2 is a recognised ISO standard [26] outweigh the process costs of converting the images. Conversion to JPEG or to compressed JP2 is violating the abovementioned policy that compression must be lossless. Thus, the corresponding alternatives have a multiplication score of 0.0 and are discarded as unacceptable alternatives.

A similar study examined the options for preserving a large collection of images scanned from sixteenth-century books held by the Bavarian State Library.[17] The collection contains 21,000 prints with about 3 million pages in TIFF-6, totalling 72 TB in size. The requirements elicitation procedure involved stakeholders ranging from the head of digital library and digitisation services to digitisation experts, library employees and employees from the supercomputing centre responsible for the storage. The resulting requirements tree is shown in Fig. 9. The considered actions were migration to JP2 with various conversion tools and leaving the objects unchanged. Storage costs do not pose significant constraints on this specific collection at the moment. The evaluation results showed that leaving the images in TIFF-6 was the preferable option, despite JP2 having advantages such as reduced storage requirements and streaming support. Storage will be monitored and the decision periodically reviewed. A detailed discussion of this case study is presented in [33].

Another related study was conducted with the State and University Library Denmark,[18] evaluating the best options to preserve a large collection of scanned yearbooks in GIF format. The storage costs were not as important and the volume not as high in the previous studies, and evaluation led to the decision to migrate the images to TIFF-6 despite the growth in file size.

These cases illustrate that a preservation action that is best in one situation does not necessarily fit in another, and shows that preservation planning has to take into account the institution-specific preferences and peculiarities.

It is worth noting that while the decision might be to leave objects unchanged, this is still a complete preservation plan and vastly different from not defining any action to be taken. On the one hand, a thorough analysis is needed before taking a decision on whether to act or not; on the other hand, the preservation plan contains monitoring conditions that can trigger a re-evaluation due to changed conditions in the future. Trustworthiness as discussed in Sect. 5 requires transparent and well-documented decisions and ongoing management.

Similar case studies have been performed for other types of objects, including databases (in cooperation with the Swiss Federal Archives), legacy text documents (in cooperation with the National Archive of the Netherlands), computer games (in cooperation with the Computer Game Museum Berlin[19] with a specific focus on emulation), and electronic art (in cooperation with the Ars Electronica in Austria).

[17] http://www.bsb-muenchen.de/index.php?L=3.

[18] http://www.statsbiblioteket.dk/english/.

[19] http://www.computerspielemuseum.de/index.php?lg=en.

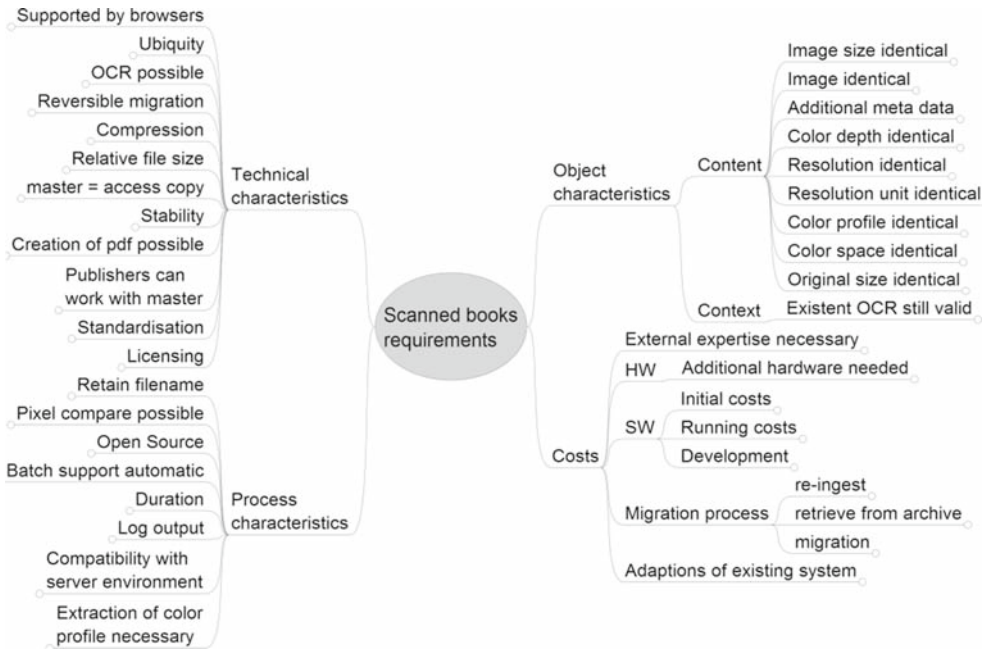

Fig. 9 Scanned book pages requirements tree

7.2 Some lessons learned

This section discusses some of the issues that arose over the past years in preservation planning case studies with a variety of different partners. While the overall acclaim is very positive and partners approve of the concepts, there are several challenges for the near future that need to be overcome.

- There is a severe lack of structured, informative and reliable information sources.
- Applying the approach of requirements specification and evaluation proved challenging.

In order to address these issues, there is a strong need for more sophisticated tool support, automated quality assessment and proactive recommendation technologies.

7.2.1 Information sources

There is a lack of well-structured information sources that can be queried and integrated automatically. While PRONOM[20] is often cited as a reference source, it does not contain sufficient levels of detail on file formats to truly support automated evaluation and risk assessment. Other sources such as the Digital Formats Website[21] represent valuable sources of information, but need to be integrated in a larger framework. The Global Digital Format Registry[22] and its successor, the Unified Digital Format Registry (UDFR),[23] promise to close this gap.

Not only analysis of file formats, but also *discovery of potential preservation actions* is a tedious process that is prone to information gaps. Registries holding information about available tools for preserving digital content are being built, but need to be populated and publicly available. Furthermore, significant experience needs to be accumulated and analysed to provide a basis for shortlisting potential alternatives. While there is considerable progress in this area [6,9], the amount of information contained in public registries is currently insufficient and still needs to be complemented by manual investigation.

7.2.2 Requirements specification

Participants in case studies were all very positive that the requirements specification in the end captured their real requirements and were very satisfied with the evaluation results and the transparent documentation that results from the planning procedure. However, requirement specification continues to remain the most challenging part of the planning workflow. This is in part due to the fact that for many institutions, this is still a new area, and thus the high-level constraints and influence factors are either not yet settled or weakly defined. For example, it is sometimes not entirely

[20] www.nationalarchives.gov.uk/pronom/.

[21] http://www.digitalpreservation.gov/formats/index.shtml.

[22] http://www.gdfr.info/.

[23] http://www.udfr.org/.

clear which standards must be followed and which are just desirable, or how to calculate costs and assess risks.

The Planets project has developed a high-level model of institutional policies, which we have started using in the abovementioned case studies. This structured documentation proved to be very valuable in the decision process as it forces stakeholders to explicitly state their preferences and constraints. A related issue is the tendency of many stakeholders to think in terms of solutions rather than problems, thus preempting decisions to be made at a later stage. Examples are requirements detailing desired file formats rather than format characteristics when no formal decision has been taken yet, or defining migration requirements when emulation should be considered as well. The structured workflow greatly helps to overcome this problem.

Furthermore, the definition of significant properties is a technically challenging and complex issue. Considerable progress has been made through early applications of the described approach and in the INSPECT project. A recent discussion summarises the state of research [30]. In the planning tool Plato, a growing knowledge base of significant property trees is being made available, both community driven and as part of a moderated procedure within the Planets project. Feedback clearly indicates that this greatly supports and eases the planning procedure; however, it incurs the risk that decision makers do not thoroughly analyse their own needs, but instead simply reuse the needs of others.

Another difficulty arises when estimating projected costs for preserving digital objects. Our approach does not preclude the usage of any specific model, but also does not provide specific support for this task; costs have to be estimated manually and entered into the system, which is seen as nontransparent and tedious. In order to improve this, we are currently developing tool support for integrating models such as LIFE with the planning tool, so that these estimates can be calculated in the planning software and properly documented as part of the planning procedure.

In general, it should be noted that there are three steps where an institution influences the evaluation outcome:

1. Requirements definition,
2. Transformation settings, i.e. definition of the utility function, and
3. Importance weighting of requirements.

Requirements definition needs to be complete and along the correct lines of measurement; transformation has to define the acceptable parameter boundaries and establish utility values for each dimension; and the importance factors need to reflect the institutional priorities. At each of these steps, there is a risk of weakly defined and weakly documented assumptions and a corresponding need for thorough analysis, automated quality checks and tool support.

Summarising these issues, *requirements specification, evaluation and transformation* are complex procedures that, at first, may overwhelm decision makers. The software tool Plato provides considerable support and enables planners to reuse experience of others through a shared knowledge base. Still, the overall complexity of the problem implies that sophisticated tool support is needed to proactively guide decision makers and help them where possible in selecting information and taking the right decision. To this end, recommendation modules are currently under investigation that shall operate on case-based reasoning concepts.

Further, the *conceptual link* between influence factors and the impact changes in these have on decision preferences is a complex and critical problem. Creating and maintaining the conceptual connection between these influence factors and the outcomes of decisions via manual monitoring is a difficult task and a largely unsolved question, and manual evaluation of experiment results can be very time consuming. The effort needed to analyse objects, requirements and contextual influence factors is in many cases prohibitive. Characterisation tools support the automated comparison of objects; yet, there is a variety of requirements which cannot be measured automatically yet.

8 Discussion and outlook

Preservation planning is a complex issue. The decision on which actions to take to keep digital content alive over time taken alone is already challenging; thorough analysis of options and complete documentation and traceability is of essential importance to ensure authenticity and trustworthiness in digital repositories. Furthermore, preservation planning needs to become a systematic and continuous management activity as opposed to the prevailing ad-hoc decision making.

In this article, we discussed the main issues and challenges related to preservation planning and provided a definition of what a *preservation plan* should contain. We then presented a repeatable planning method for digital preservation. The process strives to provide high assurance of decisions by basing them on evidence generated in controlled experimentation. Actions are evaluated by applying them to sample content that is representative for the digital objects that shall be preserved. The method provides a structured workflow for planning activities and is anchored in well-known and widely accepted models.

The approach has been evaluated in a series of case studies; ongoing case studies are further exploring the challenges that are the focus of current and future work. As discussed in the previous section, we are specifically concentrating on addressing the following issues:

1. In order to improve the evaluation of requirements, we need a flexible integration infrastructure that supports automated measurements on the range of environmental aspects that need to be considered. We are thus developing a conceptual framework and tool support that helps preservation planners to link influence factors and their impact on the decisions and preferences. Such a measure will support them in the continuous monitoring that is needed to help moving preservation planning away from an ad-hoc procedure to a regular and continuous, largely automated management activity.
2. We are developing recommendation algorithms for several steps of the workflow. Discovering potential preservation actions to consider for evaluation and the selection of representative sample objects to apply them on are challenging tasks. Recommender's support in such complex steps shall reduce the effort on the one hand, but also lower the entrance barrier for users inexperienced with preservation planning on the other hand.

Acknowledgements Part of this work was supported by the European Union in the 6th Framework Program, IST, through the PLANETS project, contract 033789. The authors want to thank the partners in the Planets project, as well as other institutions applying the Planets PP workflow and using Plato, for their comments and the fruitful discussions on the preservation plan definition and workflow evaluation.

References

1. Alves, C., Castro, J.: CRE: A systematic method for COTS components selection. In: XV Brazilian symposium on software engineering (SBES), Rio de Janeiro, Brazil (2001)
2. Ayris, P., Davies, R., McLeod, R., Miao, R., Shenton, H., Wheatley, P.: The LIFE2 Final Project Report. London, UK (2008). Research report LIFE Project, London, UK. http://eprints.ucl.ac.uk/11758/
3. Beagrie, N., Semple, N., Williams, P., Wright, R.: Digital Preservation Policies Study. Technical report. Charles Beagrie Limited (2008)
4. Becker, C., Kolar, G., Kueng, J., Rauber, A.: Preserving interactive multimedia art: a case study in preservation planning. In: Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers. Proceedings of the Tenth Conference on Asian Digital Libraries (ICADL'07), Volume 4822/2007 of Lecture Notes in Computer Science, Hanoi, Vietnam, 10–13 Dec 2007, pp 257–266. Springer, Berlin (2007)
5. Becker, C., Strodl, S., Neumayer, R., Rauber, A., Bettelli, E.N., Kaiser, M.: Long-term preservation of electronic theses and dissertations: a case study in preservation planning. In: Proceedings of the 9th Russian Conference on Digital Libraries (RCDL'07), Pereslavl, Russia, October 2007. http://rcdl2007.pereslavl.ru/en/program.shtml
6. Becker, C., Ferreira, M., Kraxner, M., Rauber, A., Baptista, A.A., Ramalho, J.C.: Distributed preservation services: integrating planning and actions. In: Christensen-Dalsgaard, B., Castelli, D., Ammitzbll Jurik, B., Lippincott, J. (eds.) Research and Advanced Technology for Digital Libraries. Proceedings of the 12th European Conference on Digital Libraries (ECDL'08), Volume LNCS 5173 of Lecture Notes in Computer Science, Aarhus, Denmark, 14–19 Sept 2008, pp. 25–36. Springer, Berlin (2008)
7. Becker, C., Kulovits, H., Rauber, A., Hofman, H.: Plato: a service oriented decision support system for preservation planning. In: Proceedings of the 8th ACM IEEE Joint Conference on Digital Libraries (JCDL'08), pp. 367–370 (2008)
8. Becker, C., Rauber, A., Heydegger, V., Schnasse, J., Thaller, M.: A generic XML language for characterising objects to support digital preservation. In: Proceedings of the 23rd Annual ACM Symposium on Applied Computing (SAC'08), Fortaleza, Brazil, 16–20 March 2008, vol. 1, pp. 402–406. ACM (2008)
9. Becker, C., Kulovits, H., Kraxner, M., Gottardi, R., Rauber, A., Welte, R.: Adding quality-awareness to evaluate migration web-services and remote emulation for digital preservation. In: Proceedings of the 13th European Conference on Digital Libraries (ECDL'09), Lecture Notes in Computer Science, Corfu, Greece, September 2009. Springer Berlin (2009)
10. Brody, T., Carr, L., Hey, J.M.N., Brown, A., Hitchcock, S.: PRONOM-ROAR: adding format profiles to a repository registry to inform preservation services. Int. J. Digit. Curation **2**(2), 3–19 (2007)
11. Carvallo, J.P., Franch, X., Quer, C.: Determining criteria for selecting software components: lessons learned. IEEE Softw. **24**(3), 84–94 (2007)
12. Carvallo, J.P., Franch, X., Quer, C.: Requirements engineering for cots-based software systems. In: Proceedings of the 23rd Annual ACM Symposium on Applied Computing (SAC'08), Fortaleza, Brazil, 16–20 March 2008, pp. 638–644. ACM (2008)
13. Dobratz, S., Schoger, A., Strathmann, S.: The nestor catalogue of criteria for trusted digital repository evaluation and certification. J. Digit. Inform. **8**(2), (2007). http://journals.tdl.org/jodi/article/viewArticle/199/180
14. erpanet: Digital Preservation Policy Tool, September 2003. http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf
15. Farquhar, A., Hockx-Yu, H.: Planets: integrated services for digital preservation. Int. J. Digit. Curation **2**(2), 88–99 (2007)
16. Franch, X., Carvallo, J.P.: Using quality models in software package selection. IEEE Softw. **20**(1), 34–41 (2003)
17. Gilliland-Swetland, A.J., Eppard, P.B.: Preserving the authenticity of contingent digital objects: the InterPARES project. D-Lib Mag. **6**(7/8), (2000). http://www.dlib.org/dlib/july00/eppard/07eppard.html
18. Guttenbrunner, M., Becker, C., Rauber, A.: Evaluating strategies for the preservation of console video games. In: Proceedings of the Fifth international Conference on Preservation of Digital Objects (iPRES 2008), London, UK, September 2008, pp. 115–121
19. Hockx-Yu, H., Knight, G.: What to preserve?: significant properties of digital objects. Int. J. Digit. Curation **3**(1), (2008). http://www.ijdc.net/ijdc/article/view/70/70
20. Hoeven, J.R., Van Der Diessen, R.J., Van En Meer, K.: Development of a universal virtual computer (UVC) for long-term preservation of digital objects. J. Inf. Sci. **31**(3), 196–208 (2005)
21. Hofman, H., Planets-PP subproject, Becker, C., Strodl, S., Kulovits, H., Rauber, A.: Preservation plan template. Technical report, The Planets project (2008). http://www.ifs.tuwien.ac.at/dp/plato/docs/plan-template.pdf
22. ISO: Software Engineering—Product Quality—Part 1: Quality Model (ISO/IEC 9126-1). International Standards Organization (2001)
23. ISO: Information technology—Multimedia Content Description Interface—Part 1: Systems (ISO/IEC 15938-1:2002). International Standards Organization (2002)
24. ISO: Open Archival Information System—Reference Model (ISO 14721:2003). International Standards Organization (2003)
25. ISO: Document management—Electronic Document File Format for Long-term Preservation—Part 1: Use of PDF 1.4 (PDF/A) (ISO/CD 19005-1). International Standards Organization (2004)

26. ISO: Information Technology—JPEG 2000 Image Coding System—Part 12: ISO Base Media File Format (ISO/IEC 15444-12:2005). International Standards Organization (2005)
27. ISO: Information Technology—Open Document Format for Office Applications (ISO/IEC 26300:2006). International Standards Organization (2006)
28. Jadhav, A.S., Sonar, R.M.: Evaluating and selecting software packages: a review. Inf. Softw. Technol. **51**(3), 555–563 (2009)
29. King, R., Schmidt, R., Jackson, A.N., Wilson, C., Steeg, F.: The planets interoperability framework: an infrastructure for digital preservation actions. In: Proceedings of the 13th European Conference on Digital Libraries (ECDL'2009), 2009
30. Knight, G., Pennock, M.: Data without meaning: establishing the significant properties of digital research. Int. J. Digit. Curation **4**(1), 159–174 (2009)
31. Kontio, J.: OTSO: a systematic process for reusable software component selection. Technical report, College Park (1995)
32. Kontio, J.: A case study in applying a systematic method for COTS selection. In: Proceedings of the 18th International Conference on Software Engineering (ICSE-18), Berlin, pp. 201–209 (1996)
33. Kulovits, H., Rauber, A., Brantl, M., Schoger, A., Beinert, T., Kugler, A.: From TIFF to JPEG2000? Preservation planning at the Bavarian State Library using a collection of digitised 16th century printings. D-Lib Mag. **15**(11/12), (2009). http://dlib.org/dlib/november09/kulovits/11kulovits.html
34. Land, R., Blankers, L., Chaudron, M., Crnkovic, I.: High Confidence Software Reuse in Large Systems, Volume 5030 of Lecture Notes in Computer Science, Chapter COTS Selection Best Practices in Literature and in Industry, pp. 100–111. Springer, Berlin (2008)
35. Lawrence, G.W., Kehoe, W.R., Rieger, O.Y., Walters, W.H., Kenney, A.R.: Risk management of digital information: a file format investigation. CLIR report 93, Council on Library and Information Resources, June (2000)
36. Maiden, N.A., Ncube, C.: Acquiring COTS software selection requirements. IEEE Softw. **15**(2), 46–56 (1998)
37. Mellor, P.: Camileon: emulation and BBC domesday. RLG DigiNews **7**(2), April (2003)
38. Mohamed, A., Ruhe, G., Eberlein, A.: COTS selection: past, present, and future. In: Proceedings of the 14th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (ECBS'07), pp. 103–114 (2007)
39. National Library of Australia: Guidelines for the Preservation of Digital Heritage. Information Society Division United Nations Educational, Scientific and Cultural Organization (UNESCO) (2003). http://unesdoc.unesco.org/ulis/cgi-bin/ulis.pl?catno=130071
40. Ncube, C., Dean, J.C.: COTS-Based Software Systems, Volume 2255 of LNCS, Chapter The Limitations of Current Decision-Making Techniques in the Procurement of COTS Software Components, pp. 176–187. Springer, Berlin (2002)
41. Ncube, C., Maiden N.A.M.: PORE: Procurement-orienteds requirements engineering method for the component-based systems engineering development paradigm. In: Development Paradigm. International Workshop on Component-Based Software Engineering, pp. 1–12 (1999)
42. nestor Working Group-Trusted Repositories Certification: Catalogue of Criteria for Trusted Digital Repositories, Version 1. Technical report. nestor—Network of Expertise in long-term STORage, Frankfurt am Main (2006)
43. Neubauer, T., Stummer, C.: Interactive decision support for multiobjective cots selection. In: HICSS '07: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, Washington, DC, USA, p. 283b. IEEE Computer Society (2007)
44. RLG/OCLC Working Group on Digital Archive Attributes: Trusted Digital Repositories: Attributes and Responsibilities. Research Libraries Group (2002). www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf
45. Rolland, C.: Requirements engineering for COTS based systems. Inf. Softw. Technol. **41**, 985–990 (1999)
46. Ross, S., McHugh, A.: The role of evidence in establishing trust in repositories. D-Lib Mag. **12**(7/8), (2006)
47. Rothenberg, J.: Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Council on Library and Information Resources (1999). http://www.clir.org/pubs/reports/rothenberg/contents.html
48. Rothenberg, J., Bikson, T. Carrying authentic, understandable and usable digital records through time. Technical report. Report to the Dutch National Archives and Ministry of the Interior, The Hague (1999)
49. Saaty, T.L.: How to make a decision: the analytic hierarchy process. Eur. J. Oper. Res. **48**(1), 9–26 (1990)
50. Strodl, S., Becker, C., Neumayer, R., 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 2007, pp. 29–38
51. Strodl, S., Rauber, A.: Preservation planning in the OAIS model. New Technol. Libr. Inf. Serv. **1**, 61–68 (2008)
52. The 100 Year Archive Task Force: The 100 year archive requirements survey. http://www.snia.org/forums/dmf/programs/ltacsi/100_year/ (2007)
53. The Center for Research Libraries (CRL), Online Computer Library Center, Inc.(OCLC): Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC). Technical Report 1.0. CRL and OCLC (2007)
54. Thibodeau, K.: Overview of technological approaches to digital preservation and challenges in coming years. In: The State of Digital Preservation: An International Perspective, Washington, DC, July 2002. Council on Library and Information Resources (2002). http://www.clir.org/pubs/reports/pub107/thibodeau.html
55. van der Hoeven, J., van Wijngaarden, H.: Modular emulation as a long-term preservation strategy for digital objects. In: 5th International Web Archiving Workshop (IWAW05), Vienna, Austria (2005)
56. van Lamsweerde, A. Goal-oriented requirements engineering: a guided tour. In: Proceedings RE'01, 5th IEEE International Symposium on Requirements Engineering, Toronto, Canada, pp. 249–263 (2001)