Impact Assessment of Decision Criteria in Preservation Planning

Markus Hamm
Vienna University of Technology
Vienna, Austria
hamm@ifs.tuwien.ac.at

Christoph Becker
Vienna University of Technology
Vienna, Austria
becker@ifs.tuwien.ac.at

ABSTRACT

Significant progress has been made in clarifying the decision factors to consider when choosing preservation actions and the directives governing their deployment. The Planets preservation planning approach and the tool Plato have received considerable takeup and produce a growing body of knowledge on preservation decisions. However, experience sharing is currently complicated by the inherent lack of semantics in criteria specification and a lack of tool support. Furthermore, the impact of decision criteria and criteria sets on the overall planning decision is often hard to judge, and it is unclear what effect a change in the objective evidence underlying an evaluation would have on the final decisions.

This article presents a quantitative approach and tool to support the systematic assessment of criteria and their impact in preservation planning. We discuss the reconciliation of different quality models and present an analysis tool integrated with the planning tool Plato. We further apply our analysis method to a body of real-world case study material and discuss the results. The outcomes provide directions to optimise and automate decision-making, watch, and policy definitions at large scales, and to lower entry barriers by focussing on those aspects that have the strongest impact.

Categories and Subject Descriptors

H.1 [**Information Systems**]: Models and Principles; K.6.4 Management of computing and Information Systems; H.3 Information Storage and Retrieval H.3.7 Digital Libraries

Keywords

Digital Preservation, Decision Making, Multiple Criteria Decision Analysis, Preservation Planning, Utility Analysis

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 treat a certain part of the content in a large repository. The planning tool Plato[1], 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 [3, 9, 20].

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 is 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 finalized 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 is 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 [2]. 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 project SCAPE[2] is set to move forward the control of digital preservation operations from ad-hoc decision making to proactive, continuous preservation management, through

iPRES2011, Nov. 1–4, 2011, Singapore.

---

[1] `http://www.ifs.tuwien.ac.at/dp/plato`
[2] `http://www.scape-project.eu/`

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 article 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.

The article is structured as follows. Section 2 describes related work in the areas of Multi-Criteria Decision Analysis, Preservation Planning and decision criteria for digital preservation decisions, and software quality models. Section 3 discusses the reconciliation of existing models. Section 4 discusses key issues in decision criteria analysis and impact assessment of criteria, while Section 5 shortly presents a decision factor analysis tool. Section 6 presents some results of applying the presented analysis approach to a growing body of knowledge created in real-world case studies. Finally, Section 7 discusses implications and presents an outlook on future work.

2. RELATED WORK

Preservation Planning is a key element of the OAIS model [12]. 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 specifiy the *'...Content Information and the Information Properties that the repository will preserve'* [15]. 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 [5, 8, 19]. 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 simple illustrative example containing three *decision criteria* and one requirement node ('Correctness') that comprises two of the criteria.

Preservation planning is a typical case of multi-criteria decision analysis [6]. In taking preservation decisions, decision makers 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 [17]. 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.

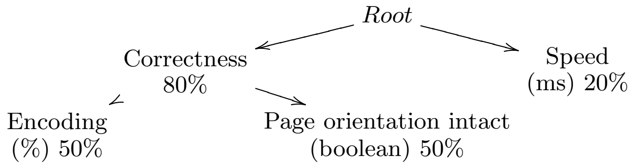


Figure 1: Highly simplified requirements tree

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 [3]. Common approaches to sensitivity analysis vary the weightings of attributes to determine the robustness of assigned weights similar to the approach presented in [4].

The planning approach supported by Plato was also applied to bitstream preservation planning [23]. Recent discussions about preservation planning presented a categorization of decision criteria according to their measurement needs [2] and analysed a series of case studies, focusing on lessons learned and open challenges [3]. 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 [18]. McKinney compared Plato to a commercial implementation that follows a slightly simpler decision model [21].

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' [16] is based on the earlier ISO 9126 family. The ISO/IEC 9126 standards [11] define a hierarchy of high-level quality attributes, where quality measures are based on procedures recommended in ISO 15939 [14]. SQUARE combines a revised quality model with evaluation procedures based on ISO 14598 [10]. It defines requirements on the specification of software product quality criteria [13]. Earlier, Franch proposed a six-step method for defining a hierarchy of quality attributes for a specific domain in a top-down fashion [7]. 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.

Characteristic	**ISO 25010 Definition**
Functional suitability	degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions, comprised of – Functional completeness: degree to which the set of functions covers all the specified tasks and user objectives – Functional correctness: degree to which a product or system provides the correct results with the needed degree of precision – 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, comprised of – Time behaviour: degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements – Resource utilization: degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements – 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
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
Reliability	degree to which a system, product or component performs specified functions under specified conditions for a specified period of time
Maintainability	degree of effectiveness and efficiency with which a product or system can be modified by theintended maintainers
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

Table 1: Software quality attributes as defined in ISO 25010 SQUARE [16]

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. Table 1 defines several characteristics relevant to preservation actions. Section 3 will discuss the relation of these 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 [2], 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. RECONCILING DECISION MODELS

3.1 A generic taxonomy

A first in-depth analysis of about 600 decision criteria of planning studies led to a bottom-up classification of criteria according to their sources of measurement. This was discussed in detail in [2]. 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. 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 [2]. Thus, this article focuses on the top-down reconciliation of top-down models with the overall classification into *action* and *outcome* criteria. In particular, this section discusses format properties, software quality, and information properties.

3.2 Format Properties

The format website run by the Library of Congress (LoC) suggests to evaluate 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 [1].

PRONOM suggests to assess 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 specitic 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 obselescence), 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) [22].

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 2 shows characteristics assembled from LoC and PRONOM (in bold letters) and links them to planning criteria extracted from several case studies. The characteristic 'impact of patents' was generalized into 'rights'. 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 6 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.

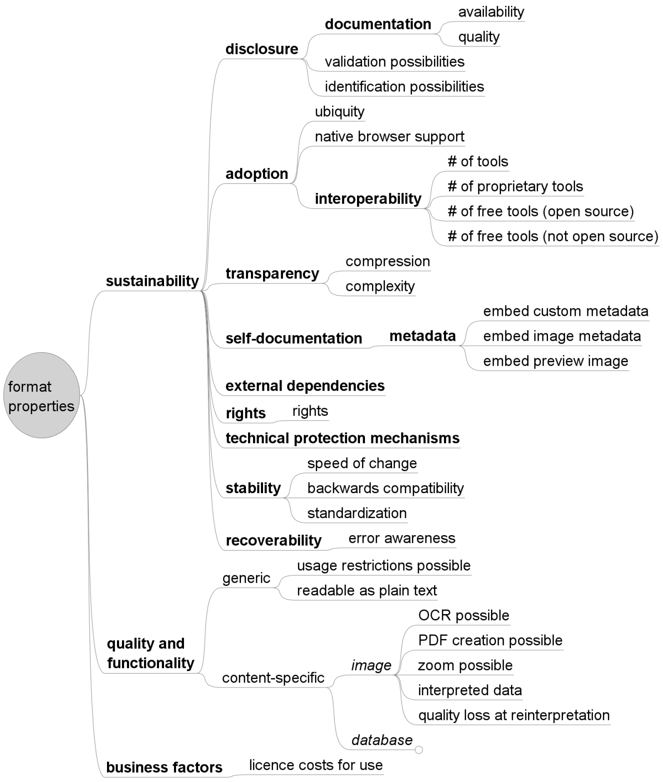


Figure 2: Format factors and associated criteria

On a more general perspective, the name *format* properties may be a bit misleading, since conceptually, this category can include any criterion referring to the *representation* of information in digital form, i.e. its encoding. This observation is particularly relevant in scenarios dealing with the preservation of large data sets instead of traditional 'file-based' objects.

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 previous analysis [2], and assigned them to the SQUARE quality model. Figure 3 illustrates a subset of criteria and their classification according to SQUARE. The ISO quality factors are given in bold.

The ISO quality characteristic *functionality* 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 as a specific category in the planning framework, as discussed below.

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.

3.5 Observations

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 discussed in [2] 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 taxonomy is not very meaningful with respect to criteria semantics. Hence, this section aimed at reconciling standard quality models with 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 in this paper 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 quality model *informs the definition* of such an objective tree. Similarly, the differentiation of the taxonomy described in [2] 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. Thus, the models presented are complementary, and a combination of them is required to model the factors that have to be considered. This unification of models in concrete decision making is achieved within the planning framework.

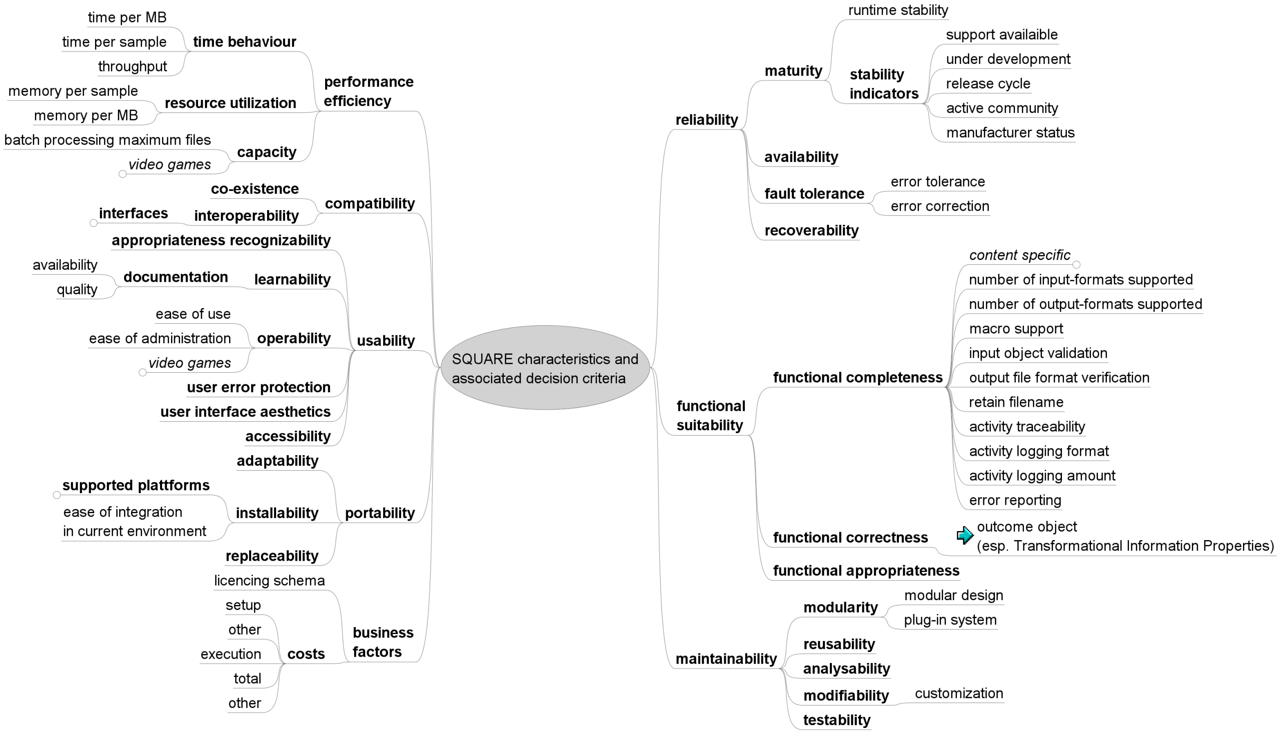


Figure 3: Metrics for SQUARE quality attributes

4. DECISION FACTORS ANALYSIS

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. We thus 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 several key questions:

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 that are *dominated*, i.e. factors that 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?

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 [2]. 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.

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.

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 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.

Table 2 summarizes and names all impact factors, designated *IF*, for single criteria. 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. Let $C = \{c_1, c_2, ...c_n\}$ be the set of criteria and $P = \{p_1, p_2, ...p_m\}$ be the set of plans considered – for example, all plans that refer to the preservation of images in a library setting. Then for a criterion $c \in C$, P_c is the set of plans using c. Thus our first impact factor IF1 represents the size of P_c: $IF1(c, P) = |P_c|$. Let thus *IF1* be the *number of plans using criterion c* and *IF2* the *percentage of plans using criterion c*, i.e. $IF2(c, P) = \frac{|P_c|}{|P|}$. Let further be *IF3* the *average total weight of c* in plans where it is used as given in Equation 1, and *IF4* the sum of total weights divided by the size of the entire set P. IF4 thus includes a discounting for criteria that are rarely used, but with high average total weights.

$$IF3(c, P) = \frac{\sum_{i=1}^{k} w_{c,p_i}}{|P_c|}, p_i \in P_c \qquad (1)$$

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 [2]. 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, i.e. $u_{c,p}(Yes) = 5, u_{c,p}(No) = 1$. If c is assigned a total weight $w_{c,p}$ of 0.25 in the given plan, the *potential output range por(c,p)* 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 $(5 - 1) \times 0.25 = 1$. If $c \notin p$, 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. In addition, $\sum_{i=1}^{k} por(c_i, p) \leq 5.0, c_i \in p$.

However, in fact no value v_c, p in this plan may actually be *No*. Thus, the *actual output range aor(c,p)* of 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 $v_c \in p$, as given informally in Equation 2, with $aor(c, p) \leq por(c, p) \forall c \in C, p \in P$. Similar calculations can be made for numeric criteria, for which thresholds define the utility function.

$$aor(c, p) = w_{c,p} \times (\max(u_{c,p}(v_{c,p})) - \min(u_{c,p}(v_{c,p}))) \qquad (2)$$

Decision criteria often are defined defensively, i.e. potential bad outcomes are considered despite the fact that they are unlikely to happen. 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*) $ror(c, p) = \frac{aor(c,p)}{por(c,p)}$ 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 [3]. We are thus interested in (a) the *rejection potential* of a criterion, i.e. the utility functons with an output range including 0, (b) the *rejection*

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

of a criterion, i.e. the amount of utility functions that reject alternatives due to a utility of 0, and (c) 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 P_c. 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.

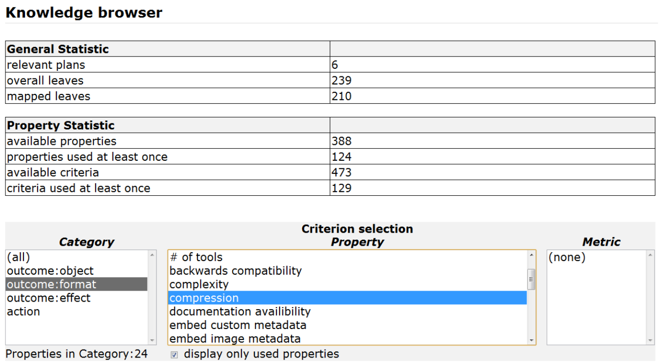
While this set of factors is mathematically simple and robust, it is clearly somewhat redundant. However, the exact factor to be used for answering a certain question has to consider a number of dimensions. To reduce the set of factors that need to be analyzed to answer specific questions and provide guidance on concrete analysis tasks, Section 6 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

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 anonymized ways by aggregating the experience that planners wish to share.

The tool loads preservation plans from the planning tool's knowledge base (provided the plan has been released by the owner and approved by a moderator). It processes and anonymizes 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.


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

Criterion selection

Category
(all)
outcome:object
outcome:format
outcome:effect
action

Property
# of tools
backwards compatibility
complexity
compression
documentation availibility
embed custom metadata
embed image metadata

Metric
(none)

Properties in Category:24

☑ display only used properties

Figure 4: Knowledge browser criteria navigation

Figure 4 shows a screenshot crop of the navigation part, where the user can browse categories and properties of interest. Upon selection of a property or its associated metric, the tool visualizes a number of analysis results. The next section discusses these in detail.

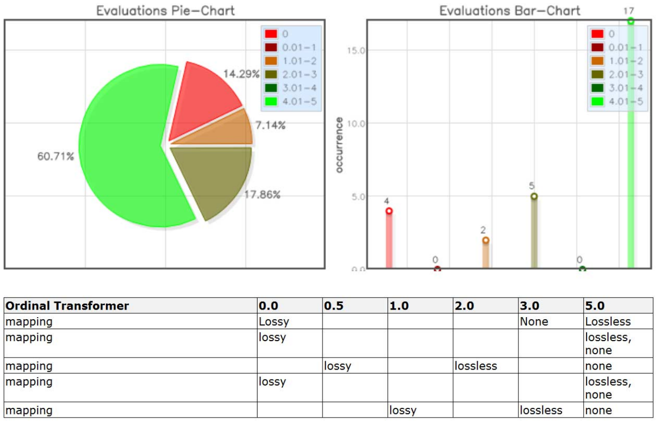
6. ANALYSIS OF RESULTS

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. Our analysis case includes the 6 plans shown in Table 4, which is a subset of the plans outlined in [3], 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.

The tool enables us to browse the criteria categories, select criteria, and analyze their properties and behaviour both in detail and through visualization. Figure 5 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

	Organization type	Planning set	Criteria	Mapped	Alternatives	Chosen action
1	National Library	Large collection of scanned images in TIFF-5 (80TB)	24	24	7	Convert to JPEG 2000
2	National Library	Large collection of scanned images in TIFF-6 (72TB)	43	35	5	Keep status quo, see [20]
3	National Library	Collection of scanned high-resoluton images in TIFF-6	35	29	3	Keep status quo
4	National Library	Small collection of scanned images in GIF	26	25	4	Convert to TIFF-6 (ImageMagick)
5	Professional photographer	Digital camera raw files (CRW,CR2,NEF)	69	67	7	Convert to DNG (lossless) with Adobe DNG Converter
6	Regional archive	Digital camera raw files (NEF)	42	30	5	Convert to TIF (Photoshop CS4)

Table 4: Selected case studies on image preservation.



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 5: Visualization of *Format compression*

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 a frequency distribution along the utility scale. On the bottom, we see the anonymized 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 [3].

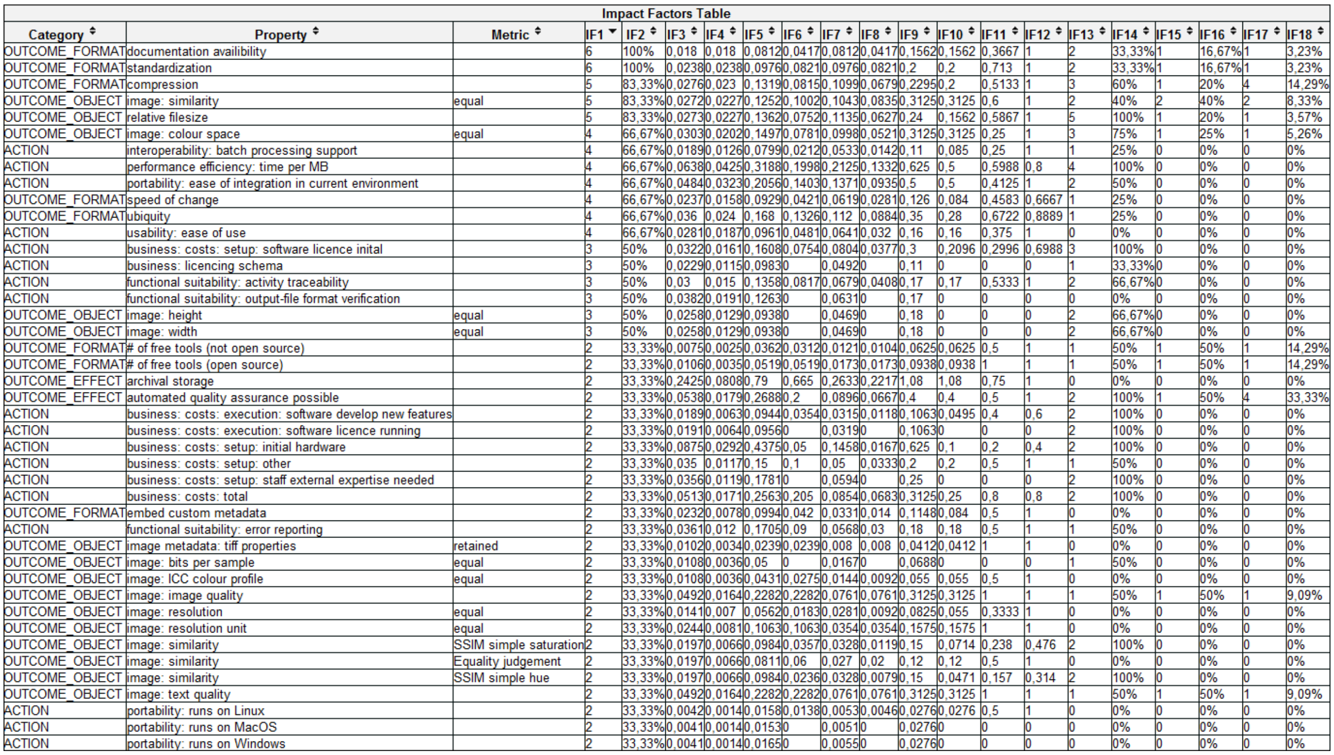
Figure 6 shows a raw view on the most frequently used criteria as displayed in the current version of the analysis tool. Clearly, the meaning of all these numbers is not immediately accessible to a decision maker and will require interpretation by systematic tools, since the question which factors to 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 paper, 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. Figure 7 shows these sets and their impact factors. While space constrains a full analysis and discussion, a number of observations can be drawn.

Format criteria are relevant in all plans, with a coverage of 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. The criteria set contains 17 utility functions with rejection potential. Every second plan in our set is affected by actual rejects caused by these criteria. Performance efficiency, on the other hand, has rejection potential, but none of the tested alternatives was rejected because of performance efficiency drawbacks.

Several aspects of actions are normally included in evaluation, but have very little impact on the decisions (Maintainability, Usability, Portability, Reliability). Business factors, which include costs and licensing, have a much higher relevance. *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



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 6: Impact factors of all frequently ($\geq 33\%$) used criteria, sorted by descending IF1 (count)

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.

7. DISCUSSION AND OUTLOOK

This article 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 (expressed in semantic models) with documented knowledge as well as experience shared by other organizations.

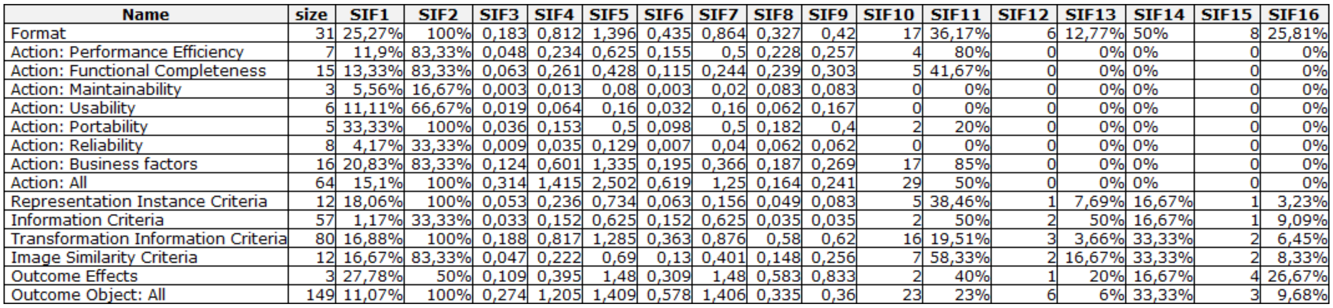
The outcome of this analysis thus provides directions to optimise and automate decision-making, watch, and policy definitions at large scales. It also enables smaller organizations to make decisions at a lower entry barrier by prioritizing essential evaluation factors. For example, criteria that have no rejection potential and a very low potential output range could be left out in preliminary evaluation or substituted with aggregated experience from analogous scenarios. While this cannot in all cases replace a full-depth evaluation, it can provide a tool for trade-off decisions between risks and costs and 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 unambigous specification of decision criteria across scenarios and organizations has additional benefits. In addition to the questions posed in the beginning of this paper, 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 such as those described in [3, 9], completing the assignment of metrics to the SQUARE quality model, 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.

Acknowledgements

Part of this work was supported by the European Union in the 7th Framework Program, IST, through the SCAPE project, Contract 270137.



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	1,396	0,435	0,864	0,327	0,42	17	36,17%	6	12,77%	50%	8	25,81%
Action: Performance Efficiency	7	11,9%	83,33%	0,048	0,234	0,625	0,155	0,5	0,228	0,257	4	80%	0	0%	0%	0	0%
Action: Functional Completeness	15	13,33%	83,33%	0,063	0,261	0,428	0,115	0,244	0,239	0,303	5	41,67%	0	0%	0%	0	0%
Action: 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%
Action: 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%
Action: 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%
Action: 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%
Action: Business factors	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%
Action: All	64	15,1%	100%	0,314	1,415	2,502	0,619	1,25	0,164	0,241	29	50%	0	0%	0%	0	0%
Representation Instance Criteria	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%
Information Criteria	57	1,17%	33,33%	0,033	0,152	0,625	0,152	0,625	0,035	0,035	2	50%	2	50%	16,67%	1	9,09%
Transformation Information Criteria	80	16,88%	100%	0,188	0,817	1,285	0,363	0,876	0,58	0,62	16	19,51%	3	3,66%	33,33%	2	6,45%
Image Similarity Criteria	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%
Outcome Object: All	149	11,07%	100%	0,274	1,205	1,409	0,578	1,406	0,335	0,36	23	23%	6	6%	33,33%	3	9,68%

Figure 7: Criteria sets and their cumulative impact factors as shown in the analysis tool

8. REFERENCES

[1] C. R. Arms and C. Fleischhauer. *The Digital Formats Web site*. The Library of Congress, accessed September 2011. `http://www.digitalpreservation.gov/formats/`.

[2] C. Becker and A. Rauber. 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, June 2011.

[3] C. Becker and A. Rauber. Preservation decisions: Terms and Conditions Apply. Challenges, Misperceptions and Lessons Learned in Preservation Planning. In *Proc. JCDL 2011*, June 2011.

[4] J. Butler, J. Jia, and J. Dyer. Simulation techniques for the sensitivity analysis of multi-criteria decision models. *European Journal of Operational Research*, 103(3):531 – 546, 1997.

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

[6] J. Figueira, S. Greco, and M. Ehrgott. *Multiple criteria decision analysis: state of the art surveys.* Springer Verlag, 2005.

[7] X. Franch and J. Carvallo. Using quality models in software package selection. *IEEE Software*, 20(1):34–41, Jan/Feb 2003.

[8] S. Grace, G. Knight, and L. Montague. *InSPECT Final Report.* InSPECT (Investigating the Significant Properties of Electronic Content over Time), December 2009. `http://www.significantproperties.org.uk/inspect-finalreport.pdf`.

[9] M. Guttenbrunner, C. Becker, and A. Rauber. Keeping the game alive: Evaluating strategies for the preservation of console video games. *The International Journal of Digital Curation*, 5(1), June 2010.

[10] ISO. *Information technology – Software product evaluation – Part 1: General overview (ISO/IEC 14598-1:1999).* International Standards Organization, 1999.

[11] ISO. *Software Engineering – Product Quality – Part 1: Quality Model (ISO/IEC 9126-1).* International Standards Organization, 2001.

[12] ISO. *Open archival information system – Reference model (ISO 14721:2003).* International Standards Organization, 2003.

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

[14] ISO. *Systems and software engineering – Measurement process (ISO/IEC 15939:2007).* International Standards Organisation, 2007.

[15] ISO. *Space data and information transfer systems - Audit and certification of trustworthy digital repositories (ISO/DIS 16363). Standard in development*, 2010.

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

[17] R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: preferences and value tradeoffs.* Cambridge University Press, 1993.

[18] W. Kilbride. Preservation planning on a spin cycle. *DPC What's New*, 28, 2010.

[19] G. Knight and M. Pennock. Data without meaning: Establishing the significant properties of digital research. *International Journal of Digital Curation*, 4(1):159–174, 2009.

[20] H. Kulovits, A. Rauber, M. Brantl, A. Schoger, T. Beinert, and A. Kugler. From TIFF to JPEG2000? Preservation planning at the Bavarian State Library using a collection of digitized 16th century printings. *D-Lib Magazine*, 15(11/12), November/December 2009. `http://dlib.org/dlib/november09/kulovits/11kulovits.html`.

[21] P. McKinney. Preservation planning: A comparison between two implementations. In *7th International Conference on Preservation of Digital Objects (iPRES2010)*, Vienna, Austria, September 19-24 2010.

[22] The National Archives of the UK. Evaluating your file formats. `http://www.nationalarchives.gov.uk/documents/information-management/evaluating-file-formats.pdf`, accessed September 2011.

[23] E. Zierau, U. B. Kejser, and H. Kulovits. Evaluation of bit preservation strategies. In *7th International Conference on Preservation of Digital Objects (iPRES2010)*, Vienna, Austria, September 19-24 2010.