SCAPE
Scalable Preservation Environments

Report on decision factors and their influence on planning

Authors
Markus Hamm, Christoph Becker (Vienna University of Technology)

November 2011

This work was partially supported by the SCAPE Project. The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

Table of Contents

1 Introduction

Over the past years, considerable effort has been invested in analysing the factors contributing to decision making in digital preservation and the constraints posed by different scenarios, and in building decision making frameworks and tools. With current state-of-the-art procedures in digital preservation, we can create plans that specify the preservation actions to be applied to a well-understood and homogeneous part of the content in a large repository. The planning tool Plato[i], created in the project PLANETS, has been applied to a number of real-world and pilot cases and is producing a growing body of knowledge (Becker & Rauber, 2011).

Consider an identified preservation problem consisting of a set of digital material that is at risk of becoming obsolete. The material is held by an organization. There are a number of possible alternatives to resolve the identified issues, and a number of objectives and constraints that have to be considered. The preservation planning approach implemented in Plato presents a systematic method and tool to create a plan for this scenario. Decision makers represent goals and constraints in a hierarchy of objectives resolving into decision criteria. They evaluate alternatives against these criteria by applying controlled experimentation and automated measurements, and take an informed decision based on the resulting objective evidence. The finalised plan is fully documented, and it is fully traceable to the reasons underlying each decision. The planning tool provides guidance and automation in the planning procedure.

Despite this progress, however, a number of significant challenges remain and pose a substantial barrier towards the successful transition of the control of preservation operations from ad-hoc decisions towards continuous management. On the one hand, preservation planning in reality still is a rather isolated affair, where knowledge is only exchanged informally. Plans created in the planning tool Plato can be shared with others by making them public, and a number of these plans are available for analysis by a growing user community. However, until now there has been no systematic assessment of the impact of decision criteria. This is partly due to the fact that the specification of decision criteria used to be entirely based on individual scenarios. This implied a substantial variation in criteria definition until recently, when a standardized method of identifying, documenting and reusing criteria with defined semantics was introduced (see Section 3). The automation in decision making processes is still limited by the fact that many information needs cannot be addressed automatically. Continuous management, however, requires systematic mechanisms and processes for information exchange and control.

The goal of the sub-project Scalable Planning and Watch is to move forward the control of digital preservation operations from ad-hoc decision making to proactive, continuous preservation management, through a context-aware planning and monitoring cycle integrated with operational systems. This systematic improvement of decision automation requires an assessment of the criticality and the exact impact of decision criteria. To provide this analysis, this deliverable presents a method and tool support for the quantitative assessment of decision criteria in preservation planning. We build upon a significant body of work collected in the last years, which includes preservation plans for different types of content, models for preservation goals and criteria, and a basic taxonomy of categories which we base our analysis upon. We conduct an analysis of key factors and decision criteria considered in preservation decisions and their quantitative influence on evaluation and decisions.

This analysis of decision factors entails not only conceptual work, but also a fully automated analysis tool that analyses preservation plans and summarises the quantitative impact of decision criteria and criteria sets upon decisions in certain scenarios. The development of this tool is the basis of the quantitative analysis activity; the module will become part of the planning component. In terms of the analysis of the planning data set, two obstacles presented themselves. On the one hand, the quality and originality of preservation plans created on the public Plato instance is not always clear, and preservation plans in general are intentionally private, i.e. protected from access since they are visible to their creators only. Analysis of plans has to focus on those plans that have been publicly released by their owners and are considered relevant and representative. Second, the majority of preservation plans have been created prior to the introduction of the taxonomy of decision criteria presented recently in (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011), which means that the criteria cannot automatically be linked across decisions. Formal classification of criteria in a manual way is thus necessary, and hence, the analysis focuses on a carefully selected set of plans that are of high quality and representative. Developing an automated analysis module presents a massive advantage: The analysis itself can be repeated at any time in the future and in fact could even be used as a trend analysis source. It will also be able to dynamically take into account new decision criteria that will inevitably be added to the planning knowledge base.

This report is largely based on the article "Decision criteria in digital preservation: What to measure and how" that appeared in JASIST 6/2011 and the conference papers "*Impact Assessment of Decision Criteria in Preservation Planning*" published at IPRES 2011 (Hamm & Becker, 2011) as well as "*Preservation Decisions: Terms and Conditions apply*" presented at JCDL 2011 (Becker & Rauber, Preservation Decisions: Terms and Conditions apply. Challenges, Misperceptions and Lessons Learned in Preservation Planning., 2011). The report is structured as follows. Section 2 provides the reader with relevant background material and references to related work. It describes key concepts of the areas of Multi-Criteria Decision Analysis, Preservation Planning and decision criteria for digital preservation decisions, and software quality models. Section 3 discusses the mapping of existing models for software quality and file format assessment to the decision criteria encountered in preservation planning. Section 4 discusses key issues in decision criteria analysis and impact assessment of criteria. Section 5 shortly presents a decision factor analysis tool that is being developed in SCAPE as part of the future planning component. Section 6 introduces the case studies we used for our analysis. Section 7 presents results of applying the presented analysis approach to a growing body of knowledge created in real-world case studies. Finally, Section 8 discusses implications and presents an outlook on future work. In particular, we will analyse how the results of this work package can be useful to guide work in other areas of the SCAPE project.

2 Background

Choosing the right treatment for a given set of objects and a specific purpose 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 intricate complexity of situations and requirements that need to be considered render this decision a delicate task. A variety of actions exist, but quality varies across tools; properties vary across content; usage and requirements vary across users and scenarios; risk tolerances, preferences, costs, and constraints vary across collections, organizations, and environments. Finally, all of these factors are subject to constant technological shifts that have to be detected and handled.

The decision maker has to achieve multiple competing objectives such as *minimize costs*, *ensure authenticity*, and *provide online access,* while considering the contextual constraints of legislation, technology and budgets. When making these objectives operational, one must not distort the balance of the whole. In complex environments with potentially changing requirements, subjective human judgment of software quality and the reliance on declared capabilities of components cannot be considered sufficient evidence for trustworthy decision making, and cannot replace objective evidence as the basis of decision making. 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. For all decisions taken, we need full evidence of reasons and documentation to ensure auditable procedures that support trustworthiness, as emphasized by the Trustworthy Repository Audit and Certification Criteria and Checklist (CRL and OCLC, 2007)(International Standards Organisation, 2011).

According to (Terzis, 2009), "the modern view of trust is that trustworthiness is a measureable property that different entities have in various degrees. Trust management is about managing the risks of interactions between entities. Trust is determined on the basis of evidence ... and is situational – that is, an entity's trustworthiness differs depending on the context of the interaction. " This applies in particular to operational preservation planning, where an entity's trustworthiness has to be validated in the context of an interaction: We need to do so in a controlled environment where the varying parameters are known and the outcomes repeatable, reproducible, and measurable.

The planning method and the supporting tool Plato have been used successfully with and without expert assistance. However, the lessons learned from the extensive real-world experience show the complexities involved in the planning activity and indicate that strong tool support and substantial knowledge is needed to successfully create a preservation plan. This section will discuss the specific issues that we deem essential for broadening the applicability of the method and point out potential for improvement of the method and tool.

There are three central, interlinked drivers that determine the decision outcomes:

1. Requirements definition,
2. Definition of the utility functions, and
3. Importance weighting of requirements.

While these aspects are closely connected, it is of central importance to have a clear understanding of the distinct nature of each of them.

Requirements definition needs to be complete; focused on the problem domain, not potential solutions; and along the correct lines of measurements that are applicable. Utility functions reflect the organization's assessment of value for each criterion. They have to define acceptable parameter boundaries and establish utility values for each dimension. Finally, the importance factors need to reflect the actual institutional priorities. These are the cornerstones of decision making and need to be explicitly separated and clearly defined. 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. Most importantly, efficient and effective evaluation and decision making depends on a number of measures to be taken on a range of sources. Manually obtaining these is tedious and error-prone; however, the coverage of automated measurements is often

unknown or insufficient (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011).
As Dappert recently discussed, there is a substantial variation in the definition of significant properties of digital objects (Dappert, Deal with conflict, capture the relationship: The case of digital object properties, 2010; Dappert & Farquhar, Significance is in the eye of the stakeholder, 2009) . The same applies to performance characteristics and measurable properties in general. This lack of standardization of property definition and measurements implies that there is no clear way of identifying measurements and requirements and providing ongoing monitoring and re-assessment of quality of service. It also leads to a lack of comparability of results across case studies. The flexibility to express and model specific aspects of the scenario, which addresses the fundamental need to take these peculiarities into account, carries considerable difficulties. The possibility to model organizational preferences and utilities is essential, but the objective *criteria* should be standardized, reusable, uniquely identified, and selected from catalogues; and correspondingly, the measurements need to be clearly defined, repeatable, and reproducible.

Case studies have shown that the manual effort needed to specify requirements, evaluate alternatives and create a preservation plan is often prohibitive. A typical case study involved several people for about a week, including a planning expert to coach the decision makers (Kulovits, Rauber, Kugler, Brantl, Beinert, & Schoger, 2009). The addressed holdings, however, constitute only a fraction of the institutions' overall content. This has the effect that for many organizations, applying the planning approach to all or even just the most valuable collections is not feasible. It is evident that substantial tool support and automation is needed to decrease the amount of manual involvement and thus make it feasible to create and monitor preservation plans and run repository operations in the large.

The upcoming ISO standard describing metrics for Repository Audit and Certification includes detailed requirements on planning procedures that have to be considered to achieve trustworthy decision making. These include, for example, the requirements to explicitly specify the '...Content Information and the Information Properties that the repository will preserve'. Clearly, such a specification needs to build upon

1. a model for specifying such properties,
2. an assessment of the possible actions that the repository can employ to achieve its goals within the constraints posed by these properties, and
3. a method to evaluate whether the repository will be able to preserve these properties, in which form, and at which costs and risks.

Models for specifying transformation information properties (as the OAIS calls them) or significant properties (as they are often referred to) have been discussed intensely over the last years. The realistic evaluation of such properties requires objective evidence, repeatable measures, and thorough documentation. The Plato approach combines such an evaluation method and supports the automated and repeatable documentation of objective evidence through controlled experimentation and automated measurements. At its heart, the so-called objective tree specifies goals and objectives of a preservation scenario and breaks these aspects down into decision criteria that can be quantitatively determined. Figure 1 presents a very simple illustrative example containing three decision criteria and one requirement node ('Correctness') that comprises two of the criteria.

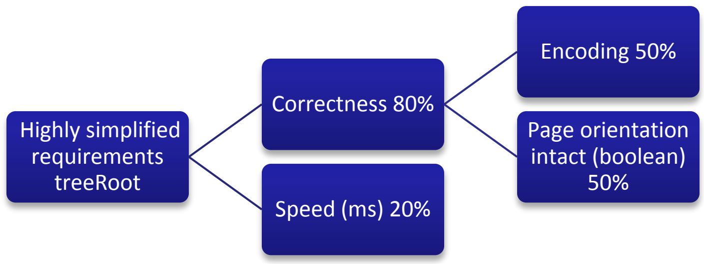


Figure 1 Highly simplified requirements tree

In taking preservation decisions, preservation planners have to reconcile potentially conflicting and initially ill-defined goals and find the optimal solution within weakly defined organizational constraints. The approach followed in Plato builds upon a widely used approach that resolves the incommensurability of multiple decision criteria by applying utility analysis. To allow a comparison across the criteria, a utility function is specified for each criterion that contains an explicit mapping to a uniform utility score ranging from 0 (unacceptable) to 5 (best). This score can then be weighted and aggregated across the hierarchy.

The combination of objective evidence measured in specific scales, subjective assessment represented in case-specific utility functions, and relative weights across the goal hierarchy, is a powerful, yet flexible model. However, it requires a profound understanding of the intricacies of decision making scenarios, and a careful distinction between the key concepts of evidence, utility, and weighting (Becker & Rauber, Preservation Decisions: Terms and Conditions apply. Challenges, Misperceptions and Lessons Learned in Preservation Planning., 2011). Common approaches to sensitivity analysis vary the weightings of attributes to determine the robustness of assigned weights similar to the approach presented in (Butler, Jia, & Dyer, 1997).

Recent discussions about preservation planning presented a categorization of decision criteria according to their measurement needs (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011) and analysed a series of case studies, focusing on lessons learned and open challenges (Becker & Rauber, Preservation Decisions: Terms and Conditions apply. Challenges, Misperceptions and Lessons Learned in Preservation Planning., 2011). Kilbride discussed the fact that decision making can be very complex, and emphasized the benefits that experience sharing would provide for organizations facing the preservation planning problem (Kilbride, 2010).

One of the key aspects in planning is the question of software quality. The ISO standard 25010 - 'Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models' (International Standards Organisation, 2011) is based on the earlier ISO 9126 family. The ISO/IEC 9126 standards (International Standards Organisation, 2001) define a hierarchy of high-level quality attributes, where quality measures are based on procedures recommended in ISO 15939 (International Standards Organisation, 2007). SQUARE combines a revised quality model with evaluation procedures based on ISO 14598 (International Standards Organisation, 1999). It defines requirements on the specification of software product quality criteria (International Standards Organisation, 2007). ISO 25010 states that it defines

- a quality-in-use model composed of five characteristics (some of which are further subdivided into subcharacteristics) that relate to the outcome of interaction when a product is used in a particular context. This system model is applicable to the complete human computer system, including both computer systems in use and software products in use.
- a product quality model composed of eight characteristics (which are further subdivided into subcharacteristics) that relate to static properties of software and dynamic properties of the computer system. The model is applicable to both computer systems and software products.

Quality attributes are defined in a hierarchic manner. The quality model divides product quality into characteristics, each of which is composed of several sub-characteristics. Section 3 will discuss their relation to decision criteria in preservation planning.

These hierarchical structuring procedures have already been used to inform the hierarchical definition of objective trees in the planning approach in Planets. But since preservation planning has a specific focus, different compared to generic cases of software product selection (Becker & Rauber, 2011), it is necessary to customize the quantitative part of evaluation, as recommended by ISO SQUARE. Hence, the next section presents a quality model that is based on ISO 25010 for the high-level generic quality model and associates it with exemplary measurable criteria that have been of concern in productive decisions in preservation planning. This reconciled quality model then enables the analysis of accumulative decision factors such as the resource utilization of preservation action components in a systematic and standardized way, while retaining the full expressiveness and flexibility of the decision making framework.

3 Mapping decision factors, quality models and metrics

3.1 A generic taxonomy for decision criteria measurement

Based on the empirical evidence gathered in Plato, a first in-depth analysis of about 500 decision criteria of planning studies led to a bottom-up classification of criteria according to their sources of measurement. This is discussed in detail in (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011) and forms the basis for the design and evaluation of a full-coverage measurement framework for digital preservation decisions. The primary distinction hereby is between criteria relating to a *preservation action* and criteria relating to its *outcome*. The latter is divided into *format properties*, *object properties* and *outcome effects* such as costs, as shown in Figure 2.

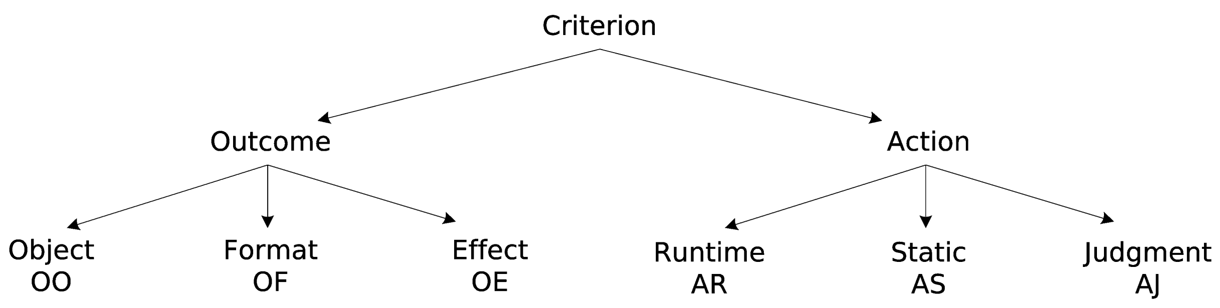


Figure 2 Generic criteria taxonomy

Fundamentally, all criteria requiring measurement refer either to the action, i.e. the component, or the outcome of an action, i.e. a rendering or transformation of a digital object. The corresponding top level categories Outcome (O) and Action (A) focus on the outcome of applying an action, and the properties of the action, respectively. Outcome criteria can be further distinguished to describe general effects of the outcome (OE), such as the expected annual storage costs that result from applying a certain action; criteria describing the format of the objects (OF); and criteria describing the abovementioned significant properties of objects (OO). Action components exhibit properties that are static and descriptive in nature (AS), properties that can be measured at runtime (AR), and some properties that depend on judgment (AJ).

The taxonomy is in principle orthogonal to the goal hierarchy and its specific structure. An evaluation objective can thus be composed of measurable criteria belonging to different categories. For instance, the general objective of minimizing costs may include both a criterion evaluating the price per object, i.e. per execution, of a component (AS), and one or more criteria specifying runtime (i.e. execution) characteristics such as memory usage or processor time used (AR), which imply a certain level of hardware expenditures.

In more detail we thus identify the following categories:

1. Properties of the **outcome** of applying a component.
    - **Object**. This category entails all desired properties of digital objects. This includes desired properties of the objects and properties that have to be kept unchanged compared to the original object. Properties of the resulting objects, such as the ability to search or edit text documents, need to be measured on the outcome of applying a preservation action. For significant properties that have to be kept intact, the base measures taken on the outcome of the preservation action have to be compared to the base measures obtained from the original object. For example, the criterion *Textual content unchanged* is measured by analyzing the original object and the outcome of the preservation action and comparing these for textual equality to get a derived measure on a Boolean scale. We thus obtain this measure by comparing the text content of the original object to the text content of the action result. Further examples of criteria in this category include *Image width is unchanged*, *Object is editable*, and *Embedded EXIF metadata are preserved*.
    - **Format.** This category comprises criteria that specify desirable characteristics of the formats that are used for representing digital content. As a significant portion of the risks to digital content lies in the form of representation and its understandability, this is often a central decision criterion. Typical criteria include standardization (e.g. *Format is standardized by ISO*), format complexity, or openness of formats. These criteria comprise compliance to institutional policies as well as preferences for low-risk formats; what an institution considers a low risk depends on its risk profile which is modelled in the utility functions. Measurements of these criteria are applied by analyzing the format of the outcome and getting additional information on known properties of certain formats from trusted external data sources such as the Pronom Technical Registry[ii] and the P2 Semantic Registry (Tarrant, Hitchcock, & Carr, 2009). Further examples of criteria in this category include *Number of viewers currently supporting this format*, *No IPR issues concerning the format are known*, and *Format is natively supported by standard browsers*.

- **Effect of outcome**. This refers to any other effects caused by the application of a certain component. Typically, these effects are calculated by organization-specific models or recognized cost models such as LIFE based on measures as model inputs. For example, storage costs will depend on organizational cost structures, but strongly correlate with the file size of objects. The file size of the output objects measured in relation to the originals can thus be used as input for a cost model computing the total annual storage costs of a collection. Further examples of criteria in this category include *Resulting archival storage costs* and *Effort for preservation watch reduced*.

2. Properties of the components, i.e. the **action** taken.
   - **Runtime.** This category entails runtime properties of action components such as performance, throughput, and memory utilization. Since these properties are highly dynamic and depend on a number of factors, measurements need to be taken in a controlled environment. Examples of this category include *Peak memory usage*, *Average processing cycles consumed per MB* and *Average memory consumed per MB*.
   - **Static.** Criteria of this category refer to properties of the action components that do not vary per execution run nor show differences when evaluated by different users; i.e., they are not subject to the evaluator's perception and can be determined objectively. These criteria can thus often be obtained from trusted sources. For example, the question whether a component is open source or not should be documented in component registries. Where not found, these criteria need to be evaluated manually with appropriate documentation. Examples of criteria in this category include *Syntactic validation is performed* and *Licensing costs of component*.
   - **Judgment.** This category is sometimes relevant, but decision criteria in this category should be kept to a minimum. It comprises criteria that cannot be objectively determined with reasonable effort. Usability is a prime example where judgment may be necessary. In digital preservation this does not have high influence on the decision, since the components to be evaluated are not to be applied by an end user. In other cases, this has more importance; but in any case, proper documentation of evaluation values is essential. Examples of criteria in this category include *Ease of component integration into existing workflow environment* and *Process log output is human readable.*

The main difference between the three categories of action criteria can be seen when considering the approaches generally assumed for measurements. Runtime criteria reflect execution properties of candidate actions and need to be empirically measured, preferably in an automated and scalable manner. Static criteria can be documented in knowledge bases, even though they will be eventually subject to changes. Criteria that require human judgment, on the other hand, have to be evaluated by an expert as part of the evaluation procedure. This judgment will inevitably be subjective; the corresponding reasoning should thus be documented to support transparency.

When a sufficient number of expert judgments have been accumulated for a certain action and a criterion, the converging average judgment may become a *static* criterion deposited in a knowledge base. Note that this would be a new, separately obtainable property distinct from the first.

The taxonomy proved complete in its expressiveness to cover all the criteria encountered in the case studies evaluated so far, since it models all relevant entities encountered in the decision process. On an analytical level, it appears that there can be only two aspects to consider: The action to take and the outcome of it. Specifying the action and the outcome in more detail resulted in the taxonomy. The presented taxonomy itself was refined from a more extensive preliminary taxonomy which

contained categories named “other” to capture any miscellaneous entries that did not fit elsewhere. These were removed because an evaluation of over a dozen case studies did not encounter any examples of such criteria.
On an empirical level, we have not yet encountered a valid decision criterion that would defy classification in one of the categories. In fact, three decision criteria encountered in one case did appear to do so, but upon closer scrutiny turned out to be ill-specified. Close inspection revealed that these criteria were in fact irrelevant to the decision process: They described the legal IPR status of the original digital objects in such a way that it was invariant of the decision process and the actions involved; no potential preservation action could have possibly changed the IPR status of an existing object. (The only way to influence that status would have been to include into the decision process the action of pursuing a legislative act; and in that case, the criteria would have been classified as output effect (OE).)
It is important to note here that although no Preservation Action can change the IPR status, it potentially has the capability to affect which actions can be chosen. Constraints on the possible way of actions to chose are reflected in the policy statements of an organisation, which are documented in the policy model associated with a plan. These policy statements reflect constraints on the preservation actions. These would, for instance, prevent the creation of derivative copies and as such exclude migration actions. As criteria, these could be categorised under “Action Static”. However, this has usually not been part of the objective tree, but considered by the planners in the selection of potential alternatives. Based on the more expressive SCAPE policy model, this previously implicit connection will be explicitly linked and directly used in the planning component to filter out candidate actions that violate IPR constraints of the organisation.
To validate the expressiveness, the construction of the preliminary taxonomy was followed by a classification of all criteria encountered in all case studies conducted so far.

The classification is based on an extensive evaluation of a number of case studies on preservation planning for a range of different types of objects which were conducted over the past with partners from different institutional backgrounds. Based on a classification of criteria and a categorization of all decision criteria encountered in case studies so far, it was demonstrated that a majority of the criteria can be evaluated by applying automated measurements in a controlled environment, and automated measurements can be used to substantially improve repeatability of decisions. This reduces the effort needed to evaluate components and thus enables scalability of planning and repository operations. It also provides substantial support for trust in the decisions since extensive evidence is produced in a repeatable and reproducible way and documented alongside the decision in a standardized and comparable form. (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011)

This classification serves as a key tool to increase automated measurements in a measurement framework. However, it does not relate clearly to the impact that decision factors and criteria sets have on the final decisions for two reasons:

1. No impact analysis is performed,
2. Decision factors are related to concerns such as risks, which may be expressed by multiple criteria measured through diverse sources (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011).

Thus, we conducted a mapping of top-down models to the overall classification into measurable *action* and *outcome* criteria. In particular, we will discuss format properties, software quality, and information properties.

3.2 Format properties

The format website run by the Library of Congress (LoC) suggests evaluating formats according to the two aspects *sustainability* and *quality and functionality*. Sustainability factors recommended are disclosure, adoption, transparency, self-documentation, external dependencies, impact of patents, and technical protection mechanisms (Arms & Fleischhauer, 2011)

PRONOM suggests assessing a given file format against each of the following characteristics and sub-characteristics:

- Capability: The support for features required or desirable to meet business requirements, such as support for specific types of content (e.g. chart support in spreadsheet),
- Quality: The accuracy of information storage, represented by Precision and Lossiness.
- Resilience: Safety over time, represented by Ubiquity (resilience against obsolescence), Stability (resilience against software updates), and Recoverability (resilience against accidental corruption).
- Flexibility: Ability to adapt to changing requirements, represented by Interoperability (with existing tools) and Implementability (the degree of difficulty to implement software for this format) (The National Archives of the UK, 2011)

The given list is not intended to be fully complete and needs customization and extension dependent on the given context. Furthermore, it is clear that most of these high-level factors are not directly measurable. While knowledge sources such as PRONOM document experts' assessments of some of these attributes, many characteristics are high-level characteristics and require assignment of more specific quantified properties to be reliably assessed. We use these factors for the high-level generic quality model and associate them with exemplary measurable criteria that have been of concern in productive decisions in preservation planning.

Figure 3 shows characteristics assembled from LoC and PRONOM (in bold letters) and links them to planning criteria extracted from several case studies. The criterion 'rights' describes the degree of openness of IP rights (including patents) that threaten the sustainability of a format. While in the case study set, only this one explicit criterion was used as a simplified risk indicator, it is a critical risk factor that should be of general concerns and only one of several possible criteria that represent the complex subject of IPR issues. These are combined in a set of criteria called "rights" which groups issues around the impact of patents (cf. (Arms & Fleischhauer, 2011)) and other IP rights such as reasonable and non-discriminatory licensing. The exact criteria set of these issues is expected to evolve as the understanding of the topic progresses in the field.
It can be seen that a combination of both models is required to cover all factors that have been used for evaluation in real-world decisions. Merging these references to a unified model as in the suggested model above leads to a more suitable model for the preservation context. Section 7 will shed some light on the actual impact that these format criteria have on real-world decisions in comparison to other decision factors such as preservation process requirements.

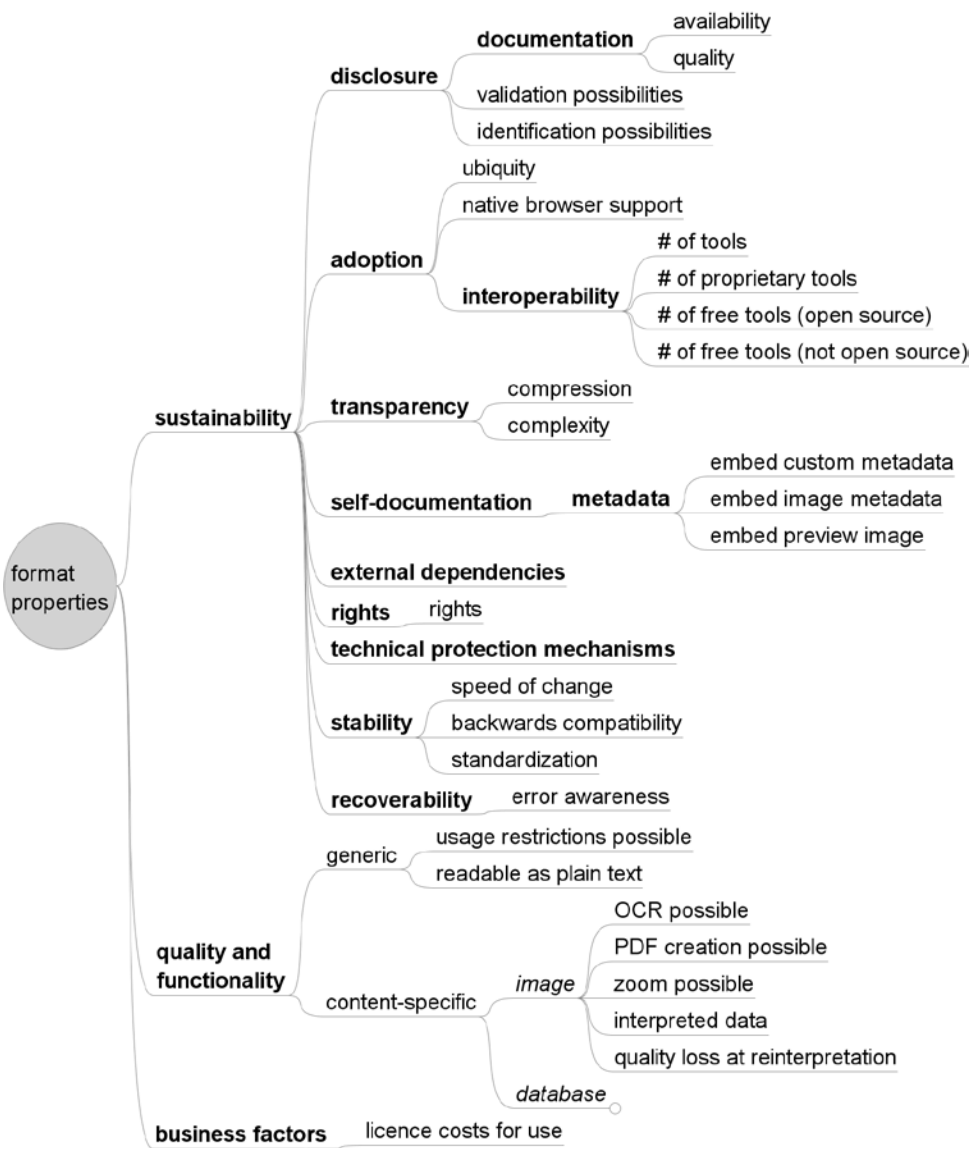


Figure 3 Format factors and associated criteria

3.3 Software Quality

The quality of software components has been analyzed extensively over the past decades, and a number of formal models have emerged. We analyzed decision criteria from planning case studies, based on our previous analysis, and assigned them to the SQUARE quality model.

To get a better understanding of the presented model, Table 1 provides the short definition of ISO 25010 quality characteristics and sub-characteristics.

Characteristic	Sub-Characteristic	Description
functional suitability		degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions
functional suitability	functional completeness	degree to which the set of functions covers all the specified tasks and user objectives
functional	functional	degree to which a product or system provides the correct results

suitability	correctness	with the needed degree of precision
functional suitability	functional appropriateness	degree to which the functions facilitate the accomplishment of specified tasks and objectives
performance efficiency		performance relative to the amount of resources used under stated conditions
performance efficiency	time behaviour	degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements
performance efficiency	resource utilization	degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements
performance efficiency	capacity	degree to which the maximum limits of a product or system parameter meet requirements
compatibility		degree to which a product, system or component can exchange information with other products, systems or components, and/or perform its required functions, while sharing the same hardware or software environment
compatibility	co-existence	degree to which a product can perform its required functions efficiently while sharing a common environment and resources with other products, without detrimental impact on any other product
compatibility	interoperability	degree to which two or more systems, products or components can exchange information and use the information that has been exchanged
usability		degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use
usability	appropriateness recognisability	degree to which users can recognize whether a product or system is appropriate for their needs
usability	learnability	degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use
usability	operability	degree to which a product or system has attributes that make it easy to operate and control
usability	user error protection	degree to which a system protects users against making errors
usability	user interface aesthetics	degree to which a user interface enables pleasing and satisfying interaction for the user
usability	accessibility	degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use
reliability		degree to which a system, product or component performs specified functions under specified conditions for a specified period of time
reliability	maturity	degree to which a system, product or component meets needs for reliability under normal operation
reliability	availability	degree to which a system, product or component is operational and accessible when required for use
reliability	fault tolerance	degree to which a system, product or component operates as

		intended despite the presence of hardware or software faults
reliability	recoverability	degree to which, in the event of an interruption or a failure, a product or system can recover the data directly affected and re-establish the desired state of the system
security		degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization
security	confidentiality	degree to which a product or system ensures that data are accessible only to those authorized to have access
security	integrity	degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data
security	non-repudiation	degree to which actions or events can be proven to have taken place, so that the events or actions cannot be repudiated later
security	accountability	degree to which the actions of an entity can be traced uniquely to the entity
security	authenticity	degree to which the identity of a subject or resource can be proved to be the one claimed
maintainability		degree of effectiveness and efficiency with which a product or system can be modified by the intended maintainers
maintainability	modularity	degree to which a system or computer program is composed of discrete components such that a change to one component has minimal impact on other components
maintainability	reusability	degree to which an asset can be used in more than one system, or in building other assets
maintainability	analysability	degree of effectiveness and efficiency with which it is possible to assess the impact on a product or system of an intended change to one or more of its parts, or to diagnose a product for deficiencies or causes of failures, or to identify parts to be modified
maintainability	modifiability	degree to which a product or system can be effectively and efficiently modified without introducing defects or degrading existing product quality
maintainability	testability	degree of effectiveness and efficiency with which test criteria can be established for a system, product or component and tests can be performed to determine whether those criteria have been met
portability		degree of effectiveness and efficiency with which a system, product or component can be transferred from one hardware, software or other operational or usage environment to another
portability	adaptability	degree to which a product or system can effectively and efficiently be adapted for different or evolving hardware, software or other operational or usage environments
portability	installability	degree of effectiveness and efficiency with which a product or system can be successfully installed and/or uninstalled in a specified environment
portability	replaceability	degree to which a product can replace another specified software product for the same purpose in the same environment

Table 1: Quality model attributes specified by ISO 25010 (SQUARE)

Figure 4 Action quality attributes and associated metrics

Figure 4 illustrates a subset of action related criteria and their structuring and classification in alignment with SQUARE. The ISO quality factors are given in bold. The ISO quality characteristic *functionality suitability* merits special attention. Functional *completeness* includes process-related features of software components such as the traceability of performed actions or the presence of mechanisms to support validation of input objects. However, content-specific features describing support of preservation action components for specific features of content also belong to this category. Functional *appropriateness* generally refers to the question whether certain preservation action components are applicable to an organization's holdings. This is generally not an evaluation criterion in planning, but rather a pre-selection criterion for creating the list of candidate actions that are evaluated. Finally and most crucially, functional *correctness* is at the heart of the quest for authenticity and represented is a specific category in the planning framework, as discussed below.

Note the node "business", which contains factors that are relevant in an organisational decision making context, but do not refer to the intrinsic quality of the software used, but to the contextual judgement about certain attributes of it that have an influence on the perceived value in an organisational context. These factors are outside the scope of the SQUARE quality model, but within the scope of decision making and preservation planning, and have to be considered alongside the intrinsic quality of the preservation action.

3.4 Information Properties and Functionality

The ISO characteristic functional correctness has an especially high relevance in the digital preservation context. Assuring that preservation action results are correct is a fundamental goal of digital preservation. This is covered by the category *Outcome Object* in the decision criteria taxonomy of Plato. Essentially, this can be further divided into

1. *Transformation Information Properties* refer to the significant properties to be preserved throughout changes of either environments or object representations.
2. *Representation Instance Properties* describe aspects of the representation, i.e. of the encoding, of information objects. This includes the file size required to represent a certain information object or the question if a representation is well-formed, valid and conforming to a certain expected format profile.
3. *Information Properties* are desired properties or features of the objects themselves.

In SCAPE terminology, the first class is the goal of the QA work package, the second category clearly is of concern in PC, whereas properties of the third category may appear in both QA (as intermediary results) and PC (as properties of interest).

3.5 Observations

This section analysed decision criteria from two viewpoints and created a mapping between these:

a. Measurable decision criteria were classified according to their sources of measures.
b. Decision factors are organised in top-down hierarchies.

The exact way of taking measures on criteria – measures which describe in a quantitative way the fulfilment of quality attributes – is a complex issue and highly domain dependent. The decision criteria taxonomy presented in (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011) and discussed above provides important information about this and enables an additional classification that can be used to guide evaluation. More specifically, this means that some attributes can be researched, documented and fed into a catalogue; some are highly or entirely context-dependent, yet, they are relevant for selection and decision making; and some require empirical measures in controlled experimentation.

However, the *measurement taxonomy* does not fully explain criteria semantics. Hence, this section aimed at mapping standard top-down quality models to decision criteria. In particular, the ISO 25010 quality model presents an international standard for modelling software quality attributes in a high-level top-down fashion. This stable standard provides a solid reference to resolve ambiguities about the meaning of certain quality attributes such as reliability, stability, etc.

Clearly, the models discussed here are all hierarchical. ISO has a hierarchical structure; the objective trees are hierarchical; the taxonomy of Plato is hierarchical. However, this does not mean that the quality model is an objective tree, or that the objective tree needs to conform to such a structure. There are many ways to structure hierarchical trees of criteria; the objective tree should contain all objectives and requirements that pertain to a certain scenario.
The differentiation of the taxonomy described in (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011) is essentially orthogonal to the ISO quality model.

The taxonomy describes measurable criteria, not the concerns they relate to – it is a bottom-up classification, whereas the ISO model is a top-down quality model. For example, the ISO quality attribute 'performance efficiency' includes dynamic runtime criteria such as time used per sample object, but also static action criteria such as the capacity of a tool, e.g. the maximum number of files in a batch process.

While every objective tree will be different, it is important to ensure consistency in the definition of measurable criteria in order to enable knowledge sharing, automation, and verification. Similarly, it is possible to improve support for the decision makers by introducing guidance about the high-level quality models that are considered of relevant in the field. The hierarchical quality models drive the top-down definition of an objective tree, while the metrics classified according to measurement sources provide the necessary quantification means at the bottom of the tree. Any decision criterion in a properly specified tree combines a standardised part – the mapping to an unambiguously specified and clearly referenced criterion – with a subjective part – the assessment of the possible values of this criterion in a utility function and the relative weight of the criterion.

The requirements definition support in the SCAPE planning component will correspondingly be supporting the combination of these two key elements in the decision making process.

4 Analysis of decision factors

As decision makers, we want to improve the efficiency of a specific decision making scenario while keeping full trustworthiness. For improving preservation planning processes in general, we want to improve efficiency over many scenarios. To advance the understanding of the field, finally, we want to gain insight into decision making processes and their key factors.

Being interested in assessing the impact of decision factors, we need to consider both single decision criteria as well as certain logical groupings of criteria. For a given set of decision criteria and plans, we want to answer key questions such as the following.

1. What is the impact of a certain criterion on the decision? Would a change in its evaluation, i.e. in the objective evidence, change preference rankings on alternative solutions?
2. Considering a specific case: How critical was this criterion in other cases? Has it led to a rejection of potential alternatives in similar cases?
3. What is the accumulated impact of a set of criteria on decisions in certain scenarios? (For example, what is the accumulated impact of criteria relating to format risks in the preservation of scanned images in large libraries? What is the accumulated impact of the resource utilization of action components in large migration decisions in archives?)
4. Are there any (sets of) decision factors which by themselves cannot change decisions, no matter which evaluation values we insert?
5. What is the minimum set of criteria that have to be considered in a given scenario?

While we are not able to fully address all questions yet, this report will advance the state of analysis by specifying concrete metrics to use and relate them to these key questions.

The questions relating to impact of a single criterion correspond to a robustness or sensitivity assessment. The previous approach to assessing sensitivity of decision makers' preferences computed variations of relative weightings to produce a robustness assessment judging the influence of tree branches on the root score. This does not address the specific scales, in particular the differences between numerical and ordinal measures. It also does not assess the sensitivity of the utility functions, which may include non-linear effects produced by the mappings. Furthermore, it does not consider reliability of measures. The combination of these aspects, however, can lead to substantial variations in the scores, as we will see below. On the other hand, the questions regarding decision cases require an accumulated assessment of the impact of multiple criteria over sets of plans, where each criterion may appear in a number of plans. To achieve this, we will define impact factors for sets of criteria. To answer the questions posed above, we need quantitative measures that consider

- the usage frequency and weight of a criterion in comparable scenarios (where a scenario is defined at least by the type of content and the type of organization), and
- the impact caused by a change in objective facts, i.e. the extent to which the utility scores of decisions including the criterion change when the evaluation facts change.

This requires us to integrate a number of properties in our assessment:

1. the number of times and frequency a criterion is used in planning cases,
2. the set of total weights of a criterion in each case,
3. the set of values collected for a criterion, and
4. the set of utility functions for the criterion.

In search for realistic, relevant and representative quantitative measures, we will define a number of impact factors for single criteria and groups of criteria. Section 6 will discuss the results obtained by their application to a set of real-world results. To consider the impact of criteria contained in a hierarchical structure, we have to consider their aggregation throughout the hierarchy. Criteria are weighted on all levels of the hierarchy in a relative fashion. To aggregate utility scores in the objective tree, the two standard weighted aggregation functions weighted sum and weighted multiplication are included in Plato. For weighted multiplication, utility values are taken to the power of the weight of the node to ensure that nodes with a weight of 0 result in a neutral element.
The *total weight* of a criterion can be easily determined by multiplying its weight with all parent weights up to the root node of the tree.

ID	Factor	Definition
IF1	Count	Number of plans using this criterion
IF2	Spread	Percentage of plans using this criterion
IF3	Weight	Average total weight of this criterion
IF4	Discounted Weight	Sum of total weights of this criterion, divided by number of all plans
IF5	Potential	Average potential output range of this criterion
IF6	Range	Average actual output range of this criterion
IF7	Discounted Potential	Sum of all criterion potential output ranges, divided by number of all plans
IF8	Discounted Range	Sum of all criterion actual output ranges, divided by number of all plans
IF9	Maximum Potential	Maximum potential output range
IF10	Maximum Range	Maximum actual output range
IF11	Variation	Average relative output range
IF12	Maximum Variation	Maximum relative output range
IF13	Rejection Potential Count	Number of utility functions with an output range including 0.
IF14	Rejection Potential Rate	Percentage of utility functions with an output range including 0.
IF15	Rejection Count	Number of utility functions actually rejecting alternatives.

IF16	Rejection Rate	Percentage of utility functions actually rejecting alternatives.
IF17	Reject Count	Number of rejected alternatives
IF18	Reject Rate	Percentage of rejected alternatives

Table 2 Impact factors for single criteria.

Table 2 summarizes and names all impact factors, designated *IF*, for single criteria. Formal definitions are provided in (Hamm & Becker, 2011). The basic impact factors of a criterion are the number of plans referring to it, the average total weight of the criterion across these plans, and the relation between these. These simple factors do not represent the actual impact that a change in evaluation has, since they do not account for the utility function. Arguably, this utility has more impact on the final result than the weighting itself (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011). More meaningful impact factors of a decision criterion can thus be quantified by considering the possible effect that a change in the objective facts that the criterion refers to has on the assessment of the criterion with respect to the decisions taken. This can be obtained by calculating the change in the final score of the objective tree root caused by a change in the criterion evaluation.

Consider a Boolean criterion c with values {Yes, No}. Let the utility function u defined in a certain plan p map *Yes* to a target utility of *5* and *No* to the target utility *1*. If c is assigned a total weight of *0.25* in the given plan, the potential output range of criterion c in plan p is given by the weighted difference between the highest and the lowest possible utility result. Hence, in our case it is *(5-1) * 0.25 = 1*. If c is not used in a plan p, the output range for *(c, p)* is considered *0*. The theoretic maximum of all output ranges here is determined by the range of the utility scale, which in the case of Plato ranges from 0 to 5. However, in fact no evaluation value in this plan may actually be *No*. Thus, the actual output range of a criterion c in plan p is given by the weighted difference between the highest and the lowest result of the utility function applied to the actual evaluation values.

To investigate how likely potential bad outcomes actually are for certain criteria and candidates, we are thus interested in the ratio between potential and actual impact. This relative output range (or Variation) corresponds to the question how far output ranges are in reality represented in the evaluation values or whether the occurring variance is much lower than the expected possible output range of a criterion. Apart from the output ranges averaged over all plans using a criterion, we can also relate the sums of potential and actual output ranges to the total number of plans to account for the frequency of usage. This is in particular relevant if we are not looking at a scenario and a criterion, but rather analyzing a set of scenarios and criteria.

Finally, a discrete, non-weighted aspect has to be considered. If a utility function contains the target 0 in the output, it has the potential to reject an alternative as unacceptable, independently of the criterion weight. This is a crucial element of the decision method. We are thus interested in

1. the *rejection potential* of a criterion, i.e. the utility functions with an output range including 0,
2. the *rejection* of a criterion, i.e. the amount of utility functions that reject alternatives due to a utility of 0, and
3. the *rejects* of a criterion, i.e. the amount of alternatives rejected.

When analyzing criteria sets, we need slightly adapted impact factors. While factors such as count and spread can be aggregated in a straightforward way, others would lead to misleading figures. For instance, simply summing up the average weights would neglect the fact that these averages are calculated based on the partial set Pc. To analyze criteria sets over the entire set P, we can thus only sum up discounted average weights. Table 3 lists the resulting impact factors for criteria sets.

ID	Factor	Definition
SIF1	Spread	Average spread of the criteria in the set
SIF2	Coverage	Percentage of plans using at least one of the criteria
SIF3	Weight	Sum of discounted average total weights
SIF4	Potential	Sum of discounted average potential ranges
SIF5	Maximum potential	Maximum compound potential ranges
SIF6	Range	Sum of discounted average ranges
SIF7	Maximum range	Maximum compound actual ranges
SIF8	Variation	Average of the relative output ranges
SIF9	Maximum variation	Average maximum of the relative output ranges
SIF10	Rejection potential count	Number of utility functions with output range including 0
SIF11	Rejection potential rate	Percentage of utility functions with output range including 0
SIF12	Rejection count	Number of utility functions rejecting alternatives
SIF13	Rejection rate	Percentage of utility functions rejecting alternatives
SIF14	Reject spread	Percentage of plans affected by a reject out of this set
SIF15	Reject count	Number of alternatives rejected
SIF16	Reject rate	Percentage of alternatives rejected

Table 3 Impact factors for sets of criteria

While this set of factors is mathematically simple and robust, it is clearly somewhat redundant. However, the exact factor to be used for answering a certain question has to consider a number of dimensions. To reduce the set of factors that need to be analyzed to answer specific questions and provide guidance on concrete analysis tasks, Section 7 will present analysis results for all factors on a set of 210 criteria from six case studies selected in a homogeneous problem space.

5 Tool support

Data relevant for impact assessment consists of real world preservation plans collected over the last years. These plans were created in the Preservation Planning Tool Plato and are available for analysis. This means that all knowledge necessary for impact assessment is already present in the data model of Plato. Data formalization established by assigning properties to decision criteria enables us to process this data automatically.

To support the systematic and repeatable assessment of decision criteria, we are developing an interactive, web-based analysis tool. This tool is compatible with the planning tool Plato and can be seen as a complementary addition to the primary planning workflow. It will thus enable decision makers to share their experience and in turn leverage the wisdom of their community's peers in anonymised ways by aggregating the experience that planners wish to share. The tool loads preservation plans from the planning tool's knowledge base. It processes and anonymises plans and presents the decision maker or analyst with a number of features that facilitate systematic analysis in search of answers to the questions posed above:

- The planner can select a set of plans to be considered, i.e. filter the scenario set to be analyzed.

- The planner can then dynamically select properties of interest. For each property, the tool calculates all impact factors described.
- The tool furthermore visualizes several attributes of interest for each property, such as the different utility functions defined in various plans, in graphical form. Finally, to enable the analysis of not single criteria, but criteria sets, the user can dynamically create hierarchical property sets that reflect natural groupings of criteria such as all format properties that are considered relevant. The user can thus analyze the properties of aggregate sets of criteria in flexible configurations. We will discuss several such sets in the next section.

Two modules used for data analysis were developed: The Knowledge Browser supports the analysis of single criteria. The Criteria Hierarchy Tool supports the analysis of criteria sets or hierarchies. These two Plato modules assess criteria impact based on preservation plans stored in Plato. As more and more plans are created using the formalised property model, Knowledge Browser and Criteria Hierarchy tool will produce more and more accurate results.

5.1 Requirements

The general requirements for the development of the modules are

- Provide general information about the database that is the subject of analysis.
- Provide criteria information
  - List all criteria and their characteristics
  - Show criteria usage statistics
  - Calculate criteria impact factors
- Provide a list of criteria sortable by impact
- Support the user at discovering and analyzing criteria set impacts.
  - Define criteria sets
  - Provide impact information for each criteria set
  - Compare different criteria set impacts

The key non-functional requirements are the following:

- *Usability*: It should be easy to view usage statistics for the criteria of interest. The tool should not hinder the user on the way to his wanted statistic. It should be easy to define criteria sets.
- *Data presentation*: Data should be clearly arranged and presented to make it easier to draw conclusions.
- *Performance*: Although the data model is of considerable complexity, significant delays at tool usage should be avoided.
- *Non-obtrusive real-time analysis*: Direct analysis of the productive database without making changes in the object model is essential.

5.2 Functionality

The following section provides a quick overview of the analysis tool, which is currently still in development. It will become a key part of the planning component and will be fully integrated in the first milestone release of the planning component.

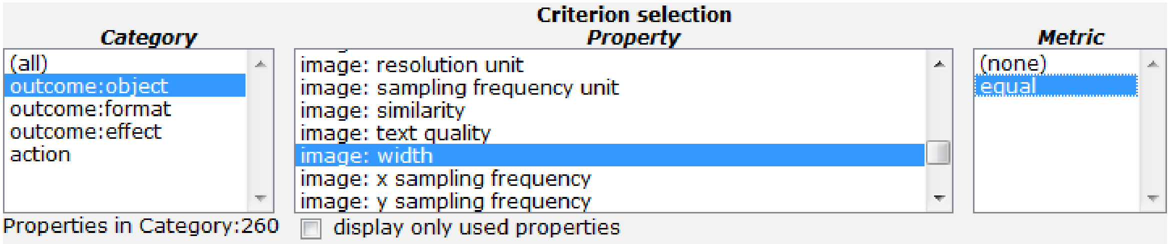
Figure 5 shows general statistics about the knowledge base of the browser, while Figure 6 shows the interface for selecting criteria in a hierarchical, flexible manner. It is possible to filter criteria for their usage in order to hide criteria that are defined in the knowledge base, but not used in the current analysis data set.

Knowledge browser

General Statistic	
relevant plans	6
overall leaves	239
mapped leaves	210

Property Statistic	
available properties	388
properties used at least once	124
available criteria	473
criteria used at least once	129

Figure 5 Knowledge Browser General Statistics



Criterion Characteristic	
property description	Width of an image, measured in number of pixels on the x-axis.
metric	Comparison of two values for equality
scale	Boolean

Figure 6 Criteria selector

Criterion Statistic	
leaves in category	47
leaves using property	6
leaves using criterion	6
AVG weight	0.325
AVG total weight	0.0238
measurements obtained	31
de facto standard : 4	
international standard : 18	
none : 9	
evaluations	31

Criterion Impact Factors	
IF1: Count	6
IF2: Spread	100 %
IF3: Weight	0.0238
IF4: Discounted Weight	0.0238
IF5: Potential	0.0976
IF6: Range	0.0821
IF7: Discounted Potential	0.0976
IF8: Discounted Range	0.0821
IF9: Maximum Potential	0.2
IF10: Maximum Range	0.2
IF11: Variation	0.713
IF12: Maximum Variation	1
IF13: Rejection Potential Count	2
IF14: Rejection Potential Rate	33.33 %
IF15: Rejection Count	1
IF16: Rejection Rate	16.67 %
IF17: Reject Count	1
IF18: Reject Rate	3.23 %

Figure 7 "Format standardization" criterion statistics

After selecting a certain criterion, the browser will display specific information about its usage. Figure 7 shows the criterion statistics of the criterion “format standardization”. What can be seen here is for example that this criterion is used in all case studies and that it has the potential to reject alternatives in one third of its usage cases. Figure 8 shows transformation rules (i.e. utility functions) for the criterion “initial software costs”. These curves define a step-wise mapping to utility scores – for instance, to gain the maximum score in the second case, the action must have initial costs of at most 500 €. They also define the minimum or maximum acceptable thresholds: In the second case, for example, anything with initial costs higher than 2500 € will be immediately rejected independent of other qualities.

Numeric Transformer	1.0	2.0	3.0	4.0	5.0
Thresholds (Steps)	10000.0	7500.0	5000.0	2500.0	0.0
Thresholds (Steps)	2500.0	2100.0	2000.0	1000.0	500.0

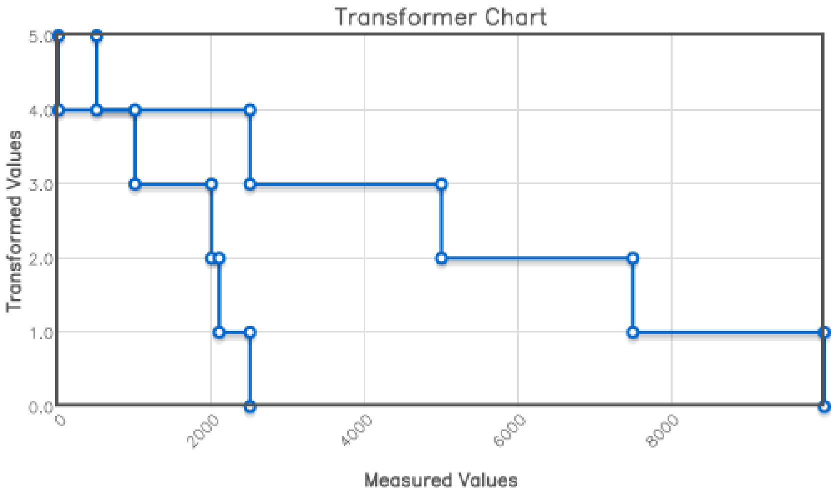


Figure 8 Simple visualisation of two different utility functions for the criterion “initial software costs”

More immediately, the browser also presents a table with all criteria that have been used at least once which includes all impact factors and can be sorted by any of the factors, as we will see in the next section. Finally, it presents a simple interface to construct arbitrary hierarchies of criteria, so that any hierarchical categorisation of properties and criteria can be represented and analysed for the impact of the contained sets of criteria on different levels of its hierarchy.

6 An Analysis of Case Studies

To illustrate the application of the above calculations and investigate the usefulness of our method and tool to answer the questions posed in the beginning, we analyse a set of related real-world case studies. We will first give an overview of recent cases and discuss the peculiarities of a set of related case studies, and then in Section 7 apply the framework presented in this report and conduct a quantitative impact analysis on a selected representative set of high-quality cases.

6.1 Overview of cases

Table 4 shows key characteristics of a number of recent case studies. For each case study, the following information is provided.

- *Organisation type* characterises the organisation carrying out the planning tasks.
- *Planning set* defines the set of objects in question (also called collection).

- *Scenario* characterises the planning setting: Some cases were coached by experts, while other cases relied only on publicly available documentation to guide their decisions. Furthermore, some cases were studying real business problems to solve a question the organisation was actively trying to tackle, while others carried out pilot studies to evaluate the usefulness of the approach for the institutions, and a third group was simply interested in the performance of potential actions for research purposes.
- *Criteria* denotes the number of quantitative decision criteria used in evaluation.
- *Alternatives denotes* the number of alternatives evaluated.
- *Key factors* distils the decision criteria that had the most critical effect on the performance of alternatives considered, e.g. by ruling out candidates because of unacceptable shortcomings.
- Chosen action denotes the recommended action resulting from the evaluation.

	Organisation type	Planning set	Scenario	Criteria	Alternatives	Key factors	Chosen action
1	National Library	Large collection of scanned images in TIFF-5 (80TB)	Coached business decision	24	7	Storage cost, standardisation, Automated QA	Convert to JPEG 2000
2	National Library	Large collection of scanned images in TIFF-6 (72TB)	Coached business decision	43	5	Colour profile complications, Lack of JPEG 2000 support	Keep status quo
3	National Library	Collection of scanned high-resolution images in TIFF-6	Coached business decision	35	3	Process costs, Native browser support	Keep status quo
4	National Library	Collection of complex PDF documents	Uncoached business decision	35	3	Migration quality, complexity of compound objects	Keep status quo
5	National Library	Small collection of scanned images in GIF	Coached evaluation	26	4	Format considerations	Convert to TIFF-6 (ImageMagick)
6	Research institution	Collection of publications in PDF versions	Uncoached evaluation	47	3	Transformation information properties, Format considerations	Migrate to PDF/A with PdfCreator
7	National Archive	Collection of legacy documents in WordPerfect versions	Uncoached Evaluation	38	3	Authentic reproduction of records	Emulate original viewer with Dioscuri
8	National Archive	Relational SQL databases	Coached pilot evaluation	67	2	Interactivity and behaviour not relevant, Documentation only	Convert to XML with SIARD
9	Computer museum (fictional)	Console video games (Nintendo SNES)	Coached evaluation research	81	4	Interactive gaming experience	Emulate with SNES9X 1.51 or ZSNES 1.51
10	Research institution	Video games for DOS	Uncoached evaluation research	44	5	Emulator compatibility, interactive gaming experience, audio/video quality	Emulate using DosBOX on Wine (Linux)
11	Professional Photographer	Digital photography camera raw files (CRW,CR2,NEF)	Coached evaluation research	69	7	Authentic object properties, colour reproduction, embedded metadata	Convert to DNG with Adobe DNG Converter (lossless compression)
12	Regional archive	Digital photography camera raw files (NEF)	Uncoached pilot evaluation	39	5	Format considerations, process control	Convert to TIF (Photoshop CS4)

Table 4 Recent case studies conducted in Plato

The number and type of stakeholders involved in decision making show some variation. For the case studies having to take productive business decisions, generally a key combination of decision makers from the organisation collaborated with software engineers and internal DP professionals. Research evaluation on the other hand was generally carried out by small academic research groups.

The first rows contain four related case studies that show several striking similarities. They all were analysing preservation actions for scanned images; they all took place in a national library; and they all were evaluating whether a migration to a more suitable format would decrease risks and lower long-term costs in return for an acceptable investment, while keeping all significant properties unchanged. Why did these cases lead to very distinct conclusions?

In the first case, storage costs were directly dependent on the file size and substantial; the file format was TIFF version 5, which is not a fully standardised format. Migration to the ISO-standardised lossless JPEG 2000 provided the opportunity to lower costs and risks without threatening the content. In the second case, the cost structures were different, and storage space less of an issue. Moreover, the images were already stored in version 6 of TIFF, which is recognised as an ISO standard. On the other hand, the particularities of the colour profiles embedded in the images made conversion risky and hindered automated quality assurance; thus, a migration would have incurred more costs than it could have saved. In the third case, the images were similarly stored in an ISO-standardised format, and thus leaving the images unchanged was a simple and safe solution. The access costs of creating derivative copies would not have been lowered with the usage of JPEG 2000, since current browsers do not natively support JPEG 2000, and the costs of migrating to JPEG 2000 were thus not considered worth the potential savings. In both cases, a monitoring task has been defined to watch upcoming browser support for JPEG 2000, as this may change the preference towards migration to JPEG 2000. Finally, in the fourth case, data volumes were relatively low and the benefit of a standardised format considered enough reason to recommend migration to TIFF-6 despite the increase in required storage.

The fact that the analysis of these closely related scenarios led to such different recommendations clearly demonstrates that a preservation action that is optimal in one situation does not necessarily address the problems of another scenario efficiently and effectively. It shows that preservation planning has to take into account the institution-specific preferences and constraints, the peculiarities of the content, and the specific context of each scenario. It also shows that the range of tools that are available for any specific migration perform differently, requiring detailed evaluation to identify the optimal solution.

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

In contrast to scanned images, digital camera files provide a different, often complex source of preservation risks, as different camera profiles contain diverse information encoded in an incompatible and often proprietary representation. Normalisation to a format such as the Digital Negative (DNG) clearly is a strong preference that comes to mind, but close examination of available migration paths and in-depth Quality Assurance is necessary to decide if a migration is possible and to select the migration path that is preferred in a specific context. The last two rows show two case studies dealing with camera raw files. In these two different settings, the photographer preferred conversion to DNG, while the archive preferred normalisation to TIFF-6. A detailed report on case 11, including a demonstration of a fully automated evaluation experiment, can be found in (Bauer & Becker, 2011).

The remaining case studies dealt with very diverse content and were conducted in a range of settings. A discussion on cases analysing options for interactive content can be found in [12].

6.2 Four cases, three solutions: Case studies for images

This section discusses four related exemplary case studies, each seeking the optimal preservation solution for large collections of scanned images. These case studies took place in four different national libraries in Europe. For each case, we will outline the scenario, the key factors and objectives, and the evaluation results. While significant properties of images are an essential aspect, they are relatively straightforward to define compared to complex objects such as video games, interactive art, or databases. We will thus focus our attention on the peculiarities that differentiate the case studies, including preservation processes and technical aspects.

6.2.1 Scanned newspapers

The first case study was carried out with the British Library and focused on a collection of 2 million images in TIFF-5 format with a size of about 40MB per image. The images were scanned from old newspaper pages; with 80TB of data volume this was the largest study in terms of size. Concerns were raised about the suitability of the Linear Tape Open (LTO) media on which the content was held, and the images were transferred to hard disk storage and reviewed. This move highlighted difficulties in 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 if 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, and
3. Original copies may be deleted.

These policies affect some of the criteria in the objective tree (in particular format properties) and on a different level the decision about migration: If it is possible to demonstrate that <u>all</u> significant properties are retained in a migration process that results in lower file size (without using lossy compression), the deletion of the original files can potentially save large amounts of storage expenses.

The objective tree as shown in Figure 9 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 estimated using the LIFE models. Table 5 shows the evaluated preservation actions and their aggregated scores. Conversion to BMP was ruled out prior to the experiment phase due to expected large file sizes and lack of compression, while GIF was discarded because of the palette size limitations.

The results show that migration to lossless JPEG2000 (JP2) using ImageMagick achieves a slightly higher root score than leaving the master files untouched. The reasons are that the long-term storage costs and the fact that JP2 is a recognised ISO standard [7] outweigh the process costs of converting the images. Conversion to JPEG or to lossy JP2 is violating the abovementioned policy that compression must be lossless, as included in the requirements tree under *File format – Compression*. Thus the corresponding alternatives have a multiplication score of 0.0 and are discarded as unacceptable alternatives. Conversion to PNG fares worse than JP2 due to significantly higher costs for big-stream preservation, despite slight advantages in terms of format properties (PNG is less complex and enjoys widespread tool support).

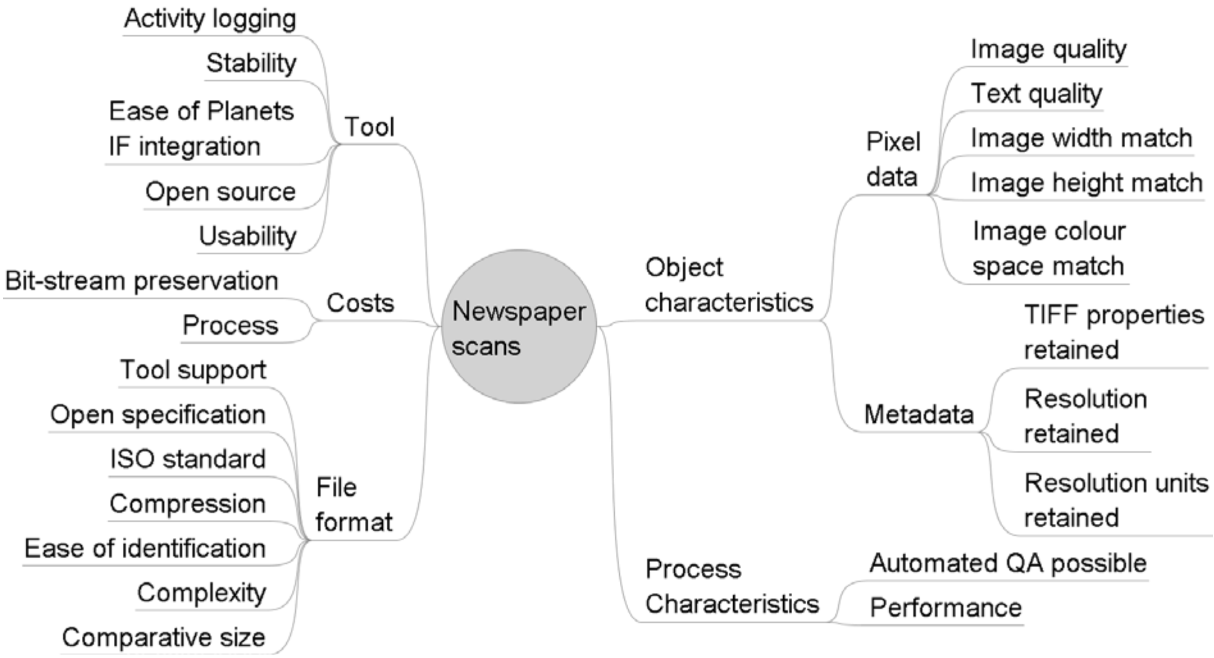


Figure 9 Scanned newspapers requirements tree

Candidate action	Weighted multiplication	Weighted sum
Keep status quo (TIFF-5)	3.01	3.46
Convert to PNG (ImageMagick)	2.72	3.27
Convert to BMP (ImageMagick)	-	-
Convert to GIF (ImageMagick)	-	-
Convert to JPEG (ImageMagick)	0.00	-
Convert to JP2 lossless (ImageMagick)	3.44	3.69
Convert to JP2 lossy (ImageMagick)	0.00	-

Table 5 Evaluation results for preservation actions on newspaper scans

6.2.2 Scanned books

A smilar study which examined the options for preserving a large collection of images scanned from 16th-century books held by the Bavarian State Library[iii] is presented in detail in (Kulovits, Rauber, Kugler, Brantl, Beinert, & Schoger, 2009). The collection contains 21.000 prints with about 3 million pages in TIFF-6, totalling 72TB 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 Figure 10. The considered actions were migration to JP2 with various conversion tools and leaving the objects unchanged. Storage itself does not pose significant constraints on this specific collection at the moment. The costs of the migration process, however, are dependent on the cost model of the computing facility to which the storage is outsourced. There, the pay-per-volume cost model depends on the volume of data that is retrieved from or (re-)ingested into the archive.

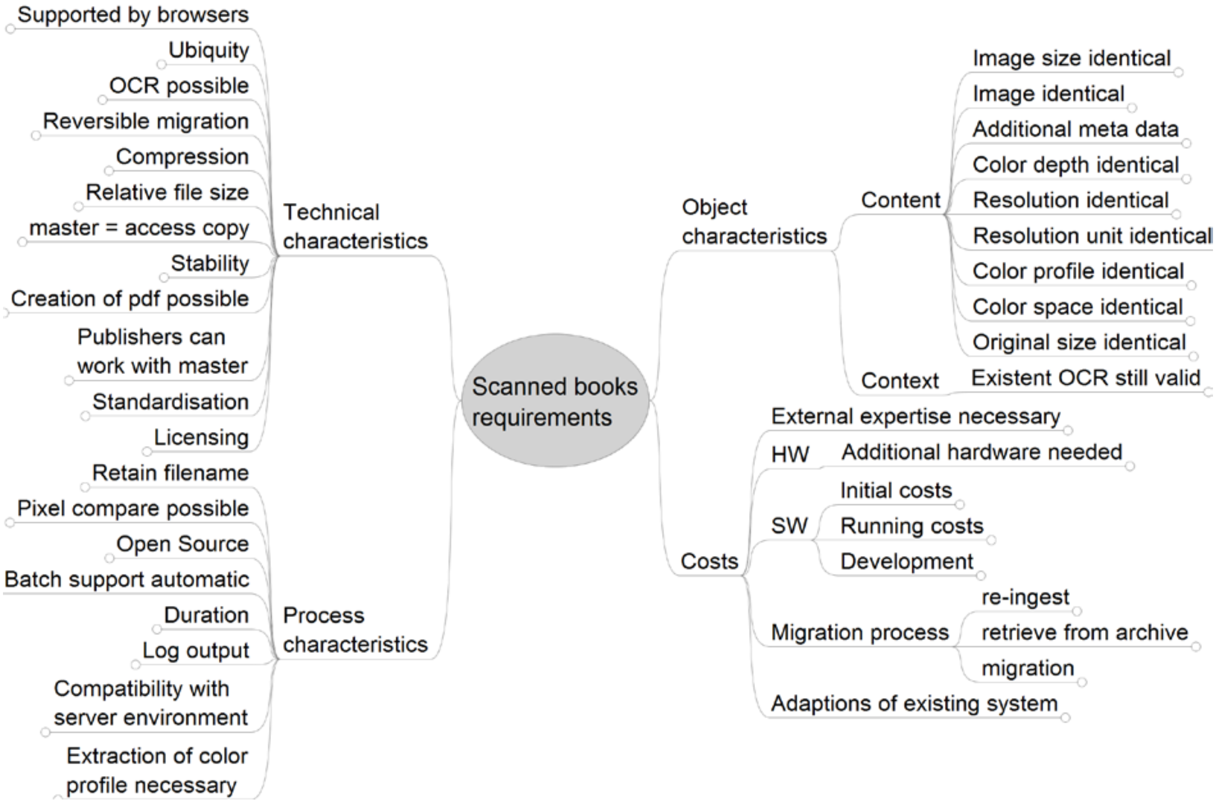


Figure 10 Scanned book pages requirements tree

Candidate action	Weighted multiplication	Weighted sum
Keep status quo (TIFF-6)	4.50	4.70
Convert to JP2 (ImageMagick)	3.71	4.09
Convert to JP2 (GraphicsMagick)	0.00	-
Convert to JP2 (Kakadu)	3.68	4.06
Convert to JP2 (GeoJasper)	3.65	4.03

Table 6 Evaluation results for preservation actions on scanned books

The evaluation results displayed in Table 6 show that leaving the images in TIFF-6 is the preferred option, even though JP2 has advantages such as reduced storage requirements and streaming support. The third alternative, conversion using GraphicsMagick, is rejected due to loss in the image data: Direct image comparison using both GraphicksMagick's and ImageMagick's compare functionality reveals that some pixel values are not identical.

This leaves four candidates that are compared using the weighted sum. The sensitivity analysis that is calculated automatically in the planning tool shows in this case that on the level of process criteria, there is sensitivity to changes in evaluation or weighting. The weighted aggregated utilities of the four alternatives with respect to the requirements branch *Process characteristics* all are between 1.04 and 1.14, and any shifts in the criteria *duration* or *costs* may eventually change the ranking of candidates within the process branch. However, this has no influence on the fact that overall, keeping the status quo is clearly preferred to the other three options; sensitivity analysis shows the robustness of the ranking on the root level. Storage will be monitored and the decision periodically reviewed.

6.2.3 Scanned negatives of aerial photographs

A third evaluation with a very similar scenario was carried out by the Royal Library of Denmark[iv], creating a preservation plan for digital safety copies representing original black-and-white cellulose nitrate negatives of aerial photographs stored as TIFF-6 images. Negatives in unstable condition are scanned in a high safety copy quality (1800 ppi, RGB, 16 bit) suitable for eventual replacement of the original material, while negatives in good condition are scanned in standard quality (1800 ppi, Greyscale, 8 bit).

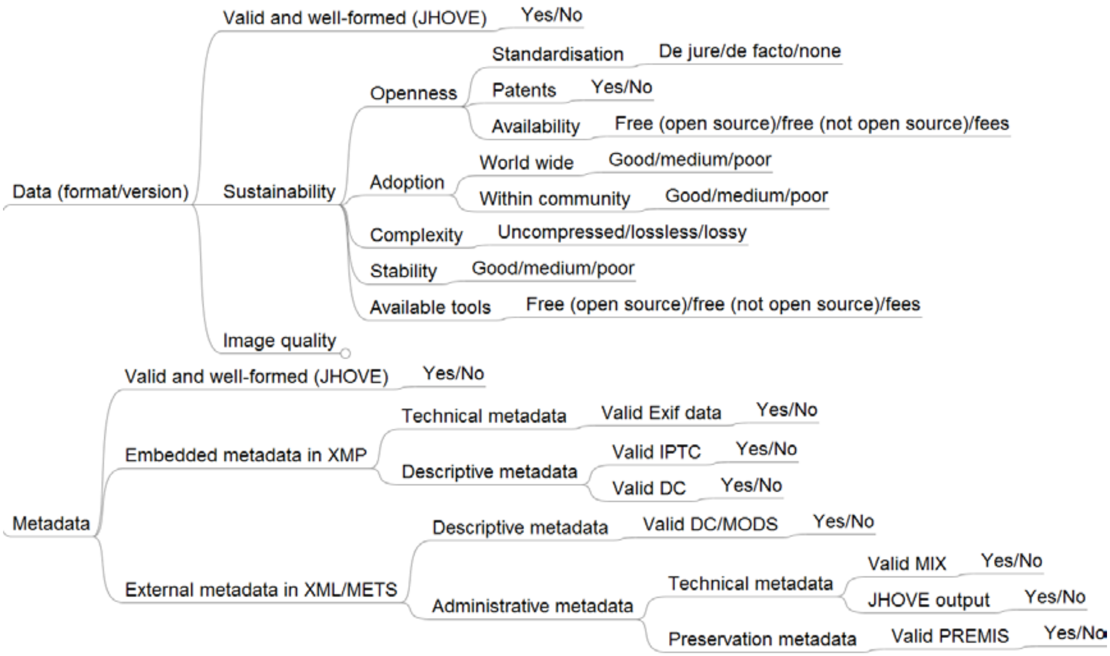


Figure 11 Aerial photographs requirements: Content, Format, Metadata

The rationale for evaluating alternative strategies to storing the large images in TIFF was again motivated by the potential cost savings on archival storage that can be achieved by the smaller file sizes of JP2. The evaluation focused on migration to JP2 and compared ImageMagick as widely available open source tool with the commercial solution LuraWave JP2 CLT (Command Line Tool).

Figure 11 shows the *object characteristics* branch defined in the study, which is separated into data and metadata. As indicated by the formulation of criteria, the evaluation procedure relied on the output of JHOVE to facilitate semi-automated evaluation of conversion quality. The evaluation values were compared manually and entered into the planning tool, but relied on the properties extracted by JHOVE. The structure of the requirements on object characteristics varies from the often-made distinction between the format and the `intellectual' properties and instead distinguishes between data (comprising both the format and the content characteristics) and the metadata, while describing the remaining aspects in a separate branch of the tree shown in Figure 12.

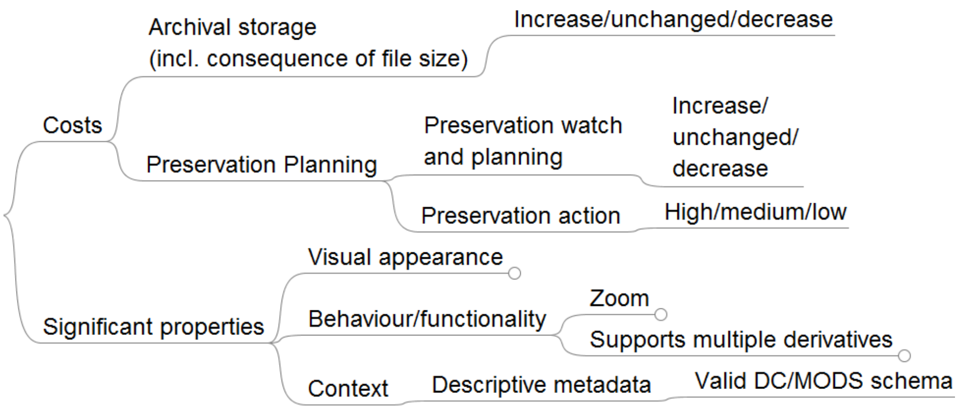
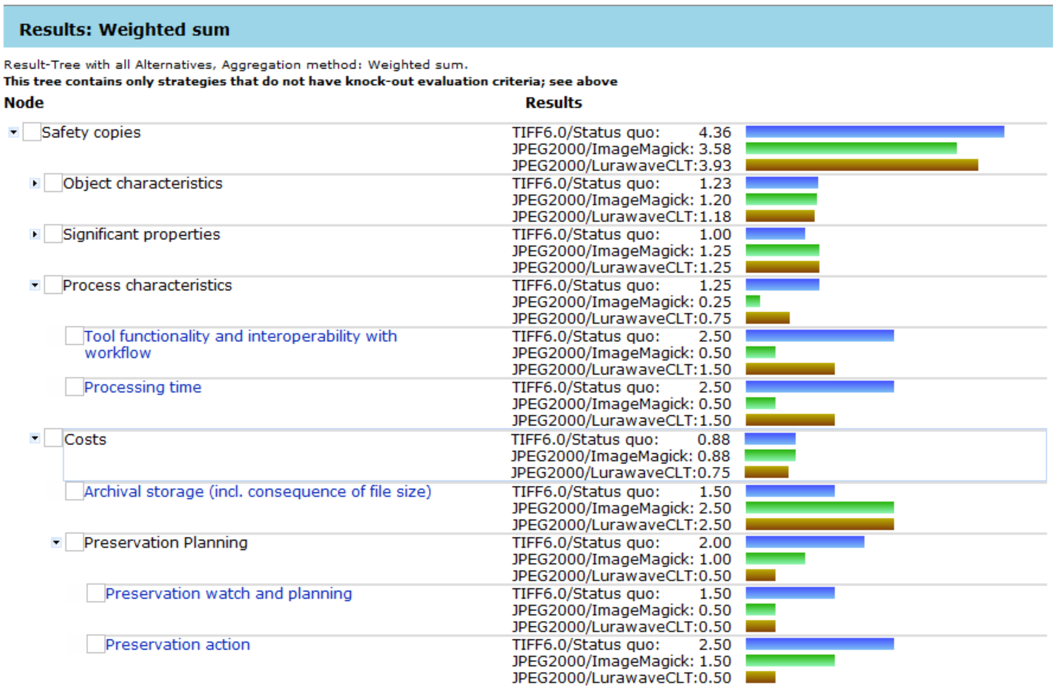


Figure 12 Aerial photographs requirements: Costs, Appearance, Behaviour, Context



Results: Weighted sum

Result-Tree with all Alternatives, Aggregation method: Weighted sum.
This tree contains only strategies that do not have knock-out evaluation criteria; see above

Node	Results
Safety copies	TIFF6.0/Status quo: 4.36 JPEG2000/ImageMagick: 3.58 JPEG2000/LurawaveCLT:3.93
Object characteristics	TIFF6.0/Status quo: 1.23 JPEG2000/ImageMagick: 1.20 JPEG2000/LurawaveCLT:1.18
Significant properties	TIFF6.0/Status quo: 1.00 JPEG2000/ImageMagick: 1.25 JPEG2000/LurawaveCLT:1.25
Process characteristics	TIFF6.0/Status quo: 1.25 JPEG2000/ImageMagick: 0.25 JPEG2000/LurawaveCLT:0.75
Tool functionality and interoperability with workflow	TIFF6.0/Status quo: 2.50 JPEG2000/ImageMagick: 0.50 JPEG2000/LurawaveCLT:1.50
Processing time	TIFF6.0/Status quo: 2.50 JPEG2000/ImageMagick: 0.50 JPEG2000/LurawaveCLT:1.50
Costs	TIFF6.0/Status quo: 0.88 JPEG2000/ImageMagick: 0.88 JPEG2000/LurawaveCLT:0.75
Archival storage (incl. consequence of file size)	TIFF6.0/Status quo: 1.50 JPEG2000/ImageMagick: 2.50 JPEG2000/LurawaveCLT:2.50
Preservation Planning	TIFF6.0/Status quo: 2.00 JPEG2000/ImageMagick: 1.00 JPEG2000/LurawaveCLT:0.50
Preservation watch and planning	TIFF6.0/Status quo: 1.50 JPEG2000/ImageMagick: 0.50 JPEG2000/LurawaveCLT:0.50
Preservation action	TIFF6.0/Status quo: 2.50 JPEG2000/ImageMagick: 1.50 JPEG2000/LurawaveCLT:0.50

Figure 13 Top level results for aerial photographs shown in Plato

An interesting aspect in the requirements hierarchy is the notion of several aspects of costs that are often neglected. The upper part of Figure 12 describes expected variations in costs in terms of archival storage (taking into account the file size), but also with respect to expected future efforts for planning and watch. The idea is that certain formats require constant attention and monitoring. While there are no exact estimates of costs – it was deemed infeasible to calculate these costs in exact figures – the directions are seen as useful indications.

The final evaluation results are shown in Figure 13 and Table 7. The analysis reveals a trade-off decision between one-time costs and running costs. TIFF-6 leads to higher storage costs, but wins in terms of watch and planning, and of course leaving the objects unchanged requires very little investment.

Candidate action	Weighted multiplication	Weighted sum
Keep status quo (TIFF-6)	4.14	4.36
Convert to JP2 (ImageMagick 6.4)	2.89	3.58
Convert to JP2 (Lurawave JP2 CLT)	3.51	3.93

Table 7 Evaluation results for preservation actions on aerial photographs

Process characteristics > Processing time

Results

Alternatives	1
TIFF6.0/Status quo	Fast
JPEG2000/ImageMagick	slow
JPEG2000/LurawaveCLT	medium

Transformer

Ordinal Value	Target Value
Fast	-> 5.0
medium	-> 3.0
slow	-> 1.0

Transformed Results

Alternatives	1	Aggregated	Comments
TIFF6.0/Status quo	5	5	status quo
JPEG2000/ImageMagick	1	1	
JPEG2000/LurawaveCLT	3	3	relative to status quo

Aggregation mode: Worst result

Costs > Archival storage (incl. consequence of file size)

Results

Alternatives	1
TIFF6.0/Status quo	unchanged
JPEG2000/ImageMagick	decrease
JPEG2000/LurawaveCLT	decrease

Transformer

Ordinal Value	Target Value
Increase	-> 1.0
unchanged	-> 3.0
decrease	-> 5.0

Transformed Results

Alternatives	1	Aggregated	Comments
TIFF6.0/Status quo	3	3	
JPEG2000/ImageMagick	5	5	
JPEG2000/LurawaveCLT	5	5	

Aggregation mode: Worst result

Preservation Planning > Preservation watch and planning

Results

Alternatives	1
TIFF6.0/Status quo	unchanged
JPEG2000/ImageMagick	Increase
JPEG2000/LurawaveCLT	Increase

Transformer

Ordinal Value	Target Value
Increase	-> 1.0
unchanged	-> 3.0
decrease	-> 5.0

Transformed Results

Alternatives	1	Aggregated	Comments
TIFF6.0/Status quo	3	3	
JPEG2000/ImageMagick	1	1	
JPEG2000/LurawaveCLT	1	1	

Aggregation mode: Worst result

Preservation Planning > Preservation action

Results

Alternatives	1
TIFF6.0/Status quo	low
JPEG2000/ImageMagick	medium
JPEG2000/LurawaveCLT	High

Transformer

Ordinal Value	Target Value
High	-> 1.0
medium	-> 3.0
low	-> 5.0

Transformed Results

Alternatives	1	Aggregated	Comments
TIFF6.0/Status quo	5	5	
JPEG2000/ImageMagick	3	3	
JPEG2000/LurawaveCLT	1	1	

Aggregation mode: Worst result

Figure 14 Example utility functions for process characteristics

6.2.4 Scanned yearbooks

The last case study was conducted with the State and University Library Denmark, evaluating the options for preserving a collection of scanned yearbooks published in the years 1965-1989. The images were stored in different versions of the GIF format. The storage costs were not as important, since the data volume was not as high as in the previously described studies. The objective tree is shown in Figure 15. In contrast to the other three cases, this study was not meant for productive decision making, but purely for evaluation purposes. Analysis and evaluation led to the recommendation to migrate the images to TIFF-6 despite the growth in file size.

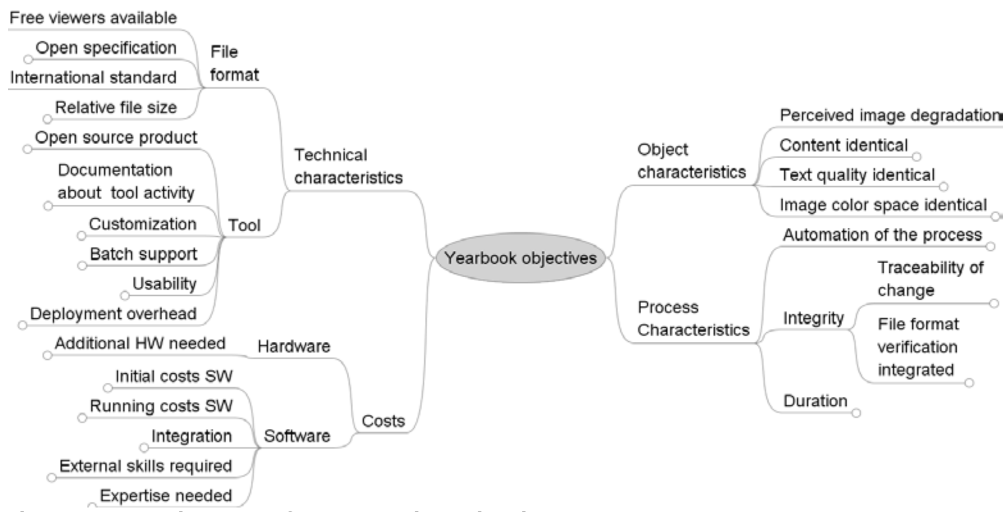
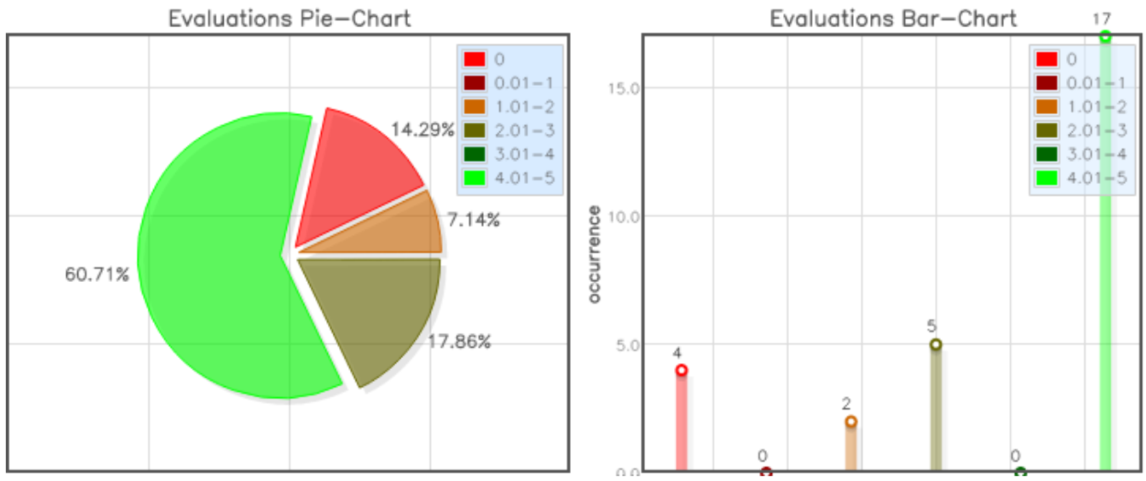


Figure 15 Requirements for scanned yearbooks

7 Results and Discussion

Our analysis case includes the plans 1, 2, 3, 5, 11, 12 shown in Table 4. This is a subset of the plans where all plans deal with image preservation. They contain a total of 239 decision criteria, of which 210 (87.9%) have been mapped to uniquely identified properties. (The remaining decision criteria all occur only in one plan and have a Rejection Potential of 0.) Out of 473 criteria currently available in the knowledge base of the planning tool, 129 are of relevance in the analysis set.



Ordinal Transformer	0.0	0.5	1.0	2.0	3.0	5.0
mapping	Lossy				None	Lossless
mapping	lossy					lossless, none
mapping		lossy		lossless		none
mapping	lossy					lossless, none
mapping			lossy		lossless	none

Figure 16 Visualization of the decision criterion “Format compression” used in five cases

The tool enables us to browse the criteria categories, select criteria, and analyze their properties and behaviour both in detail and through visualization. Figure 16 shows the tool displaying a visualization of the decision criterion *Format compression*. This is an *ordinal* criterion with the possible values *None, Lossless,* and *Lossy*. It is used frequently, in 5 of 6 plans, with an average total weight of 0.0276. The average potential output range is only 0.13, but 60% of the utility functions have rejection potential. The top left pie chart shows a distribution over utility output ranges. We can see that almost 15% of values are rejected. The top right shows the frequency distribution along the utility scale. On the bottom, we see the anonymised utility functions defined by the five plans in which this property was used. Clearly, lossy compression is always the worst case, but only in 3 out of 5 cases is it a reason for immediate rejection of an alternative. None is the option with the highest scores on average, but in one case, considered worse than Lossless compression. The accumulated knowledge can also be used to gain insight about typical preferences and support proactive recommendation of utility settings. The fact that lossy compression is in all utility functions dominated by lossless and compression-free encoding comes as no surprise as it corresponds to common knowledge in the community. In other cases, it will be valuable input for a recommender function that can base recommended utility curves for certain users on the accumulated insight of others having tackled comparable problems. In the case of lossless vs. none, it can be seen that there is no dominating value, since the preference of lossless vs. lossy compression depends on a number of factors (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011).

Impact Factors Table

Category	Property	Metric	IF1	IF2	IF3	IF4	IF5	IF6	IF7	IF8	IF9	IF10	IF11	IF12	IF13	IF14	IF15	IF16	IF17	IF18
OUTCOME_FORMAT	documentation availibility		6	100%	0,018	0,018	0,0812	0,0417	0,0812	0,0417	0,1562	0,1562	0,3667	1	2	33,33%	1	16,67%	1	3,23%
OUTCOME_FORMAT	standardization		6	100%	0,0238	0,0238	0,0976	0,0821	0,0976	0,0821	0,2	0,2	0,713	1	2	33,33%	1	16,67%	1	3,23%
OUTCOME_FORMAT	compression		5	83,33%	0,0276	0,023	0,1319	0,0815	0,1099	0,0679	0,2295	0,2	0,5133	1	3	60%	1	20%	4	14,29%
OUTCOME_OBJECT	image: similarity	equal	5	83,33%	0,0272	0,0227	0,1252	0,1002	0,1043	0,0835	0,3125	0,3125	0,6	1	2	40%	2	40%	2	8,33%
OUTCOME_OBJECT	relative filesize		5	83,33%	0,0273	0,0227	0,1362	0,0752	0,1135	0,0627	0,24	0,1562	0,5867	1	5	100%	1	20%	1	3,57%
OUTCOME_OBJECT	image: colour space	equal	4	66,67%	0,0303	0,0202	0,1497	0,0781	0,0998	0,0521	0,3125	0,3125	0,25	1	3	75%	1	25%	1	5,26%
ACTION	interoperability: batch processing support		4	66,67%	0,0189	0,0126	0,0799	0,0212	0,0533	0,0142	0,11	0,085	0,25	1	1	25%	0	0%	0	0%
ACTION	performance efficiency: time per MB		4	66,67%	0,0638	0,0425	0,3188	0,1998	0,2125	0,1332	0,625	0,5	0,5988	0,8	4	100%	0	0%	0	0%
ACTION	portability: ease of integration in current environment		4	66,67%	0,0484	0,0323	0,2056	0,1403	0,1371	0,0935	0,5	0,5	0,4125	1	2	50%	0	0%	0	0%
OUTCOME_FORMAT	speed of change		4	66,67%	0,0237	0,0158	0,0929	0,0421	0,0619	0,0281	0,126	0,084	0,4583	0,6667	1	25%	0	0%	0	0%
OUTCOME_FORMAT	ubiquity		4	66,67%	0,036	0,024	0,168	0,1326	0,112	0,0884	0,35	0,28	0,6722	0,8889	1	25%	0	0%	0	0%
ACTION	usability: ease of use		4	66,67%	0,0281	0,0187	0,0961	0,0481	0,0641	0,032	0,16	0,16	0,375	1	0	0%	0	0%	0	0%
ACTION	business: costs: setup: software licence inital		3	50%	0,0322	0,0161	0,1608	0,0754	0,0804	0,0377	0,3	0,2096	0,2996	0,6988	3	100%	0	0%	0	0%
ACTION	business: licencing schema		3	50%	0,0229	0,0115	0,0983	0	0,0492	0	0,11	0	0	0	1	33,33%	0	0%	0	0%
ACTION	functional suitability: activity traceability		3	50%	0,03	0,015	0,1358	0,0817	0,0679	0,0408	0,17	0,17	0,5333	1	2	66,67%	0	0%	0	0%
ACTION	functional suitability: output-file format verification		3	50%	0,0382	0,0191	0,1263	0	0,0631	0	0,17	0	0	0	0	0%	0	0%	0	0%
OUTCOME_OBJECT	image: height	equal	3	50%	0,0258	0,0129	0,0938	0	0,0469	0	0,18	0	0	0	2	66,67%	0	0%	0	0%
OUTCOME_OBJECT	image: width	equal	3	50%	0,0258	0,0129	0,0938	0	0,0469	0	0,18	0	0	0	2	66,67%	0	0%	0	0%
OUTCOME_FORMAT	# of free tools (not open source)		2	33,33%	0,0075	0,0025	0,0362	0,0312	0,0121	0,0104	0,0625	0,0625	0,5	1	1	50%	1	50%	1	14,29%
OUTCOME_FORMAT	# of free tools (open source)		2	33,33%	0,0106	0,0035	0,0519	0,0519	0,0173	0,0173	0,0938	0,0938	1	1	1	50%	1	50%	1	14,29%
OUTCOME_EFFECT	archival storage		2	33,33%	0,2425	0,0808	0,79	0,665	0,2633	0,2217	1,08	1,08	0,75	1	0	0%	0	0%	0	0%
OUTCOME_EFFECT	automated quality assurance possible		2	33,33%	0,0538	0,0179	0,2688	0,2	0,0896	0,0667	0,4	0,4	0,5	1	2	100%	1	50%	4	33,33%
ACTION	business: costs: execution: software develop new features		2	33,33%	0,0189	0,0063	0,0944	0,0354	0,0315	0,0118	0,1063	0,0495	0,4	0,6	2	100%	0	0%	0	0%
ACTION	business: costs: execution: software licence running		2	33,33%	0,0191	0,0064	0,0956	0	0,0319	0	0,1063	0	0	0	2	100%	0	0%	0	0%
ACTION	business: costs: setup: initial hardware		2	33,33%	0,0875	0,0292	0,4375	0,05	0,1458	0,0167	0,625	0,1	0,2	0,4	2	100%	0	0%	0	0%
ACTION	business: costs: setup: other		2	33,33%	0,035	0,0117	0,15	0,1	0,05	0,0333	0,2	0,2	0,5	1	1	50%	0	0%	0	0%
ACTION	business: costs: setup: staff external expertise needed		2	33,33%	0,0356	0,0119	0,1781	0	0,0594	0	0,25	0	0	0	2	100%	0	0%	0	0%
ACTION	business: costs: total		2	33,33%	0,0513	0,0171	0,2563	0,205	0,0854	0,0683	0,3125	0,25	0,8	0,8	2	100%	0	0%	0	0%
OUTCOME_FORMAT	embed custom metadata		2	33,33%	0,0232	0,0078	0,0994	0,042	0,0331	0,014	0,1148	0,084	0,5	1	0	0%	0	0%	0	0%
ACTION	functional suitability: error reporting		2	33,33%	0,0361	0,012	0,1705	0,09	0,0568	0,03	0,18	0,18	0,5	1	1	50%	0	0%	0	0%
OUTCOME_OBJECT	image metadata: tiff properties	retained	2	33,33%	0,0102	0,0034	0,0239	0,0239	0,008	0,008	0,0412	0,0412	1	1	0	0%	0	0%	0	0%
OUTCOME_OBJECT	image: bits per sample	equal	2	33,33%	0,0108	0,0036	0,05	0	0,0167	0	0,0688	0	0	0	1	50%	0	0%	0	0%
OUTCOME_OBJECT	image: ICC colour profile	equal	2	33,33%	0,0108	0,0036	0,0431	0,0275	0,0144	0,0092	0,055	0,055	0,5	1	0	0%	0	0%	0	0%
OUTCOME_OBJECT	image: image quality		2	33,33%	0,0492	0,0164	0,2282	0,2282	0,0761	0,0761	0,3125	0,3125	1	1	1	50%	1	50%	1	9,09%
OUTCOME_OBJECT	image: resolution	equal	2	33,33%	0,0141	0,007	0,0562	0,0183	0,0281	0,0092	0,0825	0,055	0,3333	1	0	0%	0	0%	0	0%
OUTCOME_OBJECT	image: resolution unit	equal	2	33,33%	0,0244	0,0081	0,1063	0,1063	0,0354	0,0354	0,1575	0,1575	1	1	0	0%	0	0%	0	0%
OUTCOME_OBJECT	image: similarity	SSIM simple saturation	2	33,33%	0,0197	0,0066	0,0984	0,0357	0,0328	0,0119	0,15	0,0714	0,238	0,476	2	100%	0	0%	0	0%
OUTCOME_OBJECT	image: similarity	Equality judgement	2	33,33%	0,0197	0,0066	0,0811	0,06	0,027	0,02	0,12	0,12	0,5	1	0	0%	0	0%	0	0%
OUTCOME_OBJECT	image: similarity	SSIM simple hue	2	33,33%	0,0197	0,0066	0,0984	0,0236	0,0328	0,0079	0,15	0,0471	0,157	0,314	2	100%	0	0%	0	0%
OUTCOME_OBJECT	image: text quality		2	33,33%	0,0492	0,0164	0,2282	0,2282	0,0761	0,0761	0,3125	0,3125	1	1	1	50%	1	50%	1	9,09%
ACTION	portability: runs on Linux		2	33,33%	0,0042	0,0014	0,0158	0,0138	0,0053	0,0046	0,0276	0,0276	0,5	1	0	0%	0	0%	0	0%
ACTION	portability: runs on MacOS		2	33,33%	0,0041	0,0014	0,0153	0	0,0051	0	0,0276	0	0	0	0	0%	0	0%	0	0%
ACTION	portability: runs on Windows		2	33,33%	0,0041	0,0014	0,0165	0	0,0055	0	0,0276	0	0	0	0	0%	0	0%	0	0%

Figure 17 Impact factors of all frequently (≥ 33%) used criteria of the case study set, sorted by descending IF1 (count)

Figure 17 shows a raw view on the most frequently used criteria as displayed in the current version of the analysis tool. Clearly, the meaning of all these numbers is not immediately accessible to a decision maker and will require interpretation by systematic tools, since the question which factors to consider depends entirely on the scope of interest.

Essentially, non-discounted factors will be of interest once we have decided to include a decision criterion or criteria set: They refer to the set of plans that use the criterion or set. On the other hand, if we have not decided upon inclusion or are not thinking about a concrete scenario, we need discounted factors to investigate the relative importance and the cumulative impact across multiple plans. Similarly, the pure counts are not very helpful and the corresponding indicators only become meaningful when used relatively with respect to the size of the criteria set and the size of the set of plans. However, indicators such as the *rejection potential* of criteria can provide good indicators for the criticality of a certain aspect of interest.

The raw statistics of single criteria thus present an important basis on which to assess specific criteria in certain situations. However, for the purpose of this report, logical criteria sets such as those discussed in Section 3 are much more interesting.

To illustrate the accumulated impact of such sets, we used the property hierarchy builder in the analysis tool and specified a number of criteria sets in correspondence to the models discussed above.

Table 8 shows the application of the impact factors for sets of criteria on one single plan. We discussed the case of the scanned newspapers in Section 6.2.1, mentioning the primary drivers of reducing costs through migration, the fact that primary copies could be deleted so that storage costs could *decrease* through a space-saving conversion, and the strong preference for open formats.
We can see this reflected on the decision criteria level in the impact factors for formats and *outcome effects*, which include storage costs. They have rejection potential and reject alternatives, and they also have the highest potential. Apart from these, functional correctness has high potential, but the tested alternatives perform quite equally, so that the potential is not realised.

This analysis is interesting to judge the impact of policies in the decision making process. The analysis results get much wider significance, however, when calculating impact factors across plans. Table 9 shows these sets and their impact factors in the image cases outlined in the previous section. A number of observations can be drawn.

Format criteria are relevant in all plans, with *coverage* being 100%. Their compound weight is 0.18. They achieve a maximum compound range of 0.86. On average, format properties exhaust a maximum of 33% of their utility range. This criteria set contains 17 utility functions with rejection potential. Every second plan in our set is affected by actual rejects caused by these criteria.

Name	size	SIF1	SIF2	SIF3	SIF4	SIF5	SIF6	SIF7	SIF8	SIF9	SIF10	SIF11	SIF12	SIF13	SIF14	SIF15	SIF16
Format	31	19,35%	100%	0,19	0,832	0,832	0,58	0,58	0,113	0,113	3	50%	1	16,67%	100%	4	57,14%
SQ Performance Efficiency	7	14,29%	100%	0,02	0,1	0,1	0,08	0,08	0,114	0,114	1	100%	0	0%	0%	0	0%
SQ Time Behaviour	3	33,33%	100%	0,02	0,1	0,1	0,08	0,08	0,267	0,267	1	100%	0	0%	0%	0	0%
SQ Resource Utilization	3	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%
SQ Capacity	1	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%
SQ Functional completeness	44	2,27%	100%	0,02	0,1	0,1	0,02	0,02	0,005	0,005	1	100%	0	0%	0%	0	0%
SQ Maintainability	3	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%
SQ Usability	6	16,67%	100%	0,02	0,08	0,08	0	0	0	0	0	0%	0	0%	0%	0	0%
SQ Portability	5	20%	100%	0,02	0,1	0,1	0,04	0,04	0,08	0,08	1	100%	0	0%	0%	0	0%
SQ Reliability	8	12,50%	100%	0,02	0,08	0,08	0,04	0,04	0,062	0,062	0	0%	0	0%	0%	0	0%
SQ Compatibility	5	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%
Business	16	12,50%	100%	0,06	0,3	0,3	0,16	0,16	0,05	0,05	2	100%	0	0%	0%	0	0%
SQ Funct. Correctness (outcome object)	223	4,04%	100%	0,21	0,898	0,898	0,348	0,348	0,026	0,026	4	44,44%	0	0%	0%	0	0%
SQ Funct. Correctness - RIP	12	8,33%	100%	0,01	0,05	0,05	0,04	0,04	0,067	0,067	1	100%	0	0%	0%	0	0%
SQ Funct. Correctness - IP	76	2,63%	100%	0,072	0,288	0,288	0,288	0,288	0,026	0,026	0	0%	0	0%	0%	0	0%
SQ Funct. Correctness - TIP	135	4,44%	100%	0,128	0,56	0,56	0,02	0,02	0,022	0,022	3	50%	0	0%	0%	0	0%
SQ Funct. Correctness - Img. Similarity	12	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%
Outcome Effects	3	66,67%	100%	0,44	1,48	1,48	1,48	1,48	0,667	0,667	1	50%	1	50%	100%	4	57,14%

Table 8: Criteria sets and their cumulative impact factors in the scanned newspaper case-study

Name	size	SIF1	SIF2	SIF3	SIF4	SIF5	SIF6	SIF7	SIF8	SIF9	SIF10	SIF11	SIF12	SIF13	SIF14	SIF15	SIF16
Format	31	25,27%	100%	0,183	0,812	1,397	0,435	0,864	0,327	0,42	17	36,17%	6	12,77%	50%	8	25,81%
SQ Performance Efficiency	7	11,90%	83,33%	0,048	0,234	0,625	0,155	0,5	0,228	0,257	4	80%	0	0%	0%	0	0%
SQ Time Behaviour	3	22,22%	66,67%	0,042	0,213	0,625	0,133	0,5	0,2	0,267	4	100%	0	0%	0%	0	0%
SQ Resource Utilization	3	5,56%	16,67%	0,005	0,021	0,129	0,021	0,129	0,333	0,333	0	0%	0	0%	0%	0	0%
SQ Capacity	1	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%
SQ Functional completeness	44	4,55%	83,33%	0,063	0,261	0,428	0,115	0,244	0,081	0,103	5	41,67%	0	0%	0%	0	0%
SQ Maintainability	3	5,56%	16,67%	0,003	0,013	0,08	0,003	0,02	0,083	0,083	0	0%	0	0%	0%	0	0%
SQ Usability	6	11,11%	66,67%	0,019	0,064	0,16	0,032	0,16	0,062	0,167	0	0%	0	0%	0%	0	0%
SQ Portability	5	33,33%	100%	0,036	0,153	0,5	0,098	0,5	0,182	0,4	2	20%	0	0%	0%	0	0%
SQ Reliability	8	4,17%	33,33%	0,009	0,035	0,129	0,007	0,04	0,062	0,062	0	0%	0	0%	0%	0	0%
SQ Compatibility	5	13,33%	66,67%	0,013	0,053	0,11	0,014	0,085	0,05	0,2	1	25%	0	0%	0%	0	0%
Business	16	20,83%	83,33%	0,124	0,601	1,335	0,195	0,366	0,187	0,269	17	85%	0	0%	0%	0	0%
SQ Funct. Correctness (outcome object)	223	7,40%	100%	0,274	1,205	1,409	0,578	1,406	0,224	0,24	23	23%	6	6%	33,33%	3	9,68%
SQ Funct. Correctness - RIP	12	18,06%	100%	0,053	0,236	0,734	0,063	0,156	0,049	0,083	5	38,46%	1	7,69%	16,67%	1	3,23%
SQ Funct. Correctness - IP	76	0,88%	33,33%	0,033	0,152	0,625	0,152	0,625	0,026	0,026	2	50%	2	50%	16,67%	1	9,09%
SQ Funct. Correctness - TIP	135	10%	100%	0,188	0,817	1,285	0,363	0,876	0,343	0,367	16	19,51%	3	3,66%	33,33%	2	6,45%
SQ Funct. Correctness - Img. Similarity	12	16,67%	83,33%	0,047	0,222	0,69	0,13	0,401	0,148	0,256	7	58,33%	2	16,67%	33,33%	2	8,33%
Outcome Effects	3	27,78%	50%	0,109	0,395	1,48	0,309	1,48	0,583	0,833	2	40%	1	20%	16,67%	4	26,67%

Table 9 Criteria sets and their cumulative impact factors in the image scenarios

Name	size	SIF1	SIF2	SIF3	SIF4	SIF5	SIF6	SIF7	SIF8	SIF9	SIF10	SIF11	SIF12	SIF13	SIF14	SIF15	SIF16
Format	31	12,90%	60%	0,077	0,327	0,993	0,077	0,214	0,105	0,14	7	35%	0	0%	0%	0	0%
SQ Performance Efficiency	7	11,43%	80%	0,038	0,192	0,312	0,112	0,25	0,166	0,229	4	100%	1	25%	20%	1	6,67%
SQ Time Behaviour	3	26,67%	80%	0,038	0,192	0,312	0,112	0,25	0,387	0,533	4	100%	1	25%	20%	1	6,67%
SQ Resource Utilization	3	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%
SQ Capacity	1	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%
SQ Functional Completeness	44	5,91%	80%	0,083	0,394	1,305	0,014	0,068	0,014	0,014	9	69,23%	5	38,46%	20%	2	18,18%
SQ Maintainability	3	6,67%	20%	0,002	0,009	0,045	0	0	0	0	0	0%	0	0%	0%	0	0%
SQ Usability	6	16,67%	60%	0,054	0,27	0,8	0,01	0,048	0,083	0,167	4	80%	0	0%	0%	0	0%
SQ Portability	5	12%	60%	0,037	0,182	0,8	0,001	0,006	0,025	0,05	2	66,67%	0	0%	0%	0	0%
SQ Reliability	8	5%	40%	0,009	0,044	0,175	0	0	0	0	1	50%	0	0%	0%	0	0%
SQ Compatibility	5	32%	80%	0,017	0,053	0,1	0,011	0,045	0,45	0,6	1	12,50%	0	0%	0%	0	0%
Business	16	25%	100%	0,176	0,86	1,75	0,255	0,549	0,153	0,275	19	95%	1	5%	20%	1	5,56%
SQ Funct. Correctness (outcome object)	223	9,60%	100%	0,406	1,862	3,642	1,283	3,457	0,167	0,205	62	56,36%	32	29,09%	60%	5	31,25%
SQ Funct. Correctness - RIP	12	5%	60%	0,034	0,168	0,495	0,014	0,07	0,01	0,03	3	100%	1	33,33%	20%	2	20%
SQ Funct. Correctness - IP	76	3,68%	60%	0,048	0,15	0,394	0,037	0,129	0,058	0,092	1	6,67%	0	0%	0%	0	0%
SQ Funct. Correctness - TIP	135	13,33%	100%	0,325	1,544	3,642	1,231	3,457	0,243	0,284	58	63,04%	31	33,70%	40%	3	21,43%
Outcome Effects	3	0%	0%	0	0	0	0	0	0	0	0	0%	0	0%	0%	0	0%

Table 10 Criteria sets and their cumulative impact factors in the document scenarios

Several **aspects of actions** are normally included in evaluation, but have very little impact on the decisions (Maintainability, Usability, Portability, Reliability and Compatibility).

Performance efficiency, on the other hand, has rejection potential, but none of the tested alternatives was rejected because of performance efficiency drawbacks. On the other hand, the fact that in our SQUARE categorisation, none of the image plans considered the attribute *capacity* points to a possible omission. Closer analysis reveals that (1) the maximum capacity of tools was generally tested implicitly by including the largest objects of the collection as samples (verifying size capacity), and (2) batch processing integration was intended to be done on a per-object basis, with removed the need for testing e.g. the maximum number of files that can be processed in one run, since it does not depend on the tools used. Still, it would be advisable to include an explicit notion of capacity in the criteria in future studies. This is a typical case where the explicit modelling of criteria and quality attributes can increase the quality of requirements specification: Automated quality checks can point decision makers to potential omissions where they have not considered certain aspects of SQUARE quality attributes that may be relevant. This is an obvious area of improvement for the SCAPE planning component over the existing planning tool.

Functional completeness and portability have little impact on the decisions. It should be noted that some of these have no rejects simply because the alternative actions were already discarded prior to evaluation, since their properties can be investigated prior to costly experiment runs. The same thing, obviously, cannot be said with sufficient certainty about specific significant properties or dynamic scalability characteristics of actions. Hence, these are evaluated experimentally and may lead to rejects. This means that factors that don't utilise their rejection potential aren't necessarily over-defensively specified or irrelevant, and should be as carefully documented as those that do.

Business factors of actions, which include costs and licensing, have had a much higher distinguishing effect in case studies. **Representation Instance Properties**, such as *Format is well-formed and valid*, have a high rejection potential and do lead to rejection in one case.

The most important group of criteria, of course, is concerned with **significant properties** (Transformation Information Properties), which can also seen as belonging to the **functional correctness** of performed actions. Every third plan is affected by a reject caused by a loss of authenticity in content preservation actions. The maximum compound change caused by criteria of this set is substantial with 1.28. We can further see the impact factors of the specific subset of 12 criteria describing different metrics to assess image similarity (ranging from error measures and advanced metrics such as structural similarity to subjective assessment). It can be seen that they account for a significant part of the significant properties.

Finally, all 64 **action** criteria together (without functional correctness) have not a single reject. However, they achieve an accumulated potential output change of 2.5 and realize this potential to a large degree, with a maximum compound actual range of 1.25. Thus, while they may not be responsible for an immediate reject, they should not be neglected in business environments. Among a range of alternative actions that perform acceptably well, they will often be the distinguishing factors that drive an organisation's choice.

To complement the analysis, Table 10 shows the same criteria set and the impact factors for the case studies on electronic documents (cases 4,6 and 7 of Table 4). Most of these case studies have been conducted without expert coaching, and none of these has taken productive business decisions. This may explain why a number of aspects are not considered in the set. However, it is interesting to see that the general trend of impact factors is very similar. A few differences can be observed: Several groups have rejection that only exhibit rejection *potential* in the other data set, namely performance efficiency and functional completeness.

Finally, it can be noted that SIF1, the *average spread* of criteria in a set, is rather meaningless, since it favours small or homogeneous criteria sets. For the analysis purposes of *sets* envisioned, it is generally sufficient to consider the coverage. Further analysis over the project should be able to reduce and simplify the factors.

Table 11 summarises the factors discussed above, categorises the main impact factors as H(igh), M(edium) and L(ow) and maps the distinct groups onto potential measurement sources in SCAPE. The P2 registry is what is currently used in the planning tool as a measurement source for formats. However, it has to be noted that its content is, as all other format registries, still scarce for most format groups (Becker & Rauber, Decision criteria in digital preservation: What to measure and how, 2011). Format registries are not a core focus of SCAPE; it is foreseen to integrate emerging data sources such as UDFR as they become available. It is also possible that certain format properties (such as the frequency a format occurs within a certain content type) will be covered by collection profiling and web watch services. The table also includes a judgment of volatility, estimating how high the variation of these criteria is over time and among specific collections.

Factor group	Size	Coverage	Reject potential	Rejection	Impact	Volatility	SCAPE measurement sources
Format	31	H	H	H	H	L/M	P2, UDFR, Watch services
Action: Performance	7	H	H	M	M	H	Workflow wrapper, Platform
Action: other	57	H	H	L/M	H	L/M	Component catalogue
Representation Instance Criteria	12	M	H	M	L	H	Characterisation components
Transformation Information Criteria	80	H	H	H	H	H	Quality Assurance

Table 11 Summary of decision factors on the case studies

It should be noted that these impact factors have to be interpreted with care. They are arguably representative of the Testbed scenario for Large Scale Digital Repositories (in fact, several of the case studies come from the involved organisations). It seems reasonable to assume that the general observations made will not be substantially different for similar scenarios. However, the accuracy of these factors will have to be revisited once SCAPE evaluation runs have taken place and increased the amount of data for these and other scenarios. As shown, the capability to immediately rerun the evaluation is in place in prototypical form and will become part of the planning component.

8 Outlook

This report has presented a systematic approach for analysis of decision criteria in preservation planning. We discussed the reconciliation of quality models describing properties of preservation action components as well as formats. We further presented a method and tool for quantitative impact assessment of decision criteria, and discussed results of applying the assessment to a series of real-world case studies.

One application of this analysis will be a reduction of complexity and manual effort of preservation planning through a reduction of the number of alternatives that are evaluated in depth. This can be supported by an early filtering of candidate actions based on correlating organizational goals and constraints (for example as expressed in the policy model) with documented knowledge as well as experience shared by other organizations. The outcome of this analysis thus provides directions to optimise and automate decision-making, watch, and policy definitions at large scales. It also enables smaller organizations to make decisions at a lower entry barrier by prioritizing essential evaluation factors in the evaluation phase. For example, criteria that have no rejection potential and a very low potential output range could in preliminary evaluation be substituted with aggregated experience from analogous scenarios. Automated robustness analysis could then clarify which of the substituted criteria have a decisive impact and need to be evaluated for the scenario. While this can certainly not replace a full-depth evaluation in all cases, it may provide a tool for trade-off decisions between risks and the costs of planning, and furthermore serve as a basis for solid planning.

Similarly, it can increase the focus and impact of research in characterization and Quality Assurance by prioritizing aspects that have the strongest impact. The identification and unambiguous specification of decision criteria across scenarios and organizations has additional benefits. In addition to the questions posed in the beginning of this report, questions such as 'Is a given set of criteria complete for the scenario at hand? Which other criteria may be relevant?' are of relevance and need to be systematically addressed. This can for example lead to proactive recommendation of decision factors to stakeholders, and to an intelligent monitoring service that raises alerts when certain conditions have been discovered to be of relevance by stakeholders outside an organization. This accumulated experience sharing shall advance the knowledge base of DP operations considerably and enable the transition to a continuous and continuously optimizing management activity. Current and future work is focused on expanding the data set to cover additional scenarios, completing the mapping of metrics to the SQUARE quality model for other scenarios, and using the information obtained to provide advanced, proactive decision support within the preservation planning procedure. This includes a further quantitative in-depth assessment of impact factors to arrive at a reduced and simple-to-use set of factors targeted at specific investigation scenarios.
The tools developed as part of this task will be integrated into the planning component to enable us to continuously analyse new plans as they are created. To facilitate the creation of plans where criteria are already aligned with this model from the start, the knowledge base model of the planning component will be extended to include the presented decision criteria models, including the mapping to quantified metrics. At the same time, it is crucial to ensure that the data produced by SCAPE components such as QA tools are linked to the measurable properties that are of relevance.

To enable this continuous analysis, we will also require a controlled process for the moderated publication of plans in the knowledge base, including safeguards for protecting unauthorised access.
Simplification of the impact factors into a smaller set and integration into recommendation algorithms should then enable us to leverage this knowledge in the planning component and method to reduce effort and increase automation.

The list of properties and metrics will be published on the SCAPE wiki and continuously updated.

For improving automation in the QA and PC components, it will be crucial to look at the cost-benefit relations of automation points to optimise investment of efforts. This will enable informed prioritisation and development decisions. The decision factors will influence (or even drive) the focus on which improvements should be developed in task 3 of the work package 10, Action Services. The evaluation framework used to select the tools, which will be further extended to do a gap analysis and find tool shortcomings that could be a target of improvement, is aligned with the preservation planning decision factors model since it is also based in SQUARE. The analysis of the impact factors on this quality model will help to define which improvements will have a higher return, i.e. which improvement will have a higher impact on the decision of selecting that tool above another, therefore focusing on the most important aspects for the preservation planning user community.

9 Bibliography

Arms, C., & Fleischhauer, C. (2011). *The Digital Formats Website*. Retrieved from The Libary of Congress: http://www.digitalpreservation.gov/formats/

Bauer, S., & Becker, C. (2011). Automated Preservation: The Case of Digital Raw Photographs. *ICADL.*

Becker, C., & Rauber, A. (2011). Decision criteria in digital preservation: What to measure and how. *Journal of the American Society for Information Science and Technology (JASIST) , 62* (6), 1009-1028.

Becker, C., & Rauber, A. (2011). Preservation Decisions: Terms and Conditions apply. Challenges, Misperceptions and Lessons Learned in Preservation Planning. *ACM/IEEE Joint Conference on Digital Libraries (JCDL 2011).* Ottawa: ACM.

Butler, J., Jia, J., & Dyer, J. (1997). Simulation techniques for the sensitivity analysis of multi-criteria decision models. *European Journal of Operational Research* .

CRL and OCLC. (2007). Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC).

Dappert, A. (2010). Deal with conflict, capture the relationship: The case of digital object properties. *Proceedings of the 7th International Conference on Preservation of Digital Objects (iPRES2010).* Vienna.

Dappert, A., & Farquhar, A. (2009). Significance is in the eye of the stakeholder. *Proceedings of the 13th European Conference on Digital Libraries (ECDL).* Korfu, Greece.

Hamm, M., & Becker, C. (2011). Impact Assessment of Decision Criteria in Preservation Planning. *IPRES.* Singapore.

International Standards Organisation. (1999). Information technology – Software product evaluation – Part 1: General overview (ISO/IEC 14598-1:1999).

International Standards Organisation. (2011). ISO/DIS 16363. Space data and information transfer systems - Audit and certification of trustworthy digital repositories.

International Standards Organisation. (2001). Software Engineering – Product Quality – Part 1: Quality Model (ISO/IEC 9126-1).

International Standards Organisation. (2007). Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Measurement reference model and guide (ISO/IEC 25020:2007) .

International Standards Organisation. (2007). Systems and software engineering – Measurement process (ISO/IEC 15939:2007).

International Standards Organisation. (2011). Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models (ISO/IEC 25010).

Kilbride, W. (2010). Preservation planning on a spin cycle. *DPC What's New 28* .

Kulovits, H., Rauber, A., Kugler, A., Brantl, M., Beinert, T., & Schoger, A. (2009, November/December). From TIFF to JPEG 2000? Preservation Planning at the Bavarian State Library Using a Collection of Digitized 16th Century Printings. *D-Lib Magazine* .

Tarrant, D., Hitchcock, S., & Carr, L. (2009). Where the Semantic Web and Web 2.0 meet format risk management: P2 registry. *IPRES.*
Terzis, S. (2009, April). The many faces of trust. *IEEE Computing Now* .
The National Archives of the UK. (2011). *Evaluating your file formats*. Retrieved from http://www.nationalarchives.gov.uk/documents/information-management/evaluating-file-formats.pdf.

---

[i] http://www.ifs.tuwien.ac.at/dp/plato
[ii] http://www.nationalarchives.gov.uk/PRONOM/Default.aspx
[iii] http://www.bsb-muenchen.de/index.php?L=3
[iv] http://www.kb.dk/en/index.html