FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Impact Assessment of Decision Criteria in Multi Criteria Decision Analysis for Digital Preservation

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Software Engineering & Internet Computing

eingereicht von

Markus Hamm
Matrikelnummer 0225744

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: ao.Univ.Prof. Dr. Andreas Rauber
Mitwirkung: Dr. Christoph Becker

Wien, 23.09.2011

(Unterschrift Markus Hamm) (Unterschrift Betreuung)

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Impact Assessment of Decision Criteria in Multi Criteria Decision Analysis for Digital Preservation

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Software Engineering & Internet Computing

by

Markus Hamm
Registration Number 0225744

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: ao.Univ.Prof. Dr. Andreas Rauber
Assistance: Dr. Christoph Becker

Vienna, 23.09.2011

(Signature of Author) (Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Markus Hamm
Jahngasse 4/27, 1050 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum) (Unterschrift Markus Hamm)

Abstract

Decision making is a complex issue. Although we all have a lot of practice we are not good at it. Multi Criteria Decision Analysis faces this issue and supports decision makers by specifying decision problems in a formal model which guides the way to the optimal decision. This formal problem specification adds transparency to the whole decision process.

Digital preservation is the active management of digital content over time to sustain its accessibility. Active management means that preservation decision have to be taken. The process of finding optimal preservation decisions for a set of digital objects is called Preservation Planning. Digital content to preserve contains most of the time sensitive information like cultural heritage, business contracts or historical images. Thus preservation decisions are always critical which requires a trustworthy and therefore transparent decision finding process.

Bringing this together Multiple Criteria Decision Analysis can be used to make preservation decisions in a transparent and therefore trustworthy way. The Preservation Planning tool Plato uses this approach to support the responsible planner at making the optimal decision in his scenario. Its decision finding process evaluates several possible preservation actions by a set of given criteria to identify the best action to take.

Several institutions use Plato to take real world preservation decisions. Even though we have a lot of case study data we currently do not know a lot about the criteria used to make these decisions. At the moment we only know which categories of preservation criteria exist and which aspects criteria in those categories refer to. Those categories are action and outcome separated into six subcategories. Criteria in this categories focus to assess the preservation action itself and the outcome of it. The goal of this thesis is to identify the criteria used to take preservation decisions in in detail and assess their quantitative impact on final decisions. This gained knowledge about decision criteria is aimed to help us improving the overall Preservation Planning process.

To achieve this goal we present a method for criteria impact assessment based on analysing empirical collected case study data. Case study data consists of relevant preservation decisions made in the past and their containing criteria. This introduced method consist of five major steps: Prepare case study data for automatic processing, define a set of quantitative impact factors, create tools which support us at criteria impact assessment, analyse and discuss the results.

Kurzfassung

Das treffen von Entscheidungen ist eine komplexe Aufgabe. Obwohl wir alle viel Übung darin haben sind wir nicht sehr gut dabei. Multi Criteria Decision Analysis unterstützt Entscheidungsträger dabei diese Aufgabe zu meistern, indem das Entscheidungsproblem in einem formalen Modell spezifiziert wird, welches die optimale Lösung aufzeigt. Diese formale spezifikation des Problems bringt Transparenz in den Entscheidungsfindunsprozess.

Digital Preservation ist das aktive managen von digitalen Inhalten um sie über längere Zeit zugreifbar zu halten. Aktives managen bedeutet das Konservierungsenscheidungen getroffen werden müssen. Das Verfahren um solche optimalen Entscheidungen für eine Menge digitaler Objekte zu treffen wird Preservation Planning genannt. Digitale Inhalte die konserviert werden müssen beinhalten fast immer sensible Informationen wie Kulturbesitz, Geschäftsverträge oder historische Bilder. Aus diesen Grund sind Konservierungsenscheidungen immer kritisch, was einen vertrauenswürdigen und transparenten Entscheidungsfindungsprozess nötig macht.

Wenn man diese beiden Konzepte zusammenbringt sieht man das Multiple Criteria Decision Analysis dazu benutzt werden kann um Konservierungsentscheidungen transparent und daher vertrauenswürdig zu treffen. Das Preservation Planning Tool Plato trifft Konservierungsenscheidungen auf der Basis von Multiple Criteria Decision Analysis um dem Entscheidungsträger dabei zu helfen die beste Entscheidung in seinem Szenario zu treffen. Der interne Entscheidungsfindungsprozess evaluiert die möglichen Konservierungsaktionen anhand einer Menge von Kriterien um die optimale Aktion herauszufinden.

Diverse Institutionen verwenden Plato um reale Konservierungsentscheidungen zu treffen. Obwohl wir somit viele Anwendungsfalldaten haben, wissen wir noch nicht viel über die Kriterien die zur Entscheidungsfindung dienen. Wir wissen nur welche Kategorien von Entscheidungskriterien existieren und auf welche Aspekte sich diese Kriterien beziehen. Diese Kategorien sind action und outcome mit sechs Unterkategorien. Kriterien in diesen beiden Kategorien beziehen sich darauf die Konvervierungsaktion an sich und deren Ergebnis zu bewerten. Das Ziel dieser Diplomarbeit ist es, die Kriterien die zur Entscheidungsfindung benutzt wurden zu identifizieren und den quantitativen Einfluss, den sie auf die engültige Entscheidungen haben herauszufinden. Dieses erworbene Wissen soll uns helfen den Preservation Planning Prozess zu verbessern.

Um dieses Ziel zu erreichen stellen wir eine Methode vor, die zeigt wie man den Einfluss von Kriterien auf Basis von empirisch gesammelten Anwendungsfalldaten feststellen kann. Unsere Anwendungsfalldaten sind relevante Konservierungsentscheidungen der Vergangenheit und ihre Kriterien. Die vorgestellte Methode besteht aus fünf Hauptschritten: Vorbereiten der Anwendungsfalldaten für die automatische Verarbeitung, definieren von quantitativen Einflussfaktoren,

erstellen von Tools die uns helfen diese Einflussfaktoren zu evaluieren, analysieren und diskutieren der Ergebnisse.

Contents

CHAPTER 1

Introduction

1.1 Motivation

Digital content has several advantages over its analogue counterpart. Some of these are better administration in terms of physical space needs, searchability, interactivity, easy copying, etc. Furthermore the world gets more and more connected and people have more than ever the aim to share information with others. Therefore easy shareable digital content is often preferred. Taking all these current trends into account digital content will become even more important in the years to come. When it comes to provide long term access to content a lot of challenges exist for analogue and digital content. Analogue content has to face challenges like disaster protection, discolouration of book-pages, etc. Digital content also faces its own challenges. Replicating the content to avoid data loss is only one part of solving the long term accessibility problem of digital information.

One problem is indeed the physical preservation of the digital object. Physical preservation means to sustain the bitstream of the object for a long period of time. Duplication is just one aspect of achieving this goal.

Another problem is the logical preservation of the digital object. Just having the bitstream preserved is not enough to recover the original digital object. Without knowing the meaning of the bits the preserved information is useless. The format of the object gives information how to build an information object out of the otherwise meaningless bits. But that's not all. We also have to consider the software which renders the object for a human viewer and the whole environment the software is executed in (operating system, hardware, peripheral devices). All of this influences the final performance of the object.

Taken together, it is not enough to just store digital objects and preserve them at the bitstream level. Continuous preservation decisions have to be taken and need to be executed to preserve the objects also on a logical level. The process of finding these optimal preservation decisions for a set of digital objects is called Preservation Planning.

These Preservation Planning decisions are based on several criteria which depend on the given scenario. At the moment defining these criteria is a completely manual task done by the responsible preservation planner. To ease and automate this task it would be good to know which criteria to focus on or which we may omit. Knowing the impact of criteria on final preservation decisions should lead to a more accurate and effective way of decision making.

1.2 Problem Statement

Preservation planning aims at finding the best preservation decision to take for a set of digital objects. For decision finding Multiple Criteria Decision Analysis can be used. Plato [5] is a decision support tool used in the preservation context to identify the best preservation decision to take based on Multiple Criteria Decision analysis. It enables decision makers to define and evaluate decision criteria relevant in their scenario. After running through a mature process the the outcome is the optimal decision to take presented in an transparent, understandable and documented way. Plato has been used successfully in several real world case studies. Therefore there already exists an approved way to solve this Multiple Criteria Decision Analysis problem and to find the best preservation action to take.

The problem is that the process of defining decision criteria is at the moment a completely manual task and therefore a lot of human effort is involved. Myriads of criteria have to be defined manually although we know that not all are relevant for identifying the best decision. Further all of those criteria have to be provided with a scale and they have to be evaluated and utilised. If it were possible to identify the impact of decision criteria for several scenarios and therefore separate the principal key aspects that drive the decisions from non-critical aspects, it would be possible to systematically improve the amount of automation of this process. This automation could be achieved by targeted initiatives to provide automation for the most critical aspects of evaluation. At the moment it is not intended to fully automate this process of criteria definition, because only the responsible planner knows the special requirements of his scenario. Nevertheless it should be possible to support the user at this process as good as possible to enable him to define his criteria faster and more accurately.

1.3 Methodical Approach

Our approach to identify the criteria with the most impact on the final decision is to analyse significant real world case studies collected over the last years. These case studies were created from several different institutions and span several different content types. Therefore they provide a good evidence for our analysis.

The criteria in these case studies are at the moment only presented in textual form, so dependencies between criteria with the same meaning cannot be determined automatically. For this reason a formalized property model is introduced which allows to assign each criterion an identifiable semantic and therefore leads a way to analyse the case study data an an automated way.

Having identified the semantic of all case study criteria enables us to analyse them in detail and to find out more about their impact. We present several impact factors which cover different aspects of decision criteria and help us to gain insight into criteria usage.

For analysing these impact factors two modules are added to Plato which are introduced later in more detail: Knowledge Browser and Criteria Hierarchy Tool. The Knowledge Browser supports us at analysing single criteria impact in detail. The Criteria Hierarchy Tool allows us to build hierarchies of properties and enables us to analyse these sets of criteria.

1.4 Structure of the Work

This thesis is structured as follows:

- Chapter 2 *Related Work* illustrates related work concerning this topic to establish common ground and prepare the reader for the main sections of the work.
- Chapter 3 *Impact Assessment* explains the approach we use for impact assessment in detail. It shows how the present case study data is prepared for analysis through property assignment and defines the impact factors used for impact analysis.
- Chapter 4 *Tool Support* presents the created impact analysis tools Knowledge Browser and Criteria Hierarchy Tool and explains their functionality and their technical architecture.
- Chapter 5 *Results and Discussion* reveals the impact assessment results and discusses them in detail.
- Finally chapter 6 *Summary and Outlook* summarizes the results of this thesis and presents future research questions emerged during this work.

CHAPTER 2

Related Work

2.1 Digital Preservation

Digital Preservation aims at maintaining digital objects authentically usable and accessible for long time periods. It has been stated by UNESCO as main concern at preserving the heritage for all nations [37]. Analogue objects like books, documents, etc. have been preserved for several hundreds to thousands of years more or less successfully. The main concern of this preservation was and is to keep the information alive. This main concern is also the same when digital objects come in the game, though the challenges to achieve this goal are keenly different. Digital information has to be preserved at three levels, each having it's own challenges:

1. **Physical level:** Physical storage of each bit of information - the objects bit-stream.
   The challenge at this level is first of all to ensure that the bit-stream stays unchanged over time. The change of only one bit can cause the whole information to be lost. This storage of bits without errors is mandatory but keeping this bits accessible is the second challenge. For this it has to be guaranteed that the appropriate hardware to read the stored bits from the storage media is still available and usable.

2. **Logical level:** Interpretation of the bit-stream.
   Preservation at the physical level is not enough to keep a digital object accessible because the stored bits do not give information about the object itself. Without interpretation these bits are useless. This interpretation is done by a software which makes the objects information accessible to the user. This software is embedded in other software and hardware components building a whole environment of cross-linked components. These dependencies make the whole environment fragile to modifications. This means that any changes in this environment can cause the object to be not logically accessible any more. Such an environment can consist for example of a program which is able to interpret and view doc files, an operating system which is able to run this program and finally hardware which is able to run this operating system. Neither of theses elements are stable but all belong

together. Because of this it may be that the doc document has been correctly preserved on physical level, but could not be accessed any more for for the following reasons: No programs exist any more which are able to edit this format, there exist programs but they do not work on the existing operating systems any more, the program and the operating system are okay but do not cooperate with existing hardware any more.

3. **Semantic level:** Semantic of the preserved information.
   Semantics of the preserved information changes over time because the context in which the objects exists also changes over time. Some examples of changing contexts are: Meaning of terms change (city names), interpretation of facts change (alcohol levels etc.). If those terms change, which define the context of the preserved object, the objects information gets another meaning or gets not understandable any more.

To successfully tackle the preservation problem the challenges at each preservation level have to be solved. Physical preservation is the first concern that has to be dealt with. Without accurate physical preservation all other taken actions are useless. If the correct data bit-stream to interpret gets lost, no or wrong results are the outcome in any case. Physical preservation is a well known problem for which a lot of approved solutions exists which are in general based on concepts like redundant and distributed storage, media diversity and periodical media migration. These solutions are already implemented in several systems which are successfully used by many institutions. Because of this we do not focus on this issue in this work.

The level which we focus on in our work is the logical level. It comprises a lot of research questions which are not or not fully solved at the moment. A brief overview about the main conceptual problems and the approaches to face them is presented in the following. Digital objects need a specific environment to be accessible. Because environments evolve over time actions have to be taken to keep objects accessible in these new environments. These actions aimed to preserve the objects are called preservation actions. Several types of preservation actions exist. One of the most frequently mentioned at the moment is called Migration. It's aim is to accept the environment changes and tries to migrate the digital object in a way it can be used in coming environments. By applying migration, the representation of the object changes but the information of it must not change. Another common known action is Emulation. Here the approach is to keep the original environment usable so that no adoption on the original objects has to be undertaken. The original environment is kept alive by emulating it on any new environment that comes up. This means that continually actions have to be taken also if Emulation is used. Instead of performing this action on the object to keep it usable in changing environments, the actions are taken on environment side to keep it alive and usable for rendering the original object. The most intuitive preservation action at first sight is called museum. By just preserving the original environment as it is the museum approach tries to get rid of the modification stress. The basic idea is: If the original environment does not change, the original object either does not have to change. Although this approach looks promising at first glance it does not solve the preservation problem. First of all the technical expertise for running the system has to be sustained over long term to keep the original environment alive. Additional a lot of museum-environment has to be kept because there exist a lot of different environments which have to be all uniquely preserved. Keeping all these different environment alive by buying spare parts which get even more expensive after the lifetime of these parts is an impossible task. So there is a unsuccessful end of this story.

The semantic level threat of objects does not only apply to digital objects. This threat takes place at any kind of long term preservation, digital or analogue. General strategies exists to face this problem like migration on semantic level (e.g. text replacement if a city-name changes), etc. What complicates the solution of this problem is that object contexts and actual contexts are not yet documented and therefore the detection of semantic changes is hard to grasp. This problem indicates no high risk because most of the people reading a text anyhow interpret it regarding the context is was created in. Because of this and the fact that this issue is not specifically related to digital preservation we don't focus on this level in our work.

OAIS Reference Model

OAIS stands for Open Archival Information System. It can be seen as reference model which tries to show how to overcome preservation issues. In detail OASIS defines a model of an archival information system, to be able to understand and apply concepts needed for long term information preservation. Long term means long enough to care about changing technologies and therefore changing environments. It is applicable to any archive, digital or not. As the name states it is a reference model therefore no implementation details are specified. It just cares about concepts and how they relate to each other. OASIS is defined as follows:

> "An archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community... The term Open in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted." [22]

Information is any type of knowledge that can be exchanged and is meant independent of the form (physical or digital).

OAIS was developed by the Consultative Committee for Space Data Systems based on the experiences made with the NASA's first digital archive, the National Space Science Data Center. Over its long time usage period a lot of technological changes occurred which caused a lot of problems. The OAIS model addresses these problems and presents a suggestion of a reference archive which is able to face all these problems. Because of that, this model helps to understand long term preservation in general and in specific long term digital preservation. This reference model was accepted 2003 by the International Organization for Standardization which based the resulting ISO Standard ISO 14721:2003 [27] on this model.

The functional model of an Open Archival Information System is showed in figure 2.1. It consist of 6 functional entities, defined in the OAIS specification [22] as follows:

- **Ingest:** This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive.

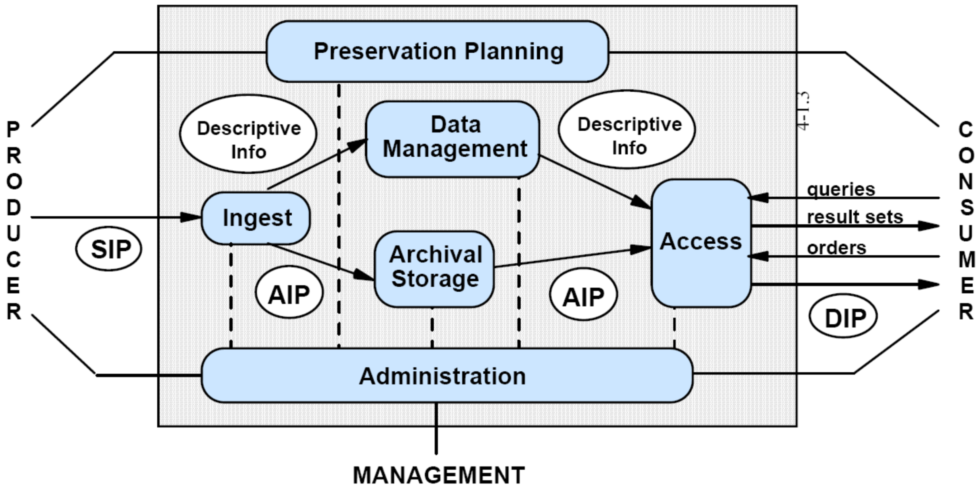


Figure 2.1: OAIS Functional Model [22]

- **Archival Storage:** This entity provides the services and functions for the storage, maintenance and retrieval of AIPs.
- **Data Management:** This entity provides the services and functions for populating, maintaining, and accessing both Descriptive Information which identifies and documents archive holdings and administrative data used to manage the archive.
- **Administration:** This entity provides the services and functions for the overall operation of the archive system.
- **Access:** This entity provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the OAIS, and allowing Consumers to request and receive information products.
- **Preservation Planning:** This entity provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete.

It can be seen that preservation planning is a main component in the functional model. It tries to tackle the logical preservation problems mentioned above. These problems arise from continually evolving environments. As stated above preservation planning should provide recommendations to ensure that the stored information remains accessible to the designated user community over long term. To be able to provide reasonable recommendations all possible options must be identified. In digital preservation these options are preservation actions to take. Knowing these possible actions to take it has to be checked how these actions align with customer and scenario specific requirements. To figure out the best preservation action to take in those specific cases some kind of decision aiding is required. This decision aiding process has to be transparent and reproducible to be trustworthy. Trustworthiness is of particular importance in preservation planning because the decisions to take are always critical. Making the wrong decision can result in complete and irreversible data loss. Multiple Criteria Decision Analysis (MCDA) or Multiple Criteria Decision Making (MCDM) is a decision aiding process which helps finding the optimal decision and is able to fulfil the trustworthy requirements stated above. It is introduced in the following chapter.

2.2 Multiple Criteria Decision Analysis

Multiple Criteria Decision Analysis (MCDA) or Multiple Criteria Decision Making (MCDM) [29] [24] [21] supports decision makers at taking the right decision in their specific scenario. This is done by analysing, structuring and specifying the decision problem in a formal way the optimal decision can be deducted from. This formalization is also intended to reveal different aspects of the decision problem which is especially useful and sometimes even necessary in very complex decision cases. A formalized way of making decision is reproducible and therefore adds transparency to the whole decision process.

When we take a look on how people make decision we can see that MCDA fits well into this process. When people have to make critical decisions they have to take into account different points of view to evaluate and compare each possible decision. What is often intuitively created because of this is a pro and contra list (decision A is cheaper but it has not such a good performance as B, etc.). This different points of view are taken into account also in MCDA and are called criteria. They are used to evaluate and compare the different aspects of each possible decision leading to the suggested decision. In contrast to this using only a single criterion (mono criterion) to evaluate different possible actions is not intuitive. Decision makers never have only a single criterion in mind for a decision problem. Every consideration has several aspects like resources, performance, security, quality, etc. Bringing this together to one criterion without formal identification of each aspect leading to this criterion brings ambiguity in the decision process instead of transparency. Maybe the final decision is the same (because all the different aspects were intrinsic taken into account) but transparency gets lost and therefore the justification why the decision was taken. MCDA lists all criteria which contribute to the final decision and therefore is a transparent and trustworthy way of making decisions.

The process of deducting a final decision from a set of well considered criteria may seem like one hundred percent objective, but this impression is not fully correct. The point is that there is still a decision maker involved in the whole process and no human being can be one hundred percent objective. Decision makers often have a inner preference for any of the options (known or not known) and therefore intuitive seek a confirmation for this. This affects the defined criteria and therefore maybe distorts objective decision finding. But even if the decision is not made one hundred percent objective MCDA is anyway a good choice because it transparently documents why the decision is taken that way. Because of this by observing the MCDA decision model you can find out the decision foundation and see if the decision was made in an objective way or not.

Summing this up decision making is a complex issue. As a prolific writer on decision making wrote, “In an uncertain world the responsible decision maker must balance judgements about uncertainties with his or her preferences for possible consequences or outcomes. It's not easy to do and even though we all have a lot or practice, we are not very good at it. [29] Multi Criteria Decision Analysis (MCDA) or Multiple Criteria Decision Making (MCDM) faces this issue and supports decision makers by specifying the decision problem in a formal model which guides the way to the optimal decision. The formal specification of the problem adds transparency to the whole process. This formal specification can be quickly explained by defining a number of possible decisions and a number of criteria by which these possible decisions are ranked. This process is executed by the responsible decision maker. The interesting part of this is the way to get to this ranking which leads to the optimal decision and therefore the decision to take. This different ways of decision making are explained in the following.

Decision making in Multiple Criteria Decision Analysis

Step 1: Define requirements and alternatives

The first step which has to be done by the decision maker before MCDA can start is to specify the decision problem for which a solution is sought in a formal way. The problem is specified by defining all possible decisions and all criteria which will be used to rank these possible actions. Each criteria is associated with a weight to identify it's user-stated impact on the decision ranking. A weight can have values between 0 and 1 and the value of all criteria weights must sum up to 1.

Step 2: Evaluate criteria

Once the weighted criteria and the possible decisions are defined evaluation of these criteria for each possible decision can begin. Criteria evaluation can be done manually by the responsible decision maker or automatically by some kind of services, etc. The evaluation outcomes are in any case criteria values for all possible decision.

Step 3: Preference modelling

Now all criteria values are evaluated, the problem is that this evaluated criteria values does not give any indication of its contextual value often stated as preference, utility, worth or simply value. We will use the term preference value in the following. Take the criterion process runtime for example, and assume that its evaluated value is 10 seconds. Although we have evaluated the criterion value we cannot indicate if this is good or bad. For a complex data warehouse query 10 seconds are extremely fast, for a user interaction delay 10 seconds are extremely slow. To be able to make a declaration about the preference of an evaluated criterion a transformation from the neutral evaluation value to a concrete preference value is necessary. This transformation is usually done by a preference function the decision maker has to define. Only he can do this, because only he knows what preference a evaluated value has in his given context. This preference functions are also known in literature as utility functions, value functions or worth functions dependent on the associated risk. Under certainty usually the term value function is used. Under uncertainty usually utility function is used. For reasons of simplicity we only use the term preference function in the following.

Another reason why it is necessary for assessing preference of evaluated criteria values is that these criteria may all have different scales and units. For example runtime is evaluated as number measured in seconds, usability is evaluated as string(good/medium/poor) and some comparisons like image similarity are evaluated as boolean. Because of this it is not possible to compare these values directly and therefore it is also impossible to aggregate them in a way which leads to the final decision.

To sum this up the preference value of an criteria must give us information about how good this criteria is fulfilled and additionally must be comparable and easy summable. The last two points are mathematically and necessary to be able to provide an aggregated ranking of the possible actions. Because of this the preference function usually transforms the evaluated value to a numeric scale with a range from 0 to 5. 0 means criteria not fulfilled, 5 means criteria fulfilled to hundred percent, any other value is something in between.

Step 4: Decision making

Now all criteria are evaluated and preference modelled, examining the best decision can start. But what is the best decision? The best decision is the one with the best criteria preference over all criteria. But how can we find out this overall criteria preference? Some criteria might have its optimum at action a, other criteria at action b and others at action c. Taking this locale optimums into account does not help a lot because the global optimum is searched for. So a method is required which aggregates all criteria and is able to state which possible decision is overall the best. Two main methods exist.

One of it is the **Multiattribute Preference Method**. The main concept is based on the assumption that a so called multiattribute preference function exists which is able to calculate an aggregated preference value over all single criteria preference values. The result of it is one numerical preference value for each possible decision. This numerical value represents the strength of each possible decision. The more preferable the possible decision the larger its numerical value. This means that all criteria preferences are aggregated to one single number for each possible decision indicating its value. Having this it is easy to rank the actions and find out the best one.

Such a multiattribute preference function can for example be based on the Weighted Sum Model (WSM) which is shortly introduced in the following. A precondition for the WSM is that all preference values are measured on the same scale (e.g. all 0-5), otherwise wrong results are produced.

The *Weighted Sum Model (WSM)* is simply based on the concept that for each possible decision each belonging criterion preference is summed up considering its weight. This leads to a final sum for each possible decision. Weighting is considered by multiplying the appropriate weight with its criterion preference value. So the quantitative result for each possible decision j is calculated as follows if we assume that we have n criteria. w is the criterion weight, c is the criterion preference.

$$r_j = \sum_{i=1}^{n} w_i * c_{ij}$$

The decision with the highest result is suggested as the best one.

The second popular method is the **Outranking Method**. These method does not calculate an overall preference value for each possible decision which can be used for ranking. Instead of this all possible decisions are successively compared to each of the other. This means that the final ranking is not achieved by creating a complete preorder on the possible decisions, instead of this it is addressed by pairwise comparisons.

The *Weighted Produce Model (WPM)* is an example method for this and uses comparison of multiplication results to rank the possible decision. The comparison base is calculated by multiplying criterion rations raised to the power of the criterion weight. The assumptions of weights are the same as stated before. The results for each decision are calculated as follows if we take the same assumptions as above. D stands for possible decision, x and y stands for any of the j possible decisions to compare.

$$r\left(\frac{d_x}{d_y}\right) = \prod_{i=1}^{n} \left(\frac{c_{ix}}{c_{iy}}\right)^{w_i}$$

$r(a_x/a_y) \geq 1$ indicates that decision x is more desirable than decision y. The decision which is better than or at least equal to all other decisions is suggested as the best one and therefore the final decision.

Sensitivity analysis

As stated above there already exist well known solutions to solve the MCDA problem in a transparent way. These solutions result in the best decision to take in a specific scenario. What we do not know at the moment is which impact each decision criterion has on the final decision. This is of interest because if we know the impact of decision criteria in certain scenarios we can optimise and automate decision making in general by identifying and prioritizing decision drivers, support responsible planners at defining their criteria by suggesting pre-assembled requirement trees or tell them what criteria they missed or should remove. Specific for the preservation context this could mean for example to know what criteria to focus on when implementing new automated measurements.

To figure out the impact of decision criteria on the final decision some kind of analysis is needed. Typically sensitivity analysis is used for this task. "Sensitivity analysis (SA) is the study of how the variation (uncertainty) in the output of a mathematical model can be attributed to different variations in the inputs of the model." [33] In general we can distinguish two types of sensitivity analysis: Local and global. Local sensitivity analysis focuses at varying only one input parameter at a time (this is the reason for the name local). The variance of the output while varying this input factor leads to the impact of this one input factor. At applying global sensitivity analysis all factors are variegated at the same time (this is the reason for the name global) depending on an user associated probability distribution. After a set of experiments the sensitivity measures can be calculated. To figure out the local sensitivity of a input parameter is quite simple. The input parameter of the model has to be varied over its range, then the variance of the output states the input parameters local sensitivity. To assess the global sensitivity is more complex because probabilities have to be considered. The general assessment procedure is as follows. First of all probability density functions have to be assigned to each input parameter of the model regarding its uncertainty. Based on this probability distribution concrete values for each input parameter, called sample values are calculated. With these couple of sample parameter the model is evaluated several times. Based on this evaluations a lot of strategies exist to derive the sensitivity measures of the input parameter. Some of these strategies are listed and explained in detail in [34].

These sensitivity analysis methods can be and are already applied in Multiple Criteria Decision Analysis. Evangelos Triantaphyllou for example uses sensitivity analysis in [36] for two purposes. The first is to to find out the criticality of criteria, by identifying how sensitive the actual ranking of the alternatives is to changes on criteria weights. For this a sensitivity analysis on the criteria weights is done. The criterion with the smallest relative weight change necessary to change the alternative ranking is the most critical. The second application of sensitivity analysis is based on a similar approach not on the criteria weights but on all criteria preference values. Therefore sensitivity analysis on criteria preference values is performed. The criterion preference value with the smallest change necessary to change the alternatives ranking is stated the most critical one. In [38] sensitivity is defined also dependable on action ranking changes. This study also concentrates on the sensitivity of the action ranking regarding to input parameter changes. The input parameter in these studies are the following: Specific changes in the evaluations of an alternative, specific changes in certain criterion-scores of an alternative, criteria weights. [18] also uses sensitivity analysis on criteria weights to identify criteria importance. Finally [20] uses sensitivity analysis on criteria weights not to identify criteria impact but to give assistance at assessing optimal criteria weights. What can be seen is that sensitivity analysis is already widely used in MCDA, most of the time performing analysis on criteria weights depending on decision ranking changes. When it comes to identify criteria impact most of the time this weight analysis results are consulted.

2.3 Preservation Planning

Because of diverse threats digital content is faced (see chapter 2.1) we continually have to take actions to keep digital objects accessible in changing environments. To figure out the best preservation action to take in a specific scenario preservation planning is used. This mandate of preservation planning is similar to the one of component selection. Component Selection is stated as the problem to find the most appropriate component in a given scenario. The only difference is the goal. Preservation planning aims to figure out the best action to take instead of the best component to select but the general underlying concepts are the same. These similarities are stated in [14].

To find out the best action to take a method is required to identify it. Preservation decisions are always critical, therefore this method must be trustworthy and reproducible. Multiple Criteria Decision Analysis (see previous chapter 2.2) is a method which can be applied to this problem successfully. In the following preservation planning and how Multiple Criteria Decision Analysis is applied in this context, is explained in more detail.

Preservation Planning and the OAIS Reference Model

The main idea of preservation planning evolves from the OAIS reference model [22]. Its role in the OAIS reference model is stated as "This entity provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete." [22]. So it tries to tackle the logical preservation problems resulting from evolving environments and gives suggestions how to deal with this problems. Its concrete functions can be seen in figure 2.2. Technology and designated community have to be monitored to identify risks which need actions to be taken. Risks can be for example formats which get obsolete, hardware or software which gets obsolete, new applied rights to formats, new format choices on consumer side, etc. Develop Preservation Strategies and Standards is the heart of preservation planning and the component where decisions are made. It aims for developing and recommending strategies how the monitored risks can be faced. Develop Packaging Designs & Migration Plans finally applies the suggested strategies from Develop Preservation Strategies and Standards.

From OAIS to concrete Preservation Planning

As we saw the main part of the OAIS Preservation Planning component is the Develop Preservation Strategies and Standards function. As mentioned above OAIS is just a reference model and therefore if we want to get concrete results we have to take this model and implement it. Plato [5] [13] is a preservation planning tool having good reputation in the digital preservation community which implements this Preservation Planning component. Its approach for doing this is showed in figure 2.3 and is explained briefly in the following.

The first step on the way to a preservation plan is to define requirements, the possible preservation actions should met. To be able to define these requirements the responsible planner has has to be clear about what kind of objects should be preserved and what are the criteria and its importance to measure the quality of preservation. The next step is to define the possible preservation actions that can be taken in this specific scenario. Afterwards the possible actions are evaluated and analysed. This whole process leads to a recommendation for a specific preservation option. This mentioned steps implement the *Develop Preservation Strategies and Standards* function of the OAIS model. Because all criteria, all preservation-actions, all evaluation and the decision analysis is documented this is a fully transparent process of decision finding which leads to trustworthy preservation planning. Based on the outcoming recommendation a preservation plan can be built. This relates to the OASIS function *Develop Packaging Designs & Migration Plans*. The monitor functions mentioned in the OAIS model *Monitor Technology* and *Monitor Designated Community* are aggregated here in the monitor activity which monitors all events which may have an effect on the preservation plan. They are changing requirements, changing technology and changing environment. Any mentioned change leads to a restart of the whole workflow recognizing the changed parameters.

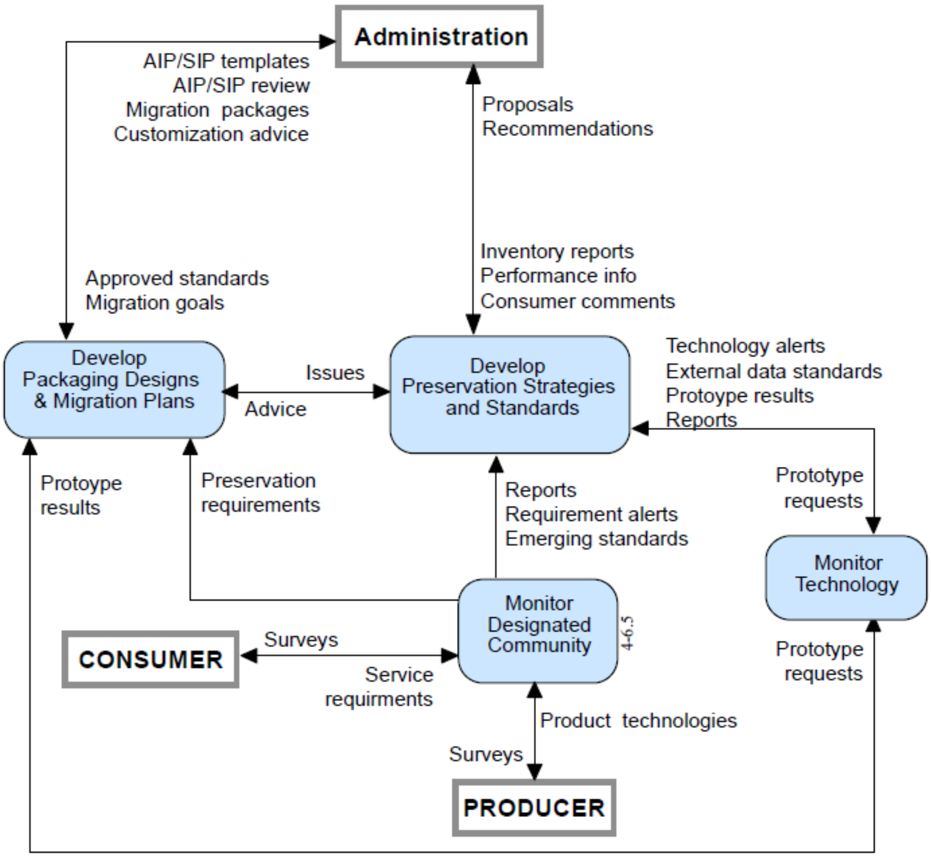


Figure 2.2: OAIS Preservation Planning Functions [22]

Plato Preservation Planning Workflow

To gain more insight in the preservation planning workflow of Plato the concrete actions of the workflow items shown in figure 2.3 are explained now in more detail. In step *Define Requirements* the responsible planner has to provide first or all some basis information about the preservation plan to create. This includes the object types which are aimed to preserve, some contextual information about the plan, planning purpose, etc. After this sample objects, which are representative for the object types to preserve, have to be uploaded and described. These sample objects are used to evaluate if preservation actions met the defined criteria. Now the main part of requirement definition starts by defining the criteria. Criteria are organised in an objective tree which means that they can be organized in an hierarchical order. Therefore it is possible to add nodes in the objective tree for structuring purposes only which are no criteria itself. For each criteria in the objective tree the responsible planner has to provide a evaluation scale. Providing scale restriction and unit for the evaluation is optional. A scale restriction can be useful for example for a criterion named school grade. School grade has an integer scale but

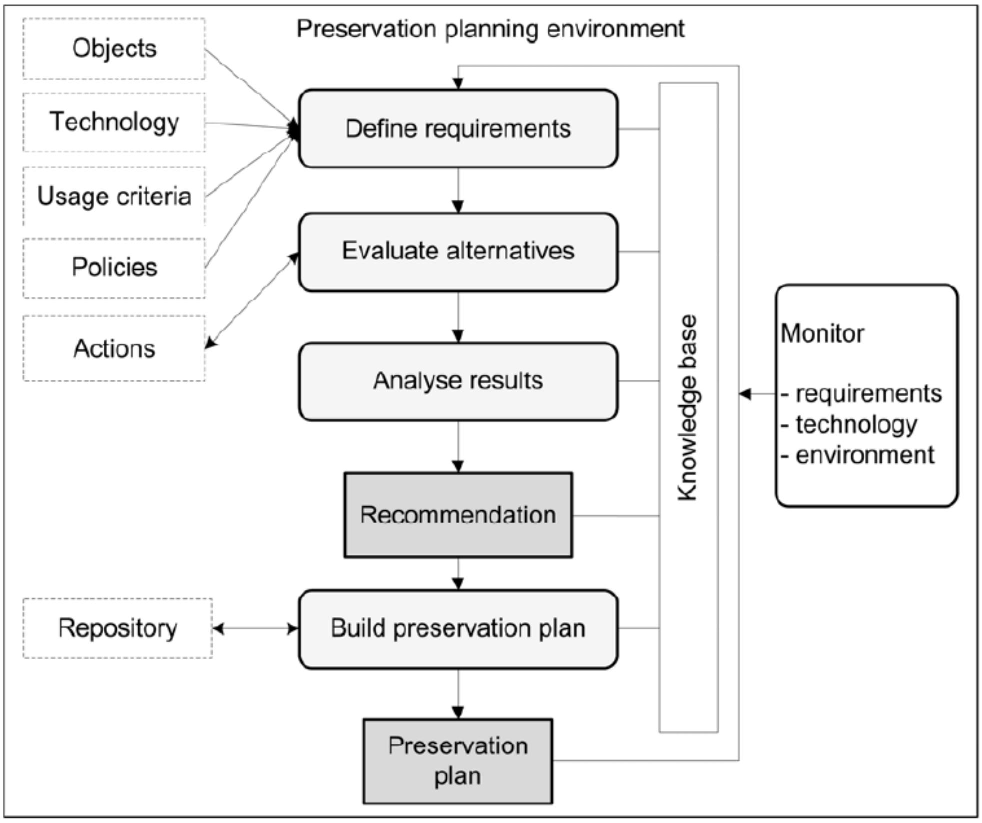


Figure 2.3: Plato Preservation Planning Approach [13]

can have only values 1-5 in Austria. To add semantics to the defined but actually only text based criteria predefined object properties properties can be assigned to criteria. These properties can be uniquely identified and contain a human understandable description. This means that adding such a property to a criteria enriches it with semantic meaning. More details about criteria and properties in digital preservation is given in the following chapter 2.3.

Once the requirements are defined they can be used to evaluate possible alternatives. This step is called *Evaluate Alternatives*. To be able to get started the responsible planner first of all has to define the preservation actions he considers in his scenario. After the alternatives are clear evaluation can start. Some criteria can easily be assessed by the responsible planner like, is the preservation action tool an open source product, etc. Anyhow most of the criteria to evaluate cannot be assessed directly by the responsible planner because they are based on the preservation action runtime behaviour or its outcome. This means the preservation action has to be executed to assess most of the criteria. This is done in the form of experiments. An experiments can be briefly be described by taking a sample object, performing a preservation action on it and measure the results of it. In detail this process is much more complicated than it sounds because the preservation actions itself usually needs a special environment to be runnable whose setup is not always straightforward. These experiments can be performed manually be the preservation planner or automatically by Plato. The manual way has the drawback that the preservation planner has to care about the preservation action installation and the correct environment setup. He also has to perform the experiments for each sample object and care about how to measure his criteria. Measuring the runtime of some kind of object migration for each sample object for example is not an easy task. To sum this up performing experiments manually is most of the time a complicated and very time consuming task. Because of this Plato offers the possibility to perform these experiments and the subsequently measurements automatically. All currently defined properties in Plato can be measured automatically. There also exist a lot of preservation actions which Plato is able to run in a controlled experimentation environment. Summarizing this means that if the responsible planner chooses an action which can be automatically evaluated by Plato, then all criteria having assigned a property are evaluated automatically.

Next we have to analyse the results of the before evaluated criteria. This is done in the step *Analyse Results*. Before analysing of the evaluated values can start their contextual meaning has to be assessed. As described in chapter 2.2 this is done by assigning a preference function to each criteria which is responsible for converting the evaluated criteria values into preference values. Preference values give information about the fulfilment of the criteria. Preference values in Plato are numbers between 0 and 5. 0 means criteria not fulfilled, 5 means criteria fulfilled to hundred percent, any other value is something in between. The final preference value for a alternative is calculated as described in the following. As described in the previous step *Evaluate Alternatives* all concrete values are assigned to or measured for the criterion. Each of these values are transformed by the preference function to preference values. Theses preference values are then aggregated to the final criterion evaluation for this alternative. Two aggregation types exist: Arithmetic Mean and Worst Result. After identifying the preference values the responsible planner has to set the importance factors for each criteria. This importance factors are relative weights which must sum up to 1 at each hierarchy level. Once all preference values and the weight of each criterion are known all data required for decision making is in place and calculation of the best preservation action can start. The best preservation action is calculated by aggregating all criteria preferences to one aggregated preference value for each preservation action. Preservation actions then can be compared and ranked by this numerical value. This method of evaluating the best decision to take is called Multiattribute Preference Method. In Plato two Multiattribute Preference Methods are used to derive the final decision. The first method uses multiplication to aggregate the preference values. The alternative preference values are raised to their weight and then multiplicated with each other. Because multiplication is used as aggregation method, all alternatives which contain at least one preference value 0 in any criteria result to 0. This method is used to eliminate alternatives with unacceptable performances. All alternatives which were not eliminated the step before are then ranked by the second applied multiattribute preference method called weighted sum model. In this method the alternatives criteria preference values are mutiplied with their weight and the summed up. This sum method results in a final ranking of the alternatives on a rational scale. The best action to take is the one with the highest score resulting from this method. Further details about this method are again given in chapter 2.2. As can be seen, this analysing process results in a recommendation which preservation action is the best to take. The analysis results are presented in a report in a

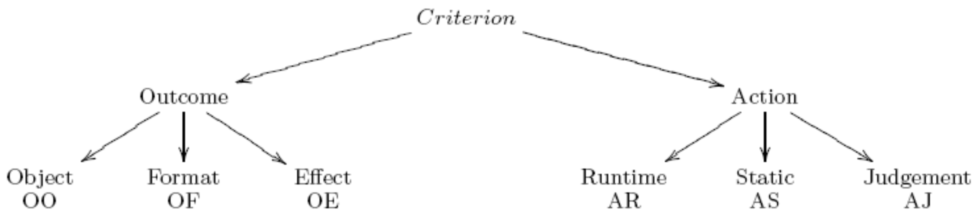


Figure 2.4: Taxonomy of criteria in digital preservation [15]

transparent way so that it is clear why this preservation actions is the recommended one.

Based on this decision for a preservation action a preservation plan can be built. The related step in the Plato workflow is called *Build Preservation Plan*. A preservation plan can be defined as:

> "A preservation plan defines a series of preservation actions to be taken by a responsible institution due to an identified risk for a given set of digital objects or records (called collection). The Preservation Plan takes into account the preservation policies, legal obligations, organizational and technical constraints, user requirements and preservation goals and describes the preservation context, the evaluated preservation strategies and the resulting decision for one strategy, including the reasoning for the decision. It also specifies a series of steps or actions (called preservation action plan) along with responsibilities and rules and conditions for execution on the collection. Provided that the actions and their deployment as well as the technical environment allow it, this action plan is an executable workflow definition." [12]

Summing this up a preservation plan describes what needs to be preserved and describes the series of concrete steps which need to be taken for this. A reasoning why this steps are the right choice is given to provide some kind of traceability and transparency. It additionally states when this defined steps have to be executed (what are the triggering events, etc.) and what are the prerequisites to be able execute this steps (system requirements, etc.). Additionally to these technical concerns the preservation plan also defines organisational responsibilities, rules and conditions for executing the preservation action.

Criteria in Digital Preservation

As the previous chapters stated preservation planning evaluates potential preservation actions objectively against scenario specific criteria. This suitability evaluation of each potential action leads transparently to the best action to take. But what are the criteria we evaluate our actions against? Do they have things in common? Can we classify them? What do we know about them? As stated in [15] decision criteria in digital preservation can be classified roughly in 2 groups: Criteria which refer to the preservation action itself (action-criteria) and criteria which refer to the outcome of the preservation action (outcome-criteria). These criteria are suggested to be further sub-classified as follows. Action can be further classified into Action Static (action properties which are static in nature), Action Runtime (action properties which can be measured at runtime) and Action Judgement (action properties that depend on judgement). Outcome can be further classified into Outcome Effect (general effects of the outcome), Outcome Format (format of the outcome), Outcome Object (significant properties of the outcome object). This taxonomy is showed in figure 2.4.

There still exist other interesting taxonomies which fit in this context and therefore can be used for guidance. For action properties we can use the ISO 25010 system and software quality model standard [28] as reference. It is based on the earlier ISO 9126 standard [26]. ISO 25010 provides guidance for quality models and defines a hierarchy of high level quality attributes. These quality attributes build a complete and extensible framework which is intended to be assessed by custom measurements. An approach how this customization to specific needs can be done is stated in [23]. These high level quality attributes are structured in characteristics and sub-characteristics. To give a brief overview the 8 characteristics are: functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability and portability. Because preservation actions are primary software actions this model fits quite well for assessing preservation actions. It does not help us defining each individual property in our context but instead of this giving us a solid base taxonomy for preservation actions which we can enrich with our own properties.

For outcome properties also exist some reference models and guidelines which help us finding the track to our final properties and the model around it. Outcome formats are an important part of preservation because they define in which way intellectual information is stored. If information about this storage format is lost the containing intellectual information cannot be extracted any more. Therefore sustainability is an important long term preservation attribute of file formats. On the other hand the storage format also defines in which quality the information can be stored and the functionality or interactions possible with this information. For assessing the outcome format several references exist. Most of them focus on the just mentioned factors. The Library of Congress provides a set of high level evaluation factors of digital formats on their website [30]. The main factors are sustainability which is further divided in 7 sub-factors, and quality and functionality. These factors are not directly measurable and require again assignment of more specific properties to be measurable. The National Archives also provides a method to assess the continuity properties of file formats [9]. This is done by assessing the file format against the following characteristics: Capability, quality, resilience/safety, flexibility. These characteristics are divided in sub-characteristics. The characteristics are not intended to be complete and need customization dependent on the given scenario. Also the Risk analysis service of the Preserv2 registry [6], evaluates the usage risk of a given format against the following characteristics: Capability, quality, resilience/safety, flexibility. Once again each characteristic is divided in sub-characteristics and needs customized measurable properties to be assessed.

How we derived our final property model from all these reference models is explained in detail in the main section of this work.

2.4 Observations and Outlook

As stated in this whole chapter, making digital content accessible and usable over a long period of time is a challenging task. These challenges arise because digital content is subjected to threats at several levels which can cause irreversible data loss if they become reality. Tackling the threats of the logical level is of special interest these days because a lot of vital but unsolved challenges exits in this area. The problems in this area mainly evolve from continually changing environments which may cause objects to be not accessible or usable any more. Preservation planning is an approach for solving these logical preservation problems by figuring out the right actions to take to keep the objects accessible over time. For preservation planning there yet exist an established tool called Plato which supports the responsible curator at taking these decisions in a transparent way. Because of the already high acceptance and usage of Plato a lot of relevant case study data exist which could be used as knowledge base for improving the preservation planning process. These case studies were originated from several institutions like libraries, archives, museums, libraries, research labs, etc. and span several content types like documents, images, games, etc. Because of the diversity of the data it is very representative. An elaborate view on two of them is presented in [11, 17].

But what are the areas of the preservation process which need improvement? First of all criteria definition for a specific scenario is at the moment a very hard and time consuming job because it has to be done one hundred percent manual. Being able to use the collected case study data to identify the most important decision criteria for specific scenarios would help a lot. This would allow to support the responsible curator at criteria definition. These support could be for example giving hints about what criteria to use or not to use in the current scenario. Knowing the most important decision criteria could be also useful to prioritize the implementation of automated criteria measurements. Beside of identifying the most important criteria, the collected case study data could also be used to do aggregation of criteria or criteria groups. Aggregation could be used on criteria and criteria groups for example to identify common weights, common scales, common utility functions, etc. These results could be again used to give the responsible planner suggestions how to define criteria weights, scales, etc. Although the analysing approach is first of all intended to be applied to the before mentioned representative case studies this approach is generic and can be repeated any time when new data is available to get new insights or just more accurate results.

As stated above there still exists a lot of relevant data in Plato. Analysing of this data would be very helpful to improve the preservation planning process. The problem is that the data is present but the semantics of it is not known. In practice this means that the criteria of the relevant case studies are at the moment text only with no ascertain meaning. Because of this no coherences can be assessed for aggregation and therefore no meaningful analysis can be performed.

To give criteria semantic, properties with a defined meaning have to be assigned to them. These properties already exist in Plato but are not optimal formalized at the moment. This means that these properties are not complete and therefore only a few criteria can be semantically enriched. For analysing the present data in a reasonable way all criteria have to be mapped to properties. Additionally the few existing properties are not structured good enough which makes finding of the fitting property for a specific criterion a hard job.

To establish a more formalized property model we need reference models which prescribe useful structures to classify our properties. Plato already uses a taxonomy to classify properties. This taxonomy is based on a main classification of properties in action and outcome and underlying sub classes action:static, action:runtime, action:judgement and outcome:object,outcome:format, outcome:effect. The main classification of criteria in action and outcome makes sense and covers all possible Preservation Planning criteria. The sub-classification of these main classes is sound as well but is maybe too strict and therefore possibly will be subject of change. There still exist another interesting property models which fit into this context which have to be taken into account. For example the ISO 25010 standard [28] is interesting for action evaluations, The Library of Congress [30] and The National Archives [9] format evaluation factors are intersting for outcome:format evaluations, etc. So there already exist some interesting formalization models for digital preservation properties. The problem is that they are not homogenized at the moment. Homogenize all these property models is part of the main work.

Summing this up there are steps required to be able to start further analysis on the present preservation planning data of Plato. These steps are: Homogenize property models, define properties of the homogenized model as complete as possible, map properties to criteria to give them a semantic meaning. After this first steps are done the semantic of the present data is known. Therefore now all kind of impact or aggregation analysis can be performed.

CHAPTER 3

Impact Assessment

This chapter introduces a procedure to assess the impact of decision criteria in digital preservation. This procedure is based on an empirical analysis of collected case study data. In detail this data consists of relevant preservation plans collected in the Preservation Planning Tool Plato over the last years. Relevant plans were sorted out thoroughly from all numerous available plans in Plato to build an excellent basis of plans which is able to deliver meaningful results. Plans created by trusted institutions and other reliable sources were specifically selected. In addition plans were selected to cover a wide range of content types in order to achieve best results for diverse scenarios. The following chapters explain the procedure of impact assessment on the relevant plans in more detail.

3.1 Introduction

In our approach impact assessment of decision criteria is based on analysing significant case studies collected over the last years. There already exists research executed on part of this data which basically focuses on the assessment of criteria usage frequency. These usage frequency can be seen as very basic and inexact impact factor. In [15] thirteen significant case studies were analysed to identify the most frequent used criteria categories. The usage frequency of a criterion category is determined by the number of case study criteria belong to it. Criterion categories are based on the current taxonomy used by Plato which classifies criteria in: Outcome Object, Outcome Format, Outcome Effect, Action Runtime, Action Static and Action Judgement. Figure 3.1 shows the criteria category frequency distribution over all case study plans and visualizes it via a pie chart. Criteria belonging to the Outcome Object category are used the most. This is not surprising because those sort of criteria fitting in this category refer to checks which verify that the original object was preserved correctly and therefore has identical characteristics than the transformed one. Criteria belonging to the categories Action Runtime, Action Static and Action Judgement criteria have 25 percent share on overall criteria. This criteria concentrate mainly on evaluating the executing action. Outcome Format criteria follows with 11 percent share. Outcome Effect criteria are only little used.

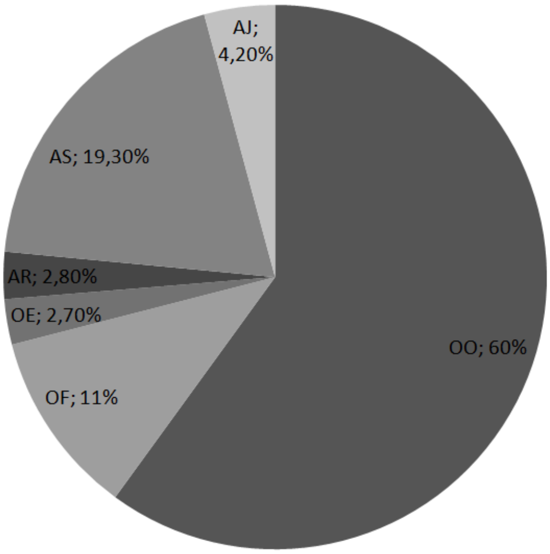


Figure 3.1: Existing research based on criteria usage frequency [15]

The results of existing research presented above only gives an overview about usage frequency of criteria and their belonging categories. In this theses we are interested in identifying concrete criteria impact. Criteria impact can be defined as contribution a criteria makes at finding the optimal decision. To calculate this impact taking only criteria usage frequency into account is not enough. Just because a criterion or criteria category is used very often does not mean that it has a high impact on the final decision. There are a lot of criteria factors aside usage frequency which need to be taken into account when it comes to assess criteria impact. Those factors are for example criteria weight, variance, etc. Finally decision criteria impact is always scenario dependent because influential criteria differ from scenario to scenario. For example decision criteria used to evaluate image preservation strategies are different from the ones used to evaluate video game preservation strategies. Thus identifying the most influential decision criteria without taking scenarios into account would end up in misleading results.

The starting point of our work and the tasks which have to be fulfilled to reach our final goals are shown in figure 3.2. The steps *extract properties*, align properties into reconciled model and *map case study criteria to properties* are described in chapter 3.2. Step *Define impact factors* is described in chapter 3.3. *Create analysis tools* step can be found in chapter 4. Finally *analyse case study data* and *results* are presented in chapter 5.

The data we base our analysis on is case study data in the form of preservation plans. These preservation plans contain decision criteria which describe objectives to be tested. Several things prevent us from analysing this data right away. These are listed in the following.

To be able to analyse present case study data we have to identify criteria with the same meaning. Without knowing which criteria belong together aggregation of data makes no sense. The fundamental problem is that case study criteria are stored text only with no identifier attached which can be used to identify common criteria in an automated way. For example a criteria re-

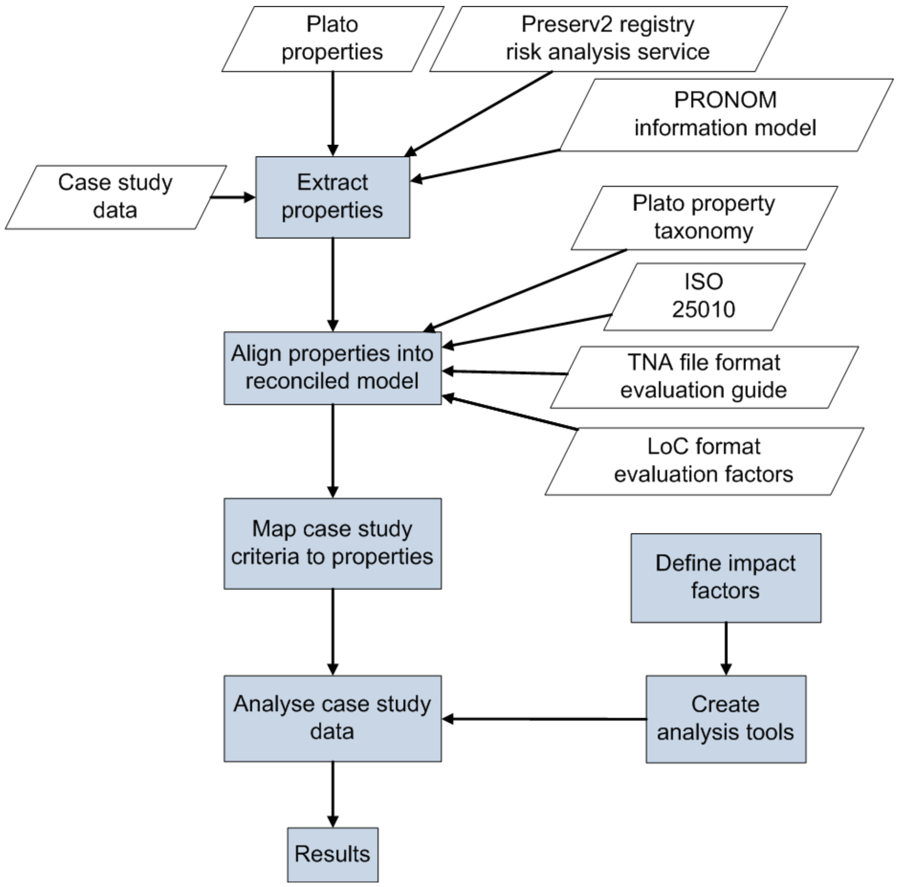


Figure 3.2: Thesis tasks

lating to the width of an image may be named *image width* in one plan, *picture width* in a second or even *photo width in another*. Human beings can indeed read such decision criteria consisting of text, understand their meaning and therefore identify common ones. Machines cannot understand the meaning of text. Thus they require a unique identifier which helps them to identify which criteria have the same meaning and therefore belong together. This unique identifier has to be attached to criteria by human beings because only they are able to understand the text.

Tagging criteria with unique identifier is done by assigning them predefined properties of the preservation context which have a predefined meaning. Each property consists of one unique identifier and a human understandable explanation. Consequently assigning properties to criteria identifies their meaning and prepares them for automatic processing. Criteria with the same unique identifier have the same meaning and therefore can be aggregated.

To cover nearly all criteria with related properties a huge number of them has to be defined. Because of this numerous properties it makes sense to introduce some kind of hierarchical structuring. This alignment of defined properties in a structured way is called property model. It can be useful in several ways, for example to find the appropriate property for a given criterion much easier.

There already exists a property model in Plato. It is based on a shallow hierarchy which classifies each property in one of the main criteria categories *outcome:object*, *outcome:format*, *outcome:effect*, *action:static*, *action:runtime* or *action:judgement*. The drawback of this model is that it is populated sparsely. Only properties which are automatically measurable are contained in this model. This comes from the fact that the focus at model creation was to support the user at criteria evaluation. The focus in this thesis is completely different. We want to achieve the maximum possible property coverage of all case study criteria to be able to analyse them in detail. Impact analysis based on the given data only makes sense if a very high percentage of all criteria have a property assignment. To achieve this, we need to extract properties based on the given case study criteria. Doing this the number of properties will emerge drastically. Thus we need a better and more precise way of structuring these properties resulting in the need of a more formalized property model.

Formalized property models still exits apart from the one actually used by Plato. Those can be consulted for guidance at building the new structured property model. These models are promising but they need to be homogenised and customized first to fit into our context.

In summary it is necessary to enrich case study criteria with semantic. This is inevitable for analysing their inherent data in an automated way. Semantic is attached to criteria by assigning related properties to them. Properties have a predefined meaning and are aligned in a property model for better maintenance. There already exist several property models applicable to use in the preservation planning context, which have to be taken into account. The next chapter describes the process of building a unified property model which can be used to assign properties to criteria and therefore make them meaningful for analysis.

3.2 Building an unified property model

As an analysis of preservation planning criteria presented in [15] showed, criteria in digital preservation can be classified as into two main categories called outcome and action. Those can be further sub classified in outcome:object, outcome:format, outcome:effect, action:runtime, action:static and action:judgement. “The analysed criteria differ in the information sources they depend on to obtain measurements, i.e. the source and type of measurement and what entity it needs to be applied on.” [15]. Section 2.3 discusses this taxonomy in more detail. [15] takes a more detailed look on this and defines the categories and subcategories as follows:

> “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)."

This taxonomy is currently used in the preservation planning tool Plato and therefore is our starting point for building a unified property model.

As stated in chapter 2.3 there already exist other interesting taxonomies which fit quite well in this context and therefore can be used for guidance. For the whole action category the ISO 25010 software components quality attribute model can be used as reference [28]. For outcome:format category the following inputs have been identified as relevant: The library of Congress format evaluation factors [30], The National Archives file format evaluation guide [10], Preserv2 registry Risk analysis service [6]. For outcome:effect no model has to be created because only two properties are in this category yet and therefore there is no need for structuring. Outcome:object properties are very object centred. Thus object characteristics like content-type or object-type are used for structuring its containing properties.

The property model created is derived from all collected case study criteria. The approach used to build this model is an incremental one. First of all, the meaning of each criterion has to be understood by a human being, taking into account its textual description, its alignment in the plans criteria structuring, its scale and if available additional comments. If those data was not sufficient to determine the meaning of a criterion the original decision makers were consulted to resolve ambiguities and provider clarification. Dependent on the now identified criteria meaning it has to be decided how a useful property can be created out of it or with which still existing property it matches. The next decision to make relates to the Scale. It has to be decided which scale is the most appropriate. Finally the property itself has to be defined in a clear and understandable way to be reusable in future plans. After the property is defined, its appropriate position in the hierarchical property structure has to be found. A new identified property is is either incorporated by creating a new property or merging it with existing ones. After all properties were created we were able to map case study criteria to its relating properties.

Although this process is inherently subjective, we tried to reduce this subjective factor by discussing and aligning the created model at different stages with colleagues and partners. The desired target for property creation was to achieve a 100% criteria coverage with defined properties. We refrained from achieving this goal because some criteria were only used in one of the case study plans and were so plan specific that creating a public available property out of them did not make sense.

The next sections explain in more detail the creation of all property model parts.

Action Properties

Action properties refer to properties of the preservation action. In the taxonomy used by Plato this class of properties is further divided into *action runtime*, *action static* and *action judgement*. The *action runtime* category can be described as "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." [15]. *Action* *static* as “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.” [15]. *Action judgement* as “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.” [15].

The ISO standard 25010 - Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models [28] is based on the earlier ISO 9126 family [26]. Its purpose is to define a hierarchy of high-level quality attributes with evaluation procedures. Because preservation planning has a specific focus, different compared to generic cases of software product selection, it is necessary to customize the quantitative part of evaluation. This is a normal process expected and recommended by ISO 9126 and 25010 and already carried out by Franch and Carvallo as shown in [23]. Thus the main input from ISO 25010 to our unified action quality model is the hierarchical structure which needs to be filled up with properties derived from our case study criteria. The ISO hierarchical structure is defined in [28] as follows:

- **functional suitability:** Degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions.
  - **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.
  - **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.
  - **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.
  - **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.
  - **appropriateness recognisability:** Degree to which users can recognize whether a product or system is appropriate for their needs.
  - **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.
  - **operability:** Degree to which a product or system has attributes that make it easy to operate and control.
  - **user error protection:** Degree to which a system protects users against making errors.
  - **user interface aesthetics:** Degree to which a user interface enables pleasing and satisfying interaction for the user.
  - **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.
  - **maturity:** Degree to which a system, product or component meets needs for reliability under normal operation.
  - **availability:** Degree to which a system, product or component is operational and accessible when required for use.
  - **fault tolerance:** Degree to which a system, product or component operates as intended despite the presence of hardware or software faults.
  - **recoverability:** Degree to which, in the event of an interruption or a failure, a product or system can recover the data directly.

- **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.
  - **confidentiality:** Degree to which a product or system ensures that data are accessible only to those authorized to have access.
  - **integrity:** Degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data.
  - **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.
  - **accountability:** Degree to which the actions of an entity can be traced uniquely to the entity.
  - **authenticity:**
- **maintainability:** Degree of effectiveness and efficiency with which a product or system can be modified by the intended maintainers.
  - **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.
  - **reusability:** Degree to which an asset can be used in more than one system, or in building other assets.
  - **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.
  - **modifiability:** Degree to which a product or system can be effectively and efficiently modified without introducing defects or degrading existing product quality.
  - **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.
  - **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.
  - **installability:** Degree of effectiveness and efficiency with which a product or system can be successfully installed and/or uninstalled in a specified environment.
  - **replaceability:** Degree to which a product can replace another specified software product for the same purpose in the same. environment

Taking a more detailed look on the two introduced action taxonomies reveals the differences between the two used approaches. The current Plato taxonomy is strongly adapted to the digital preservation context because it was created based on the measurement needs of digital preservation action criteria. Therefore its classification is based on the source (action judgement) and type (action static, action runtime) of measurement. The intended purpose of ISO 25010 is to evaluate the quality of software components. Preservation actions are software components as well thus ISO 25010 can be used to evaluate them. The classification used is based on inherent quality properties of the software which should indicate how good the software meets stated needs like performance efficiency, compatitlity, usability, etc. Hence the Plato taxonomy and ISO 25010 use different approaches for classifying action properties. Plato taxonomy uses a bottom-up classification based on action measurements and their possible automation. In contrast ISO 25010 uses a top-down approach which classifies based on action goals and their meaning. At action criteria definition responsible planners mainly have their main goals in mind which they want to evaluate. They usually do not care about in which way this criteria can be measured. Consequently we decided to use ISO 25010 to classify our action properties.

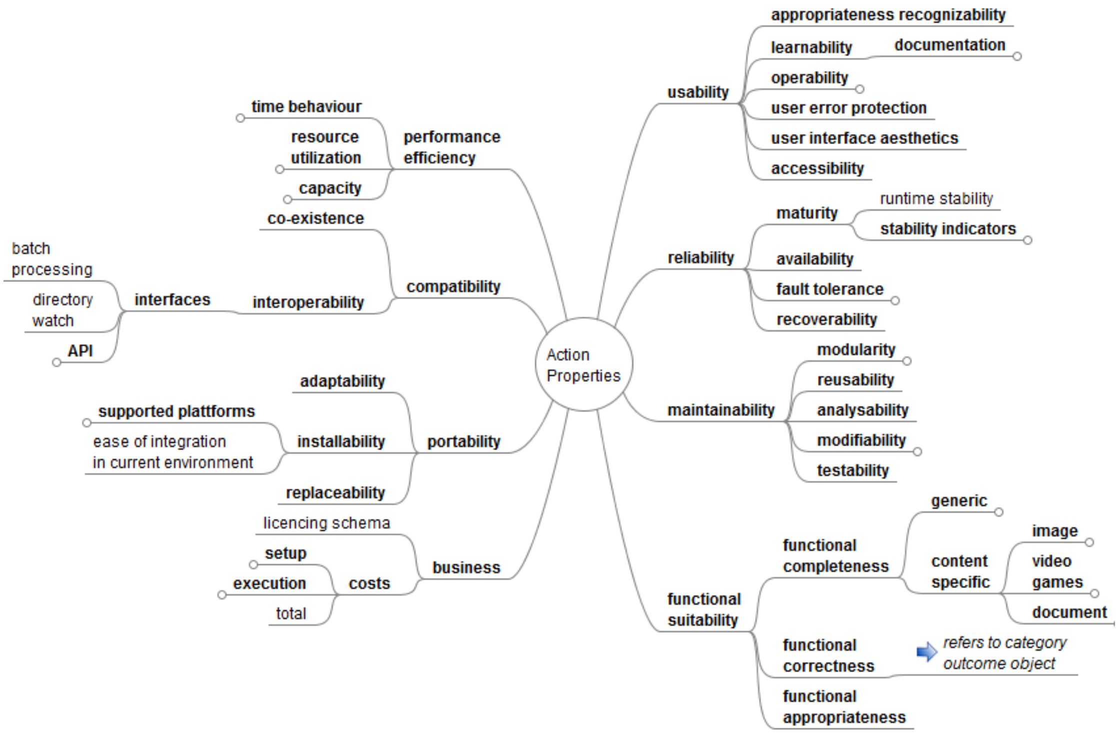


Figure 3.3: Action properties: High-level classification

Having established ISO 25010 as the main classification, the next step is to derive appropriate properties from case study criteria and align them into the given categorization. Figure 3.3 shows the high-level structuring of all action properties. Structural nodes are presented in bold letters, properties in standard letters.

At aligning properties into the two level ISO 25010 categorisation additional structural nodes were added to build a clear and logical hierarchical structure. The concrete reason for adding each of the additional nodes is explained in the following. The category *functional correctness - functional completeness* in general refers to features of the preservation action. A separation between *generic* and *content-specific* features is created to distinguish between features applicable to all kind of actions like error reporting, number of input-formats supported, etc. and features which are very specific to content-specific action tools. This separation helps the responsible planner at property selection so he is able to navigate straight forward to his desired property without having to browse properties which he is not interested in. The content-specific separation is further segmented by the different content types the action tool is applied to. This helps to filter the domain-specific features. For example, the indication of *XMP sidecar support* is very specific for image migration tools. Thus this property is located in the *content-specific - image* category. *Category compatibility - interoperability* is extended by a structural node *interfaces* to make clear that all the properties following refer to interfaces of the action tool. Category *portability - installability* is extended by a node *supported platforms* following by properties which indicate on which concrete plattforms the action tool is installable. *Usability - learnability* is extended by a node *documentation* because all following properties deal with documentation concerns. *Reliablity - maturity* adds one additional structural node *stability indicators* to distinguish between properties which give direct stability feedback and properties which are just stability indicators. For example the measured *runtime stability* of an action tool delivers direct stability feedback. On the other hand knowing that the *tool manufacturer is still in business* is good to know because this infers us that the software bugs are fixed and we can get support if needed but it. Anyhow based on this fact we cannot directly infer that the tool is stable.

The ISO main categorization fits quite well for aligning all criteria related properties but one aspect is missing. This relates to business concerns. ISO 25010 is intended to measure only software quality therefore it is clear that it does not cover business concerns. At making preservation decisions business requirements like *software licence costs* or *hardware obtaining costs* of course play an important role at decision making. Thus a new main category *business* is created with one subcategory *costs*. Taking into account the economic paradigm costs are further classified into *setup costs* and *execution costs*. *Setup costs* cover the inital costs, *exection costs* the running costs incurring at applying the action tool.

One category of the used taxonomy is of special interest for us. It is *functional suitability - functional correctness* which is defined in ISO 25010 as “Degree to which a product or system provides the correct results with the needed degree of precision.” [28] As this definition states, it contains properties used to assure the functional correctness of the performed action. The task of assuring that the preservation action produces correct results is the main goal of digital preservation and therefore is very relevant. Thus properties belonging to this category are put in an own main property category called *outcome object*. This category is discussed later in this chapter in more detail.

Having discussed the property structuring we now take a more detailed look on the action properties. Figure 3.4 shows all defined action properties except the ones belonging to category *functional suitability - functional correctness*. Those are showed in figure 3.4. Again structural nodes are presented in bold letters, properties in standard letters

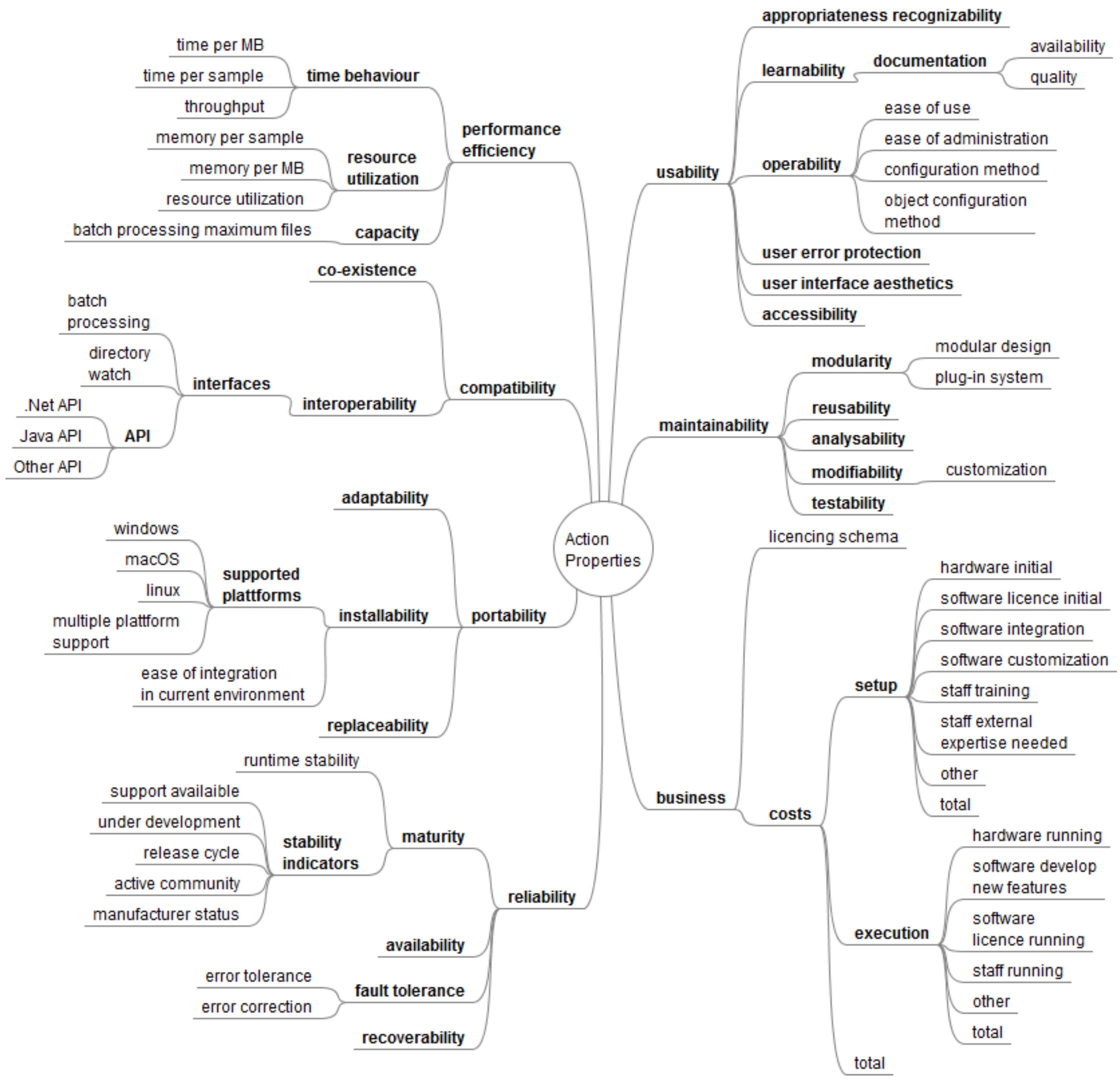


Figure 3.4: Action properties

As we can see some categories are more populated with properties than others. A high category population indicates its high suitability in the preservation context. Just to make this clear this has nothing to do with the impact on final decisions. Dense category population just indicates that the category covers a preservation relevant topic. Take for example the category *usability - accessibility* which is not populated at all. Thus we can conclude its low suitability in the preservation context. This can be explained by the fact that preservation actions are usually tools which function fully automatic or are at least used by a very limited group of people. On the other hand *functional completeness - functional suitability* is dense populated category. This high indicated suitability is not surprising because the supported features are of course of interest for nearly every preservation action.

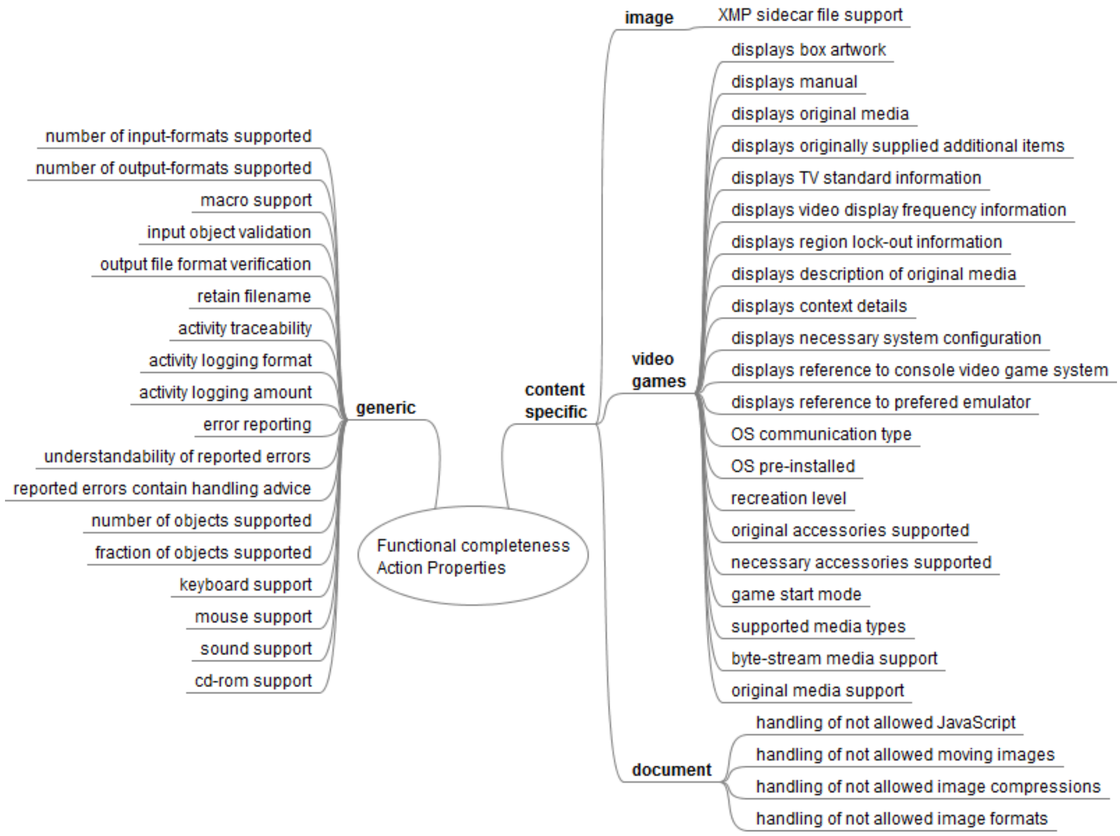


Figure 3.5: Action properties: Functional suitability - functional completentess

An extract of a few defined action properties is listed in table 3.1.

Outcome Format properties

Format properties refer to representation characteristics of the outcome. These representation characteristics are classified in the actual used taxonomy of Plato by the category *outcome format* which is defined as follows: “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.” [15]. Because this taxonomy at the moment only has 12 properties assigned and contains almost no structuring it needs to be improved. Several models dealing with formats are presented in the following, which are used for guidance to build a better, more formalized format property model.

Property	Description	Scale	Unit
time per MB	Elapsed processing time used per MB, measured in milliseconds.	Positive Number	ms/MB
batch processing support	Indicates whether the action-tool supports batch processing.	Boolean	
activity traceability	Indicates to which degree the activities performed by the action-tool are traceable.	Ordinal: good/acceptable/poor/no	
licencing schema	Licence applying to the action-tool.	Ordinal: openSource/closedSource	
hardware initial costs	Initial hardware costs arising when using the action-tool (e.g. new hardware needs to be acquired to get the tool running).	Positive Number	Euro

Table 3.1: Extract of action properties

The Library of Congress format evaluation factors

The Library of Congress (LoC) provides information about digital content formats on their website. As part of this a list of format evaluation factors are presented in [30]. These are:

- **Sustainability**
  - **Disclosure:** Degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content. A spectrum of disclosure levels can be observed for digital formats. What is most significant is not approval by a recognized standards body, but the existence of complete documentation.
  - **Adoption:** Degree to which the format is already used by the primary creators, disseminators, or users of information resources. This includes use as a master format, for delivery to end users, and as a means of interchange between systems.
  - **Transparency:** Degree to which the digital representation is open to direct analysis with basic tools, such as human readability using a text-only editor.
  - **Self-documentation:** Self-documenting digital objects contain basic descriptive, technical, and other administrative metadata.
  - **External Dependencies:** Degree to which a particular format depends on particular hardware, operating system, or software for rendering or use and the predicted complexity of dealing with those dependencies in future technical environments.
  - **Impact of Patents:** Degree to which the ability of archival institutions to sustain content in a format will be inhibited by patents.
  - **Technical Protection Mechanisms:** Implementation of mechanisms such as encryption that prevent the preservation of content by a trusted repository.
- **Quality and functionality:** Quality and functionality factors pertain to the ability of a format to represent the significant characteristics of a given content item required by current and future users. These factors will vary for particular genres or forms of expression for content.

It is clear that these high-level factors are not directly measurable and require assignment of more specific properties to be quantifiable.

The National Archives file format evaluation guide

The National Archives (NA) also provides a method to assess the continuity properties of file formats. This is done by assessing the given file format against each of the following characteristics and sub-characteristics preseted in [9]:

- **Capability:** How well your business requirements are met?
- **Quality:** How accurately your information is stored?
  - **Precision:** Is data represented to a sufficient precision?
  - **Lossiness:** Does the format intentionally throw information away?
- **Resilience/Safety:** How resilient your information is to time?
  - **Ubiquity:** How widespread is the use of the format?
  - **Stability:** How long will the format be supported by software?
  - **Recoverability:** How resilient is the format to accidental corruption?
- **Flexibility:** How well you can adapt to changing requirements?
  - **Interoperability:** How much existing software can access the format?
  - **Implementability:** How easy is it to write software to interact with the format?

The given list is not intended to be fully complete and needs customization and extension dependent on the given context. All this characteristics are high-level characteristics and therefore have to be evaluated by more detailed measurable characteristics.

Preserv2 registry: Risk analysis service

The Risk analysis service of the Preserv2 registry [6], evaluates the usage risk of a given format. The evaluation is based on extracted data from the PRONOM registry [10] and DBpedia [2]. Especially interesting for us are the criteria which are used to evaluate the usage risk and how they fit in our format properties model. These criteria presented in [6] are:

- **Ubiquity**
  URI: http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/ubiquity
  Scale: Ordinal: most widely used, widely used, occasional, specialized, deprecated, obsolete
- **Documentation Quality**
  URI: http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/documentation_quality
  Scale: Ordinal: high, complete, suitable, poor, unusable

- **Stability**
  URI: http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/stability
  Scale: Ordinal: stable, compatible, not compatible, unstable

- **Identification Type**
  URI: http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/identification
  Scale: Ordinal: positive specific, positive generic, tentative, unidentifiable
  (this refers to automatic identification)

- **Format Type**
  URI: http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/format_type
  Scale: Ordinal: compressed, lossless, lossy

- **Rights**
  URI: http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/rights
  Scale: Ordinal: ipr protected, open, proprietary

- **Complexity**
  URI: http://p2-registry.ecs.soton.ac.uk/pronom/risk_categories/complexity
  Scale: Ordinal: low, medium, high

These factors do mix up a few issues. For example, format type is about compression (maybe just a strange choice of words). In Addition an exact and unambiguous specification of the presented scales is clearly necessary. Some evaluation examples for different pdf-formats can be found in [35] and [31].

PRONOM information model

"PRONOM is an online registry which stores technical information about file formats, software products or other technical components which are required to support long-term access to electronic records and other digital objects of cultural, historical or business value." [10] This information is made publicly available via a web interface. In particular interesting for us is the information model which is used to store this information [19] because the here used attributes to store file format information can help us to define appropriate format properties.

Consolidated Format Model

The consolidated format model is presented in figure 3.6. Structural nodes are presented in bold letters, properties in standard letters. The process of model building is described in the following.

As mentioned before Library of Congress format evaluation factors are high level factors which are intended to be evaluated by concrete properties. We use these factors as main structural nodes in our consolidated format model. This base structure serves as starting point for further elaborations through other models. Only the factor *impact of patents* suggested by the LoC was not used directly because it appeared too delimiting. Patent protection is just one aspect of applying rights therefore a structural node *rights* was added instead of it.

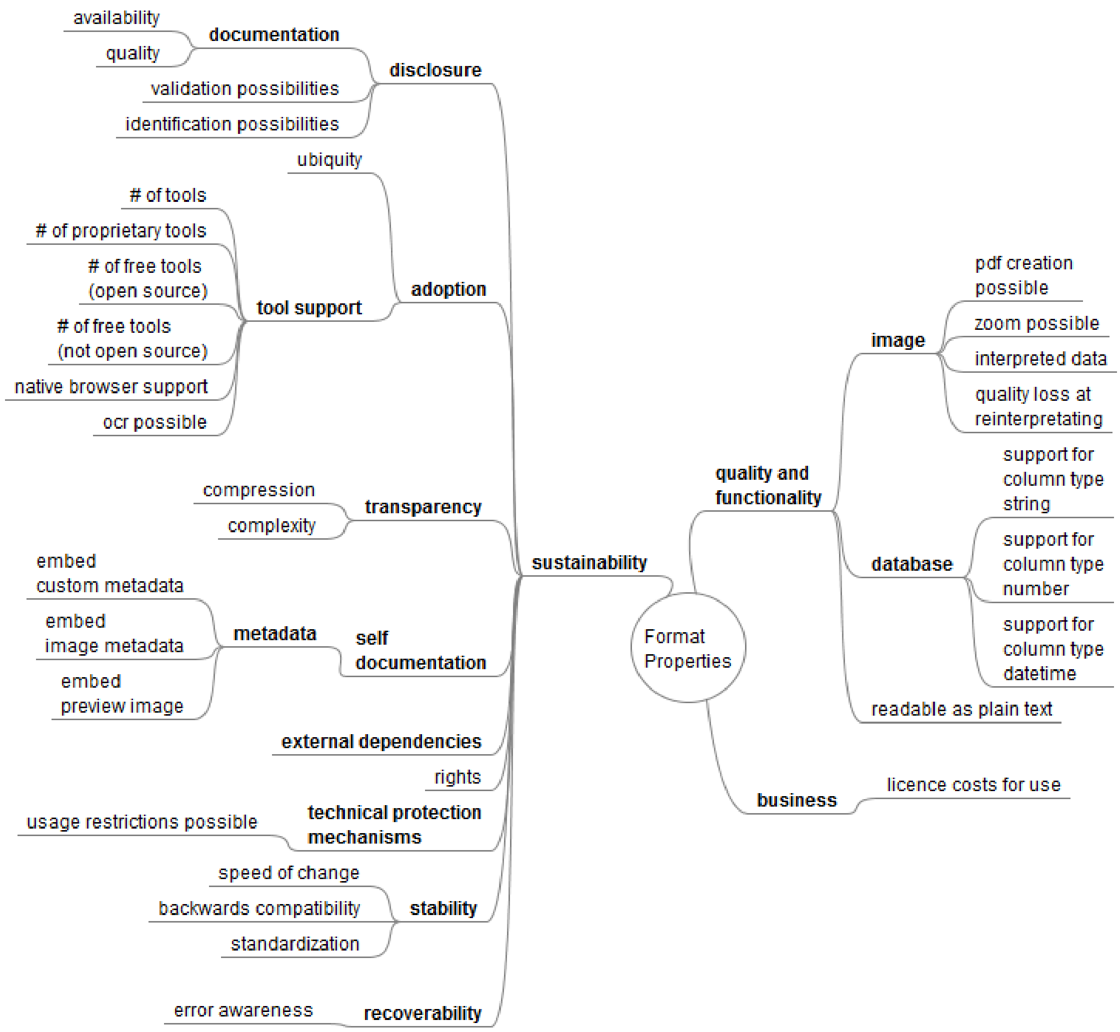


Figure 3.6: Format properties

The National Archives also suggest factors to assess the continuity of file formats. If we compare those two approaches we see that they use different viewpoints from which the format evaluation task is tackled. The LoC factors mainly focus on the format representation itself. For example the factors concentrate on how sustainable a format is by sub factors like format adoption, format disclosure or format stability. On the other hand NA suggested factors mainly focus on the format related information continuity or in other words on the ability of the given format to ensure information continuity. This can be seen by factors like *resilience/safety* which assess how resilient your *information* is stored based on the given format. The viewpoint of NA evaluation factors is also expressed their usage statement: "This guidance is primarily aimed at information and IT managers who need to assess the usage of file formats in different business situations across their organisation and beyond." [9]. In our consolidated model we used the

LoC way of structuring our properties because the focus on format representation seemed more intuitive to us for defining format related criteria.

The risk analysis service of Preserv2 provides a list of factors to assess the usage risk of formats. The difference to the factors presented from LoC and NA is that these factors are not high-level factors but concrete factors at property level. They furthermore have a defined scale which proved useful to identify their concrete meaning. Preserv2 factors were used to align the case study extracted properties and their scale with these accepted and approved factors.

The Pronom Information Model specifies a set of properties used to store file format information into Pronom registry. Like Preserv2 factors, Pronom file format properties were used to align the case study extracted properties with those approved properties.

In the following we describe the process of building our consolidated format properties model. As mentioned before we used the LoC evaluation factors as main categorisation and starting point for further actions. The given separation into *sustainability* factors and *quality and functionality* factors worked out quite well because it separates two of the main format concerns of preservation planning: How sustainable is the format? and what are the capabilities of it? This can be seen by the fact that all case study aggregated properties except one relate to one of these two main categories. The property which did not fit into this classification is *licence costs of use*. This is a business related property which is not covered by any kind of format assessment. Thus we create an own main category for it called *business*.

Category *sustainability - stability* was added to adhere all format stability factors which indeed are a concern of sustainability. For example, the standardization status of a format defines its stability and therefore its sustainability over the next years. Finally category *sustainability - recoverability* was added. It refers to the *continuity - recoverability* factor proposed by The National Archives. Properties in this category should indicate how resilient a format is regarding errors. If a file format is more aware of errors it becomes more sustainable. Even if errors occur in a file, it is more probable that the information is not lost.

Having established the main category levels, the additional sub-classification was able to be added in order to structure all properties in a meaningful way. This additional sub classification is added to demarcate specific property groups resulting in a more logical structuring. For example, an own *documentation* class was introduced in category *sustainability - disclosure* to specifically separate achieved disclosure through documentation. Under *sustainability - adoption* a new class *tool support* was added to specifically show that adoption measured through tool support is different than general assessment of ubiquity. Adoption assessed through ubiquity and through tool support can deliver different results. For example, trying to assess Microsoft Word format adoption, tool support (measured by the number of tools supporting this format) is rather low because finally only Microsoft Office Word can handle the format in a perfect way. On the other side ubiquity is very high because almost everyone deals with Microsoft Office files. Thus the way high level factors are evaluated effects their final result. Another structuring node introduced to further specify the category *sustainability - self-documentation* is *metadata*. It is intended to group properties refering to format self documentation through metadata. As denoted above the main class *quality and functionality* deals with properties relating to features the format supports. Format features are usually content specific thus additional classification based on content type was introduced here. Thus two classification nodes called *image* and *database* were added.

Property	Description	Scale
standardization	Standardization of the outcome format.	Ordinal: international standard/de facto standard/none
compression	Compression used in outcome format.	Ordinal: none/lossless/lossy
# of tools	Number of all tools which support the outcome format.	Positive Integer
documentation availability	Availability of the outcome format documentation.	Ordinal: yes-free/yes-pay/no
ubiquity	Ubiquity or popularity of the outcome format.	Ordinal: ubiquitous/widespread/specialised/obsolete

Table 3.2: Extract of outcome format properties

Based on the resulting model each property was assigned to its relating category. An extract of a few resulting outcome format properties is listed in table 3.2.

Outcome Object Properties

Outcome object properties refer to characteristics of the outcome object produced by a preservation action. These properties not only include characteristics of the outcome object itself, but also characteristics indicating if the outcome object is still conform to the input object. In the taxonomy currently used by the preservation tool Plato, this category of properties is defined as: "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." [15] As stated in this definition all outcome object Properties are based on simple object properties. So we first take a detailed look on object properties and how they can be hierarchical structured, and then how these object properties are used to form outcome object properties.

Object Properties

Due to the fact that object properties are very content specific, the main classification of our created object property model is given by the content types of our case study preservation objects. These are: image, document, video game, and database.

Having established the main classification case study criteria were aggregated to properties again but in a slightly different way than the times before. The difficulty was to extract the appropriate object properties based on the given criteria text. Those criteria texts often contained constraints regarding the original object which made it hard to identify the appropriate object property. For example it is easy to derive the object property *image width* from criteria like *image width equals original* or *image width match*. It is even harder to figure out that the criterion *all pages have been migrated*, refers to a property *number of pages* whose evaluation has to be compared between input and output. Based on this example we see that object related user criteria either refer to the outcome object itself or to a relation between original and outcome object. This differentiates outcome object criteria from action or outcome format criteria. Keeping this in mind in the following we only focus on simple object properties and their alignment in a property model coming back to this issue later in chapter 3.2.

The final object property model resulted from aligning all derived object properties into the given taxonomy based on content type. In the following the different parts of this property model are discussed in detail.

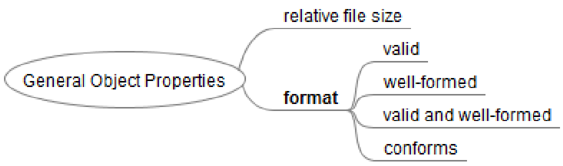


Figure 3.7: General Object Properties

General Object Properties At property alignment into the given structure based on object content types, further structural nodes were added to align the properties in more logical therefore easier to understand groups. First of all there were a few properties which were generic enough so they needed no content type specific grouping. Almost all of these relate to format concerns of the outcome object and therefore were grouped into a own class named format. It contains properties which assure that the outcome object is valid, conform and well-formed. There was only one property named relative file size which does not belong to a specific content type and also does not relate to format concerns. It therefore was placed on the root level of the hierarchy. These properties are shown in figure 3.7

Image Object Properties Image objects were analysed with accuracy due to the availability of case studies for this type of objects.

at the most accurate because a lot of case studies were present for this type of objects. The resulting object properties for images are shown if figure 3.8. To deal with the huge set of image properties reasonable, additional structuring elements had to be added. Structuring nodes were already used by the responsible planner by defining their criteria in the form of an objective tree. This structuring was taken into account with small adoptions for property structuring too. An own section for metadata properties was added to differentiate between concrete object data and describing data. This section is further divided based on the different metadata standards used in images: exif, iptc, xmp, dc, mets, etc. Furthermoe additional subclasses were added in this metadata classes to differentiate between the different aspects this metadata is describing. For example exif is additionally splitted into picture taking conditions, gps data, IFD0 related data and tiff related data.

Document Object Properties The next content type we take a look at is document. The structuring of object related criteria in document plans was very different. Thus a consolidated structuring based on all present hierarchies was constructed. The resulting properties are presented in figure 3.9. Document properties were structured by the following main categories:

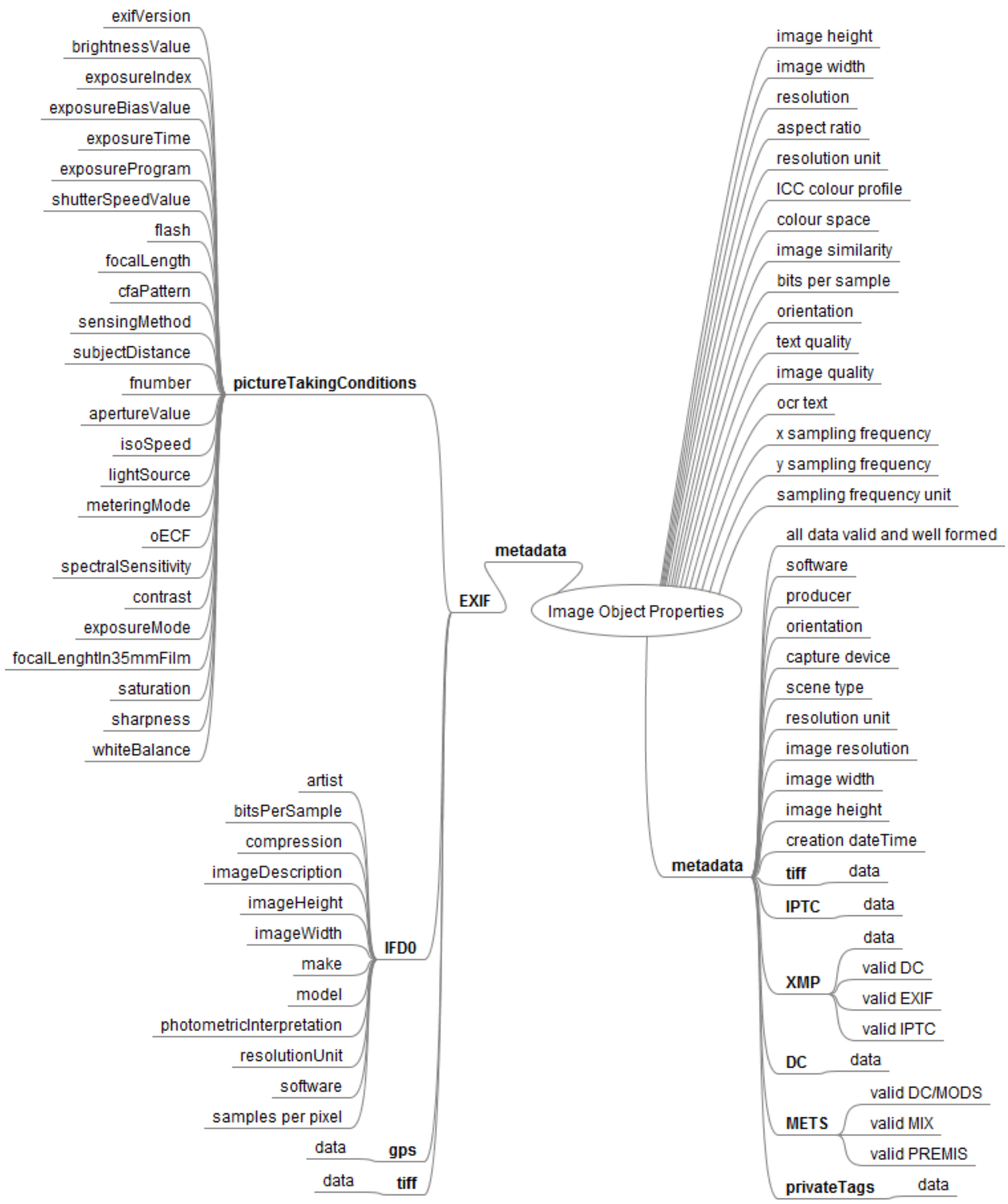


Figure 3.8: Image Object Properties

content, structure, layout, metadata, features. Category content contains properties caring about the conveyed information of the document. It contains properties concerning about text, tables, figures, fonts, etc. Category structure deals with all elements used for structuring the comprised information but not with the information itself. Representing properties are page numbering, table of contents, etc. Category layout cares about the way the overall document is presented. Therefore properties like page size, or background are relevant. Structural node metadata refers

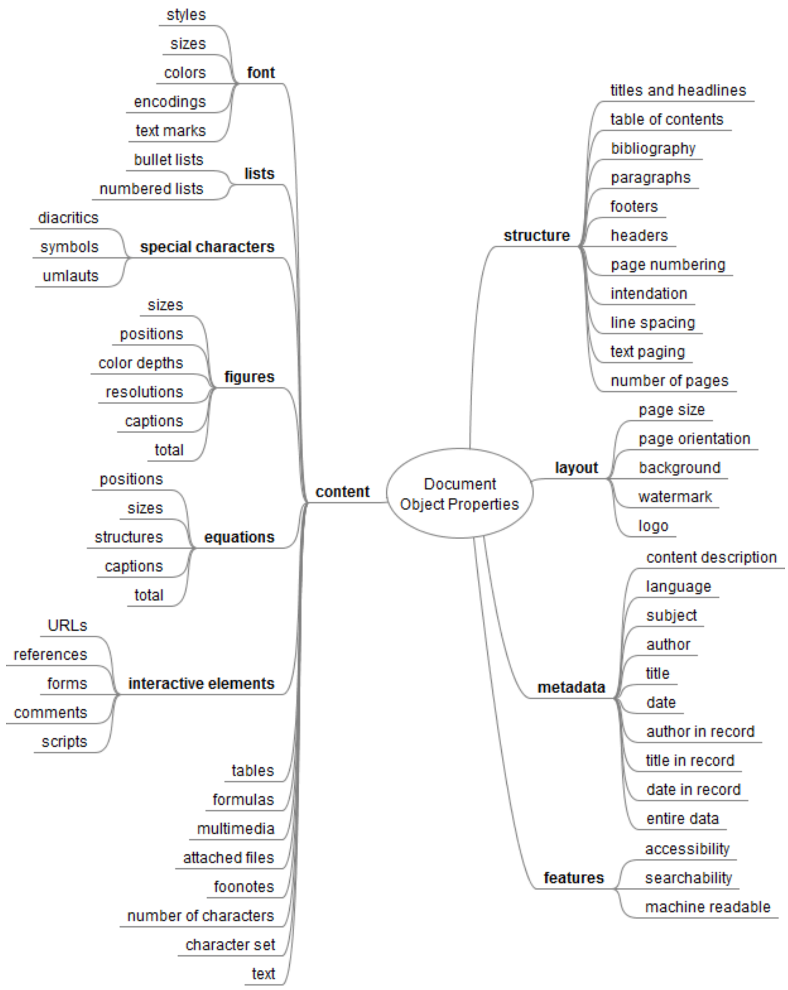


Figure 3.9: Document Object Properties

to all descriptive data of the document. Category features refers to all operations the document support like accessibility and searchability.

Video Game Object Properties Video games are a little bit different than the content types we have discussed before. Video games are in general more interactive and less static. Therefore its object properties are more focused on game performance, features and interactivity aspects. The resulting model is presented in figure 3.10. Interactivity aspects were summarized by a corresponding section interactivity whose main section is input. It tries to cover all supported input features by separating between personal computer standard hardware, personal computer gaming hardware and original hardware. Personal computer standard hardware refers to simple mouse and keyboard, personal computer gaming hardware to PC related gaming hardware like gamepads, joysticks, etc. Original hardware refers to the original hardware used to play

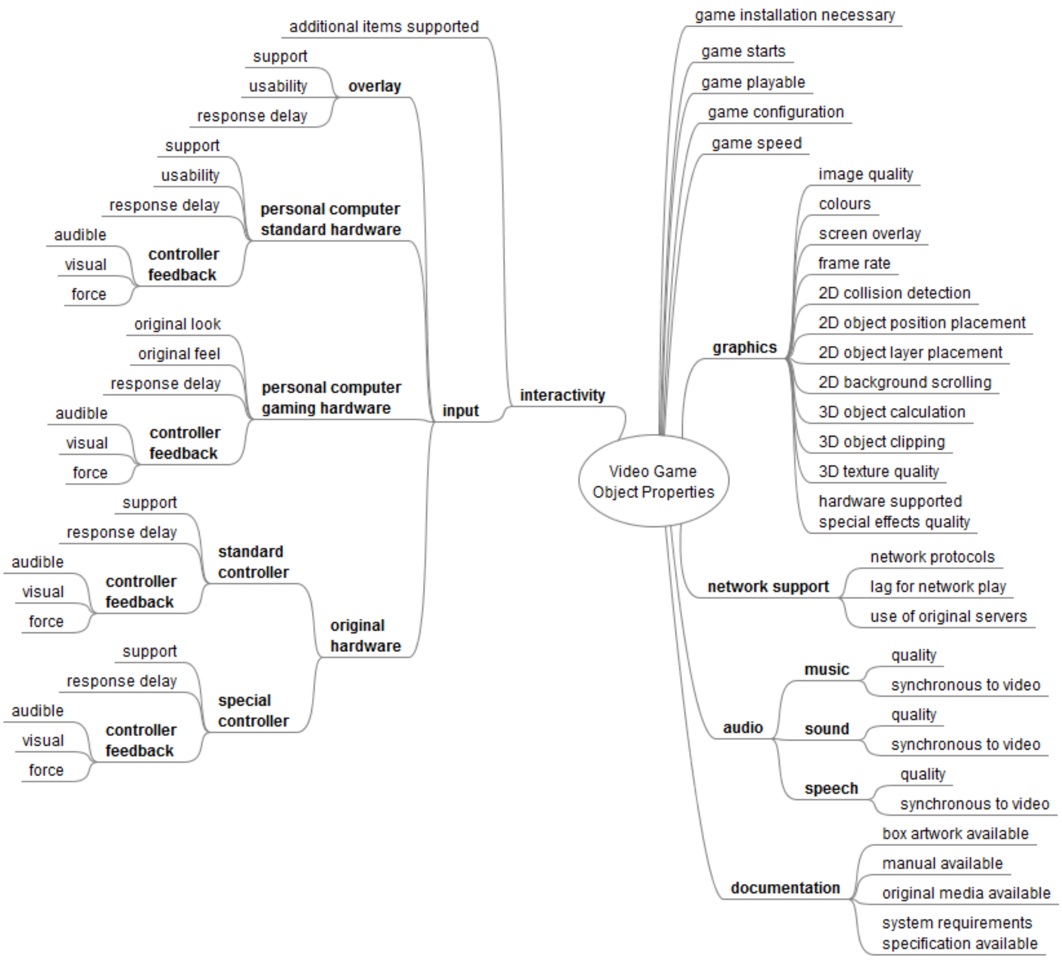


Figure 3.10: Video Game Object Properties

the game in its original environment. Performance aspects are covered in graphics and audio level. Additionally classes network support and documentation were added to group network and documentation features.

Database Object Properties Database properties can be again categorised in properties relating to content and properties relating to context. Content relating properties are for example tables, view, sequences, etc. stored in the database. Context properties are for example database name, database purpose, etc. Contextual database properties can be further grouped to achieve more clearance in where the properties belong to. For this additional groups archival process, data dictionary, data model and traceability were added. Despite the two main categories there is one additional category called representation which contains representation relevant properties like character or time encoding. The resulting property model is shown in figure 3.11.

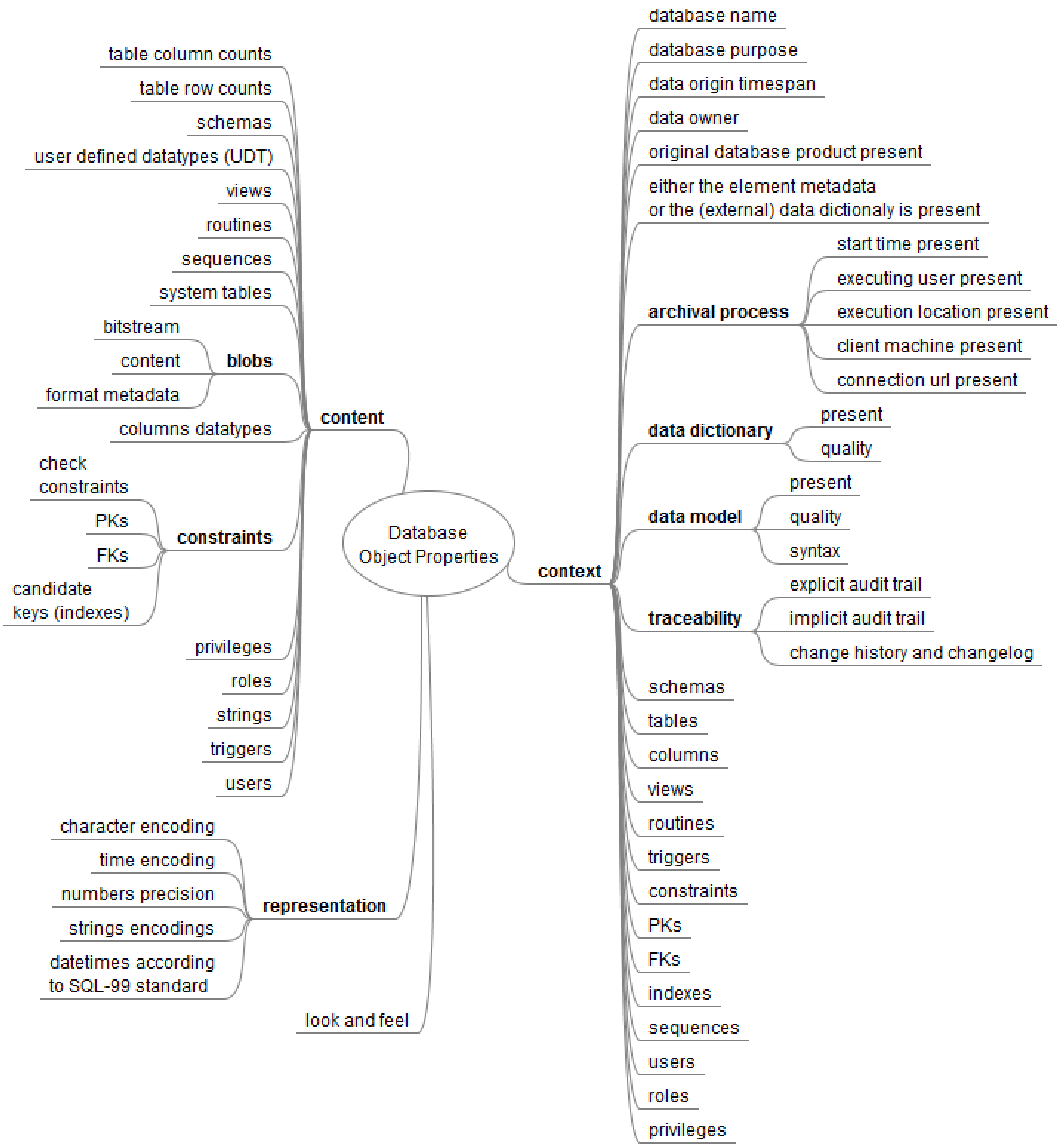


Figure 3.11: Database Object Properties

Outcome Object Properties based on Object Properties

As adumbrated before, outcome object relating preservation planning criteria are used in two major ways. They are used to assess simple outcome object characteristics like *image width*, *document searchable*, etc. and to check if the outcome object still conforms to the input object (e.g. image width equals). This conformity check is necessary because all the information to preserve is contained in the object, and therefore it is clear that it has to be checked whether this information was preserved correctly or not.

Taking this criteria usage into account reveals the question how outcome object properties have to be constructed based on object properties to be able to cover all possible criteria. By using the object properties defined in the previous chapter in a simple way it is possible to refer only to simple characteristics of the outcome object like image width, document searchable, etc. As seen before this is not enough to cover all criteria. Thus we also have to provide properties which check if the original object was transformed correctly to the outcome object? Testing if the outcome still conforms to the given input is done by comparing the object properties of input and output object. This conformance check is done by predefined metrics.

Taking this into account Outcome Object Properties can be classified into three major groups:

1. *Information Properties*
   Properties which refer to characteristics of the outcome object itself, or in other words simple object characteristics of the outcome object. This type of property is often used to check if the outcome supports specific features like *is this document searchable*, *is this document copyright protected*, etc. But it can also be used to check simple outcome object characteristics like image width, image height, etc.

2. *Representation Instance Properties*
   Properties relating to the outcome object representation. These criteria are very similar to Information Properties because they also relate to information about the outcome object, but the information focuses on representation information. Example properties are: validity or conformance of the outcome object format or simply the file encoding.

3. *Transformation Information Properties*
   Properties which indicate whether the outcome object conforms to the input object and therefore was preserved correctly. The presented classes up to now only focused on the outcome object itself - so about the features or its representation. But the main goal of digital preservation is to keep objects accessible and authentic. To see if an object stayed authentic we need to compare it against the original. But how do we do this? By comparing the significant object properties of the input object against the ones of the outcome object. This category is called Transformation Information because these properties check if the information from the input object was transformed correctly into the outcome object. Example properties are for example image width equal, document font sizes equal, etc.

As the analysis of all case study criteria showed, criteria regarding the outcome object always relate to on one of these three presented property types. A huge part of it relates to Transformation Information Properties.

Transformation Information Properties are based on the comparison of object properties extracted from the input and the outcome object. For this comparison, metrics have to be defined. For example, to check if the width of an image stayed the same through the preservation action, the object property image width can be compared by a simple metric called equal. This metric is be measured on a simple Boolean scale evaluating to true or false. Even more complex metrics exits but in fact we do not care how easy or how hard it is to evaluate a metric. If input and outcome can be compared in any way we can define a metric for it. Take for example the case we want to check if all tables in a document stayed the same. Tables in a document cannot be assessed on a simple scale but they are a valid part of a document so we can define a property for it. Consequently we can also define a comparison metric for it called equalJudged which is measured on an ordinal scale: equal/unsignificant changes/significant changes. Although the property document tables cannot be measured on a given scale yet it can still be compared for equality at least by human beings.

To identify appropriate metrics for our yet defined object properties we consulted our case study data again. For example the criteria *image width equals original* was used to add the metric equals to the object property image width. A more complex example may be the video game related criteria *image quality* evaluated on an ordinal scale: *nothing displayed/severe errors on whole image/errors noticeable but don not affect gameplay/near perfect/no difference to original noticeable*. In this case a metric *videoGameImageQualityCompliance* was attached to the property *image quality* which is measured on the same scale than the given by the criterion.

In conclusion there are three types of Outcome Object Properties which which are based on simple object properties. Information Properties and Representation Instance Properties simply refer to object properties of the outcome object. The difference between these two criteria is that they focus on different information aspects. Transformation Information Properties compare object properties of the input to object properties of the outcome via predefined metrics.

Property	**Description**	**Scale**	**Unit**	**Metrics**
relative filesize	Factor for relative output file size, calculated as: (size of output file / size of input file)	Positive Number	out/in ratio	
format: valid	Indicates whether the format of the object is valid.	Boolean		
image: width	Width of an image, measured in number of pixels on the x-axis.	Positive Integer	pixel	equal
image: colour space	Colour model of the decompressed image data.	Free Text		equal
document content: bullet lists	Bullet lists used in the document.	NOT MEASURABLE		equalJudged

Table 3.3: Extract of object properties

An extract of a few object properties with associated scale, their meaning and possible metrics is shown in table 3.3. Properties with the scale *NOT MEASURABLE* are complex properties which cannot be evaluated on a simple scale. Thus they cannot be used as Information or Representation Instance Properties. Nevertheless those properties can be used for comparison reasons as Transformation Information Properties by applying any of the associated comparison metrics. An example for this is the property *document content: bullet lists*, which represents all bullet lists of a given document. Although it can be compared and judged by human beings it cannon be evaluated on a simple scale. On the other hand object properties with no given metric cannot be used as Transformation Information Properties because no defined way to compare input to outcome is defined. They can only be used as Information Properties or Representation Instance Properties regarding their information context.

Outcome Effect Properties

Outcome object properties evaluate refer to consequences of the produced outcome. It is defined in the current taxonomy of Plato the following way: "Effect of outcome. This refers to any other effects caused by the application of a certain component." [15].

Analysis and property aggregation of case study criteria showed that only three properties fall in this category. Nevertheless this category plays an important role at decision finding because effects like the costs resulting from applying a preservation action are covered within. Only one structural node was added to group the two cost related properties together. All Outcome Effect Properties are shown in figure 3.12.

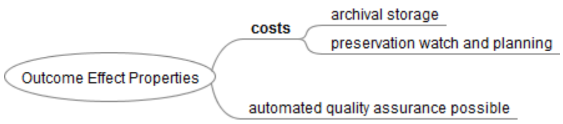


Figure 3.12: Outcome effect properties

Property	Description	Scale
archival storage	Effects on archival storage costs.	Ordinal: increase/unchanged/decrease
preservation watch and planning	Effects on preservation watch an planning costs.	Ordinal: increase/unchanged/decrease
automated quality assurance possible	Indicates if the outcome supports automated quality assurance.	Boolean

Table 3.4: All outcome effect properties

All defined outcome effect properties are listed in table 3.4.

3.3 Impact factors

Once sufficient properties are defined we are able to map case study criteria to defined properties and therefore make this data ready for analysing. We want to achieve several goals at analysing this data. In general we want to get insight into the the decision making process and its key factors. From the viewpoint of a decision maker we want to improve the efficiency of the preservation planning process while keeping full trustworthiness. For automation reasons we want identify the minimal representative set of criteria for a given scenario to accelerate the currently decelerate criteria definition step. Based on these goals several questions arise like: What is the impact of a single criterion? What is the impact of criteria sets? When is a given criterion added to the minimal representative set of a scenario? When a criteria can be stated as critical?

To answer the above raised questions in a clear and quantitative way we have to define metrics. Due to the defined goals impact factors for single criteria and criteria sets are defined as metrics. Criteria have a lot of characteristics which have to be considered at defining quantitative measures. They are: usage frequency, utility function, assigned values and evaluations. The following two chapters describe the defined impact factors in detail. The results of their application to our real-world case-study data is discussed in chapter 5.

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 criteria having the potential to reject alternatives.
IF14	Rejection Potential Rate	Percentage of criteria having the potential to reject alternatives.
IF15	Rejection Count	Number of criteria actually rejecting alternatives.
IF16	Rejection Rate	Percentage of criteria actually rejecting alternatives.
IF17	Reject Count	Number of rejected alternatives.
IF18	Reject Rate	Percentage rejected alternatives.

Table 3.5: Impact factors for single criteria. [25]

Impact factors for single criteria

This section is largely based on our work Impact Assessment of Decision Criteria in Preservation Planning [25].

Table 3.5 shows all the factors considered important for impact assessment of single criteria. Their meaning and their calculation is explained in the following.

IF1(Count) and IF2(Spread) are simple factors only considering the usage frequency of a criterion. Their calculation is straight forward and do not need further explanation.

More details about the usage of a criterion can be deducted from IF3(Weight) and IF4(Discounted Weight). Both factors care about the criterion weight. IF3 is calculated by averaging all criterion weights. The higher the value the more probable it is that the criterion has a high impact. IF4 additionally takes into account the usage frequency of the criterion. If a criterion is used more often it is discounted less than if it is used only rare. This is taking into account by summing up all criterion weights and dividing them by the number of all plans. This calculation results in lower values for rare used criteria.

Weight is often seen ans very important for the impact assessment. Although this is true there are still other factors which are at least as important as the weight. One of these factors is the potential output range. It indicates the evaluation change a criterion could potentially cause. It depends on the utility function assigned to criterion. Arguably this factor has more impact on the final decision than the weight [15]. Take for an example an utility function of a criterion which maps Yes to the final value 5 and No to 1. Let further the weight of this criterion be 0.25. The *potential output range (por)* is given by the weighted difference between the highest and the lowest utility result. In our example the result would be $0.25x(5-1)=1$. A high potential output range indicates a high potential effect on the final decision, a low potential output range exactly the opposite. It can have a value between 0 and 5. This value is most of the time much nearer to 0 than to 5 because weighting is also considered at calculation. Based on this potential output range, IF5(Potential) averages all criteria potential output ranges and IF9(Maximum Potential) results in the maximum potential output range. IF7(Discounted Potential), similar to IF4, takes into account the usage frequency by dividing all criterion potential output ranges in sum through the number of all plans.

Knowing the potential output range of a criterion the actual produced output range is of special interest. It indicates the evaluation change a criterion actually caused. So in fact the real world change this criterion was responsible for. The calculation is again based on the utility function but now deals with the concrete criterion values instead of potential ones. The *actual output range (aor)* is defined as the maximum alternative evaluation minus the minimum alternative evaluation of a criterion. Because of this evaluation it is clear that actual output range is always less than or equal to the potential output range, which itself as the name states indicates the maximum possible output range. A high actual output range indicates a high effect on the final decision, a low output range exactly the opposite. IF6(Range) averages all criteria actual output ranges, IF10(Maximum Range) results in the maximum actual output range. IF8(Discounted Range) takes the usage frequency into account by dividing the sum of all actual output ranges by the number of all plans.

Decision criteria utility functions are often defined very defensive taking all bad situations into account, although this worst cases are unlikely to happen in real world. To investigate how likely those potential bad outcomes actually are we are interested in the *relative output range (or variation)* of a criterion. Relative output range of a criterion is calculated by dividing the actual output range by the potential output range. A variation of 1 indicates that the whole potential output range is used by actual criterion evaluations. A very low variation indicates that the actual evaluation does not spread a lot. Finally a value 0 indicates that all criterion evaluations are equal. IF11(Variation) averages all criteria relative output ranges. IF12(Maximum Variation) states the maximum relative output range of all matching criteria.

When dealing with impact factors, also discrete non-weighted aspects have to be considered. One aspect of this are rejections a criterion can cause. Any criterion evaluating to 0 causes the rejection of alternatives. This is a crucial part of the used decision method [16]. These rejections are fully independent of the criterion weights. Taking this into account the rejection potential of a criterion is given by the possible evaluation to 0. So every criterion having a utility function which is able produce a 0 value has a rejection potential. IF13(Rejection Potentail Count) and IF14(Rejection Potential Rate) relate to this rejection potential. IF13 counts the decision criteria which have an alternative rejection potential. IF14 relates this count to all criteria matching the given criterion, resulting in the percentage of decision criteria which have an alternative rejection potential.

Beside the criteria which can cause potential rejections, we are even more interested in the criteria actually rejecting alternatives. A criterion causes a rejection if it actually rejects at least one alternative. In other words, the concrete criterion evaluation for at least one alternative is 0. IF15(Rejection Count) and IF16(Rejection Rate) relate to this actual rejections. IF15 counts the decision criteria actually rejecting alternatives. IF16 relates this count to all criteria matching the given criterion, resulting in the percentage of decision criteria actually rejecting alternatives.

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	Average compound weights
SIF4	Potential	Average compound potential ranges
SIF5	Maximum potential	Maximum compound potential ranges
SIF6	Range	Average compound actual 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 criteria having the potential to reject alternatives.
SIF11	Rejection Potential Rate	Percentage of criteria having the potential to reject alternatives.
SIF12	Rejection Count	Number of criteria actually rejecting alternatives.
SIF13	Rejection Rate	Percentage of decision criteria actually rejecting alternatives
SIF14	Rejection Spread	Percentage of plans affected by a reject out of this set.
SIF15	Reject Count	Number of alternatives rejected.
SIF16	Rejection Rate	Percentage of alternatives rejected.

Table 3.6: Impact factors for *sets* of criteria [25]

We can even go one step further and not only counting the criteria actually rejected values, but the concrete number of alternative rejected by a given criterion, in other words the rejects of a criterion. This is considered in IF17(Reject Count) and IF18(Reject Rate). IF17 counts the number of alternatives actually rejected by a given criterion. IF18 relates this count to the overall number of alternatives evaluated for this criterion, resulting in the percentage of alternatives actually rejected by this criterion.

Impact factors for sets of criteria

The impact factors for sets of criteria mainly base on the factors for single criteria. Only two of the presented factors are new because they can only be applied to criteria sets. Although a strong correlation between single criteria and criteria-set factors exist, aggregating single factors to set factors is not always as straight forward as expected. The different aggregation methods are explained in the following.

Most of the presented impact factors for criteria sets sound similar to the impact factors for criteria set. This could give the impression that criteria set factors can be deduced easily from single criteria factors by just calculating the arithmetic mean over all criteria of the set. But his impression is deceptive. While factors such as spread can be aggregated in a straightforward way be simply building the arithmetic mean, using the same aggregation for other factors would lead to wrong results. For instance, simply summing up the average weights to get the aggregated weight would neglect the fact these averages are calculated based on a partial set of plans. To analyse criteria sets over the entire set of plans, we thus only sum up discounted average weights. The different aggregation methods needed for each criteria set factor are discussed in the following.

Simple arithmetic mean is used to aggregate factors SIF1(Spread), SIF8(Variation) and SIF9(Maximum Variation). Arithmetic mean is appropriate in this case because the aggregation is done at criteria level with no plan-grouping, etc. to be considered.

This is different for factors SIF3(Weight), SIF4(Potential), SIF5(Maximum Potential), SIF6(Range) and SIF7(Maximum Range). As the factor descriptions state they are based on an aggregation on plan level. The different criterion values (e.g. weight) of criteria in the set are summed up on plan level. The results of this are is then used to do the wanted aggregation. For weight, potential and range arithmetic mean is used for aggregating these plan values, for Maximum Potential and Maximum Range the maximum is used as aggregation method. This aggregation on plan level is necessary to get meaningful results. The average weight of all criterion weights in the set is not very helpful. But the weight all criteria together have in a plan is a good statement of the relevance in it. So building the average and the maximum on base of this delivers in much more interesting results. Just to remark, criteria set factors Weight, Potential and Range can also be calculated by just sum up the single criteria factors Discounted Weight, Discounted Potential and Discounted Range. This is because the plan aggregation is still considered in the single criteria discount factors.

Factors SIF10(Rejection Potential count), SIF12(Rejection Rate Count) and SIF15(Reject Count) are calculated by simply summing up the correspondent single criteria factors. Simple building a sum is appropriate here because each criterion rejection is independent from any other. Therefore to measure the rejections of a set of criteria can be calculated by the sum of each criterion rejection.

At the first sight factors SIF11(Rejection Potential Rate), SIF13(Rejection Rate) and SIF16(Rejection Rate) seem to be calculable by building the arithmetic mean of the correspondent single criteria factors. But this assumption is not correct. Consider two criteria C1 and C2 both defined in a criteria set. Let C1 refer to 5 criteria of these 1 produces rejections, and C2 refer to 10 criteria of these 5 produce rejections. Therefore the single criterion impact factor Rejection Rate is for C1 0.2 and for C2 0.5. The arithmetic mean of these two values is 0.35. But if we calculate the impact factor as it should be calculated for set of criteria, we would have to divide the number of criteria caused a rejection by the number of all criteria. This is 6 divided by 15, which results to 0.4. This is the wanted result. The difference in results can be explained by the fact that the number of criteria and criteria which cause rejection need to be considered aggregated through all criteria in the set. This aggregation over the whole set produces a wanted kind of weighting by the criteria counts.

SIF2(Coverage) and SIF13(Rejection Rate) do not relate directly to any of the previous presented single criteria factors and therefore need to be calculated separately. This is because these two factors have a plan correlation which is not covered by our presented single factors.

CHAPTER 4

Tool Support

4.1 Introduction

This chapter introduces two tools which support us at analysing our case-study data.

Data relevant for impact assessment consists of real world preservation plans collected over the last years. All these plans were created in the Preservation Planning Tool Plato and are available for analysing. 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. What is missing up to now are tools which are able to analyse this data and among other things support us at calculating our previously introduced impact factors.

Because of this two Plato-Modules used for data analysing 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 the currently available 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.

4.2 Knowledge Browser

Requirements

The main task of the Knowledge Browser is to let the user gain insight in the internal preservation planning knowledge stored in Plato. Preservation planning aims to find out the best preservation action to take in a given scenario. To evaluate the different preservation actions decision criteria are used. Taking this into account, decision criteria are the decision driver of the preservation planning process and therefore of particular interest. Thus the Knowledge Browser especially focuses on providing criteria usage statistics.

The Knowledge Browser has the following **functional requirements**:

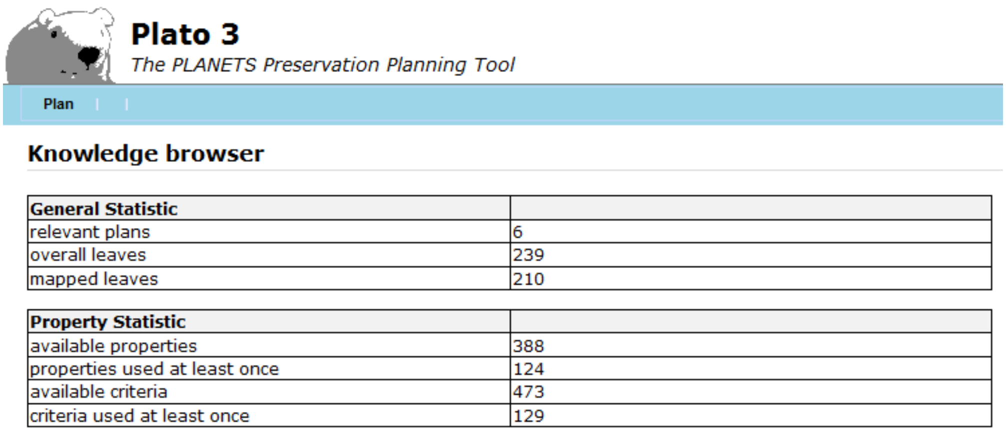
- Provide general information about the data base that is the subject of analysis.
- Provide criteria information
  - List all criteria and their characteristics
  - Show criteria usage statistics
  - Calculate criteria impact factors
- Provide a list of criteria sortable by impact

The **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.
- *Visualization:* Data should be displayed and visualized in a clear, appealing and understanding way.
- *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.

Functionality

Based on the given requirements stated above, the Knowledge Browser displays and visualizes statistics about actual relevant preservation plans and their criteria.

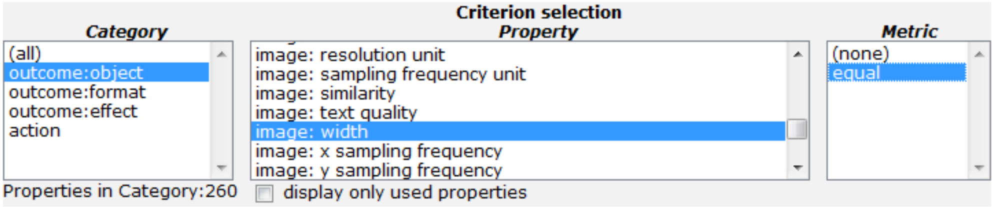


General Statistic	
relevant plans	6
overall leaves	239
mapped leaves	210

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

Figure 4.1: Knowledge Browser General Statistics

As shown in figure 4.1, the general statistics give information about the data which is analysed. It is separated in General Statistics and Property Statistics. General Statistics inform about the current data available in Plato. Property statistics inform about the available properties, the available criteria, and their usage.



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 4.2: Knowledge Browser Criterion Selection

Beside this general information a lot of criterion relevant statistics are available. As shown in figure 4.2 the criterion of interest can be selected by multiple select boxes. Feedback about the characteristics of the selected criterion is given in real time.

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 4.3: Knowledge Browser Criterion Statistics

Dependent on this selection a set of criterion statistics and the previous introduced impact factors for single criteria are calculated and displayed (see figure 4.3).

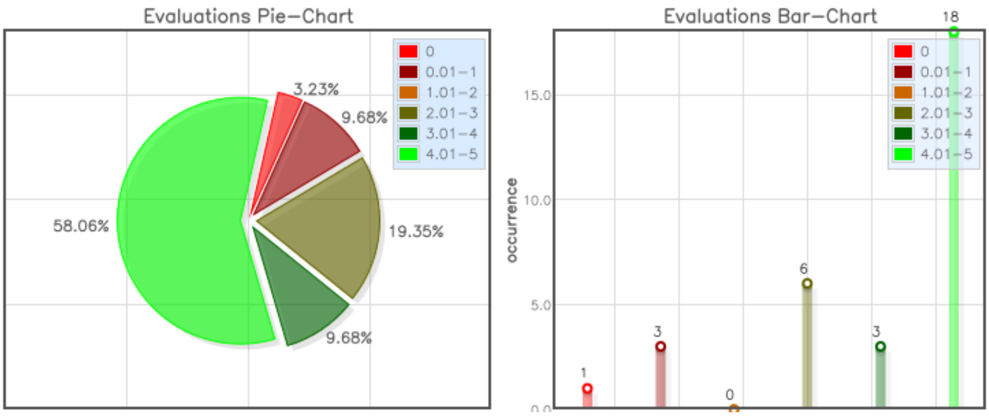


Figure 4.4: Knowledge Browser Criterion Evaluation Charts

Evaluations of criteria are of special interest because their value is used directly in alternative assessment. To better demonstrate the evaluation distribution, evaluations are visualized via a pie and a bar chart. In the actual evaluation model of Plato each criterion is evaluated on a range from 0 and 5. Indeed visualizing each single value in the chart would lead to confuse and unclear results. Furthermore we are not interested to show evaluations in such a detailed way. To be able to visualize the output in a meaningful way possible evaluations are packed into 6 groups (0 which indicates a reject, 0-1, 1-2, 2-3, 3-4 and 4-5). The pie chart is suitable for visualising the evaluation share of each evaluation. The bar chart is is suitable for showing the exact number of evaluation for each potential evaluation. These two presented charts are shown in figure 4.4.

The last part of the criterion statistics refers to transformer or utility functions the selected criterion uses. Utility functions transform measured criterion values to the solution space by evaluating them on an importance scale from 0 to 5. To be able to see what transformer were used for this criterion, all transformer are visualized via a table and a chart (see figure 4.5).

To make requirement coverage complete, the Knowledge Browser finally contains a table called impact factors table which lists all eighteen impact factors for all used criteria. This table is sortable by each column to be able to analyse this data in more detail.

Supporting frameworks and tools

Plato already uses a lot of frameworks and technologies to ensure its current functionality. Mentioning all of these would be beyond the scope of this chapter. Most of these technologies are anyhow now already standard technologies and therefore well known. This chapter only introduces technologies which are not already used in Plato.

To be able to support the Knowledge Browser functionality presented above and to fulfil the non functional requirements usability and visualization mentioned above two new technologies are used. One of it is based on the actually Plato used JSF framework Richfaces [8]. Richfaces

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

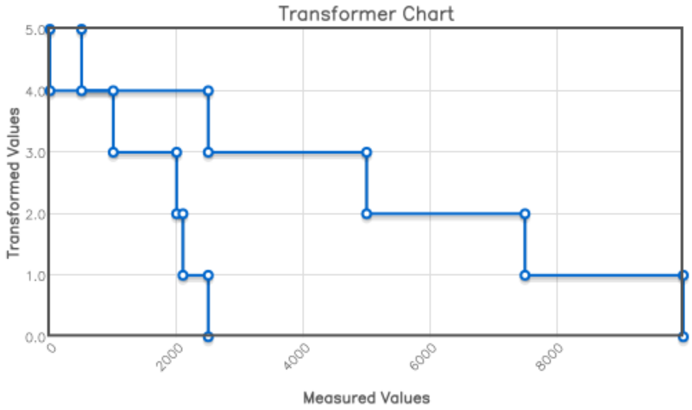


Figure 4.5: Knowledge Browser Criterion Transformer

is already used by Plato but the usage of its AJAX functionality is new. AJAX functionality is required for dynamic refill and re-rendering of the criterion select-boxes (see figure 4.2). When a category is selected the displayed properties change resulting in a refill and reload of the items of the properties select box. A reload of the properties select box is also required if the *display only used properties* select box is checked or unchecked. The same reload procedure effects the metrics select box. Displayed metrics depend on the selected property and therefore also have to be dynamically reloaded. Based on the overall selected criterion other elements like criterion descriptions, charts, etc. have to be reloaded. To support this dynamic reloading of specific page elements in real time AJAX has to be used.

To visualize all criterion evaluations the JavaServer Faces JSF chart library jsflot is used [4]. Jsflot was preferred over chart libraries based on Apaches MyFaces [1] or Primefaces [7] because we wanted to use an independent chart library easy to plug in in current environment without causing any dependency problems. Plato currently uses Richfaces and therefore adding any new JSF component library in addition has a high conflict potential. Despite the easy environment integration of jsflot it additionally does not depend on special browser features or plug-ins (flash, silverlight, etc.) like the most of the currently available JSF chart libraries. It is solely based on JavaScript and therefore renders in standard browsers without problems.

Fetching and analysing data

The Knowledge Browser makes preservation planning knowledge stored in Plato accessible to the user. Accordingly the task of the Knowledge Browser is to fetch the stored data, analyse and process it and present it to the user in an appropriate way. Data fetching takes a long time if a lot of data to analyse is available in Plato. On the other hand data processing and presenting is not a time consuming task and can be done in little time. Data fetching is the bottleneck of the analysing process which indeed caused problems in the first implemented Knowledge Browser version.

Plato preservation planning criteria are organized in a hierarchical structuring consisting of nodes used for structuring and leaves representing decision criteria. The data relevant for analysis are mainly the plan leaves representing the conceptual decision criteria of a plan because they contain all criteria relevant data. Nodes are mainly used as back-reference to the relating plan and to calculate the total weight of a criterion which is defined as the multiplication of each node/criterion weight from the root to the wanted criterion. Consequently, taking into account the current Plato data model we need to fetch the whole decision tree (including plan leaves and nodes) for each plan to be able to do our calculations. Fetching data this way for the first time resulted in a waiting time of 14.3 seconds, at only having 15 plans stored in the database. Taking into account that much more plans will be analysed in future this waiting time is not acceptable and need to be shortened. This is also necessary to fulfil the performance requirements stated above.

To identify the long waiting time cause we installed JBoss Profiler [3] which helped us to identify the bottleneck more concretely. The Hibernate mapping used in plan leaf and plan node entity is defined in a way, every time a plan leaf or a plan node is fetched from database the whole criteria tree is fetched from database as well. This comes from the defined parent and child relations. Fetching ten plan leaves therefore results in fetching the whole plan tree ten times. This recursion is the main cause for the long waiting times.

To solve this problem changing the actual used data model of Plato is not an option because it was constructed this way to support the main functionality of plan creation. The Knowledge Browser is just an additional module which should not affect the main task of Plato in a negative way. Taking this into account we created a view on plan leaf table which includes only data and references to other tables which we need for calculating our statistics. Doing this also removes the parent and children relations which causes the time consuming recursions. Using this view do not touch the actual data model of Plato but instead introduces a new class VPlanLeaf which maps to the just introduced plan leaf view and is used as base for all further calculations. Although the plan leaf itself contains most of the data needed for criteria analysis two things needed for calculation cannot be extracted just from the plan leaves. These are the corresponding root node of this leaf which references the corresponding plan, and the total weight which can only be calculated taking into account the nodes on the way to the root node. Both of this properties relate to the parent nodes of this leaf. Adding such a reference to the view would solve this problem. The problem of this solution is that it would again introduce the recursion problem we are trying to get over. To circumvent this recursion problem when pulling out one entry, we created two SQL stored functions which extract the root node and the total weight for each plan leaf on database level. This two results are just added to the plan leaf view as simple value columns not as references. Doing the necessary recursion on database level prevents the recursion on application level which caused this long waiting times.

Applying the presented problem solution resulted in a performance enhancement by factor 5. Fetching statistic relevant data via the view for the above mentioned 15 plans now only takes about 3 seconds.

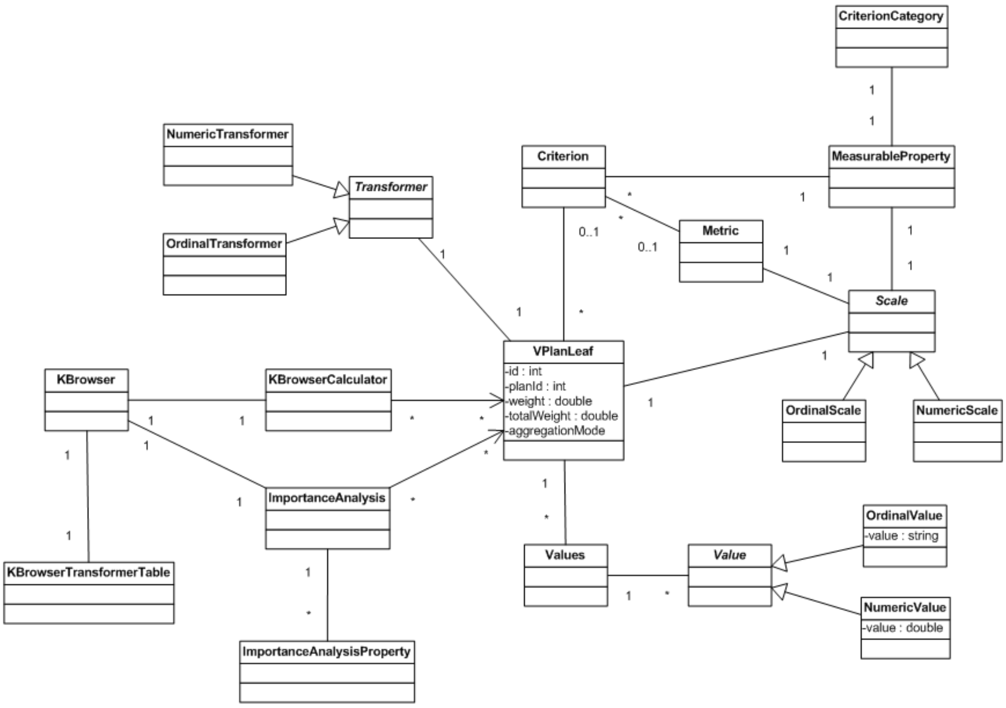


Figure 4.6: Knowledge Browser class diagram

Architecture

The presented view VPlanLeaves containing all leaves respectively decision criteria is used as starting point to calculate relevant Knowledge Browser statistics. This view contains references to all calculation relevant data entities like Transformer, Criterion, Values, etc. The class diagram of the Knowledge Browser is presented in figure 4.6.

The classes KBrowser, KBrowserCalculator, KBrowserTransformerTable, ImportanceAnalysis and ImportanceAnalysisProperty are used for processing and visualization of the data provided through VPlanLeaf and its related classes. Class KBrowser is the backing bean of the Knowledge Browser view. It is responsible for displaying and visualizing data, as well as handling user selection events and re-rendering accordant parts of the page. KBrowserCalculator class is responsible for almost all impact factor calculations. KBrowserTransformerTable is a helper class used to display the transformer table and transformer chart. Finally ImportanceAnalysis and ImportanceAnalysisProperties are helper classes used to visualize the impact factors table.

Having examined the purpose of each class we now take a more detailed look at KBrowserCalculator class which is the calculation module of KBrowser and encapsulates the whole calculation logic. At KBrowser start-up all VPlanLeaves are fetched from database and passed to the KBrowserCalculator. The calculator uses these leaves for the calculation of all factors. The once passed leaves are stored by the calculator in session scope for performance reasons. This session caching prevents from unnecessary repeated database access. When no specific criterion is selected by the user all plan leaves are used for factor calculation. When a specific criterion is selected, the matching leaves are filtered first and then only these are taken into account for factor calculation. The impact factor calculation is based on the single criteria impact factors presented in chapter 3.3.

4.3 Criteria Hierarchies Tool

Requirements

The main task of the Criteria Hierarchies Tool is to support the user at discovering and analysing criteria set impacts.

Functional requirements:

- Define criteria sets
- Provide impact information for each criteria set
- Compare different criteria set impacts

Non-functional requirements:

- *Usability:* Easy and intuitive definition of criteria sets
- *Visualization:* Clear arranged impact factor output to make it easier to draw conclusions.
- *Non-obtrusive real-time analysis:* Direct analysis of the productive database without making changes in the object model.

Supporting frameworks and tools

Just like in Knowledge Browser, Criteria Hierarchies Tool also uses the previously introduced Richfaces AJAX functionalities to make assigning criteria to leaves more user friendly.

Functionality

Based on the given requirements stated above the Criteria Hierarchies Tool allows to define criteria sets in the form of criteria hierarchies. Aligning criteria in a hierarchical structure has the advantage that each hierarchical level itself contains a set of criteria which can be analysed on its own. Comparing the impact factors of different hierarchical levels can led to new insights in the criteria set impact distribution.

The Criteria Hierarchies Tool allows the definition of criteria hierarchies. As shown in figure 4.7 the user can align wanted criteria in a tree structure. This structure consists of structural nodes containing sets of criteria and leaves representing criteria. Next to each structural node criteria set impact factors are displayed which are calculated by the aggregation of all criteria of this hierarchical level. This is especially useful to asses impacts of parts of the hierarchy criteria. Taking this into account the root node displays the impact factors for all criteria used in the hierarchy.

Figure 4.7: Criteria Hierarchies Tool Criteria Tree

Criteria Hierarchies Summary

Name	size	SIF1	SIF2	SIF3	SIF4	SIF5	SIF6	SIF7	SIF8	SIF9	SIF10	SIF11	SIF12	SIF13	SIF14	SIF15	SIF16
outcome format	8	20,83%	66,67%	0,1056	0,386	1,56	0,3135	1,512	0,7062	0,8	4	40%	2	20%	33,33%	5	11,11%
action	8	27,08%	66,67%	0,0557	0,274	0,8312	0,059	0,2372	0,125	0,2623	7	53,85%	0	0%	0%	0	0%
outcome object TIP	5	56,67%	100%	0,0911	0,4169	1,0938	0,2984	1,0938	0,6373	0,8	11	64,71%	5	29,41%	33,33%	5	5,49%

Figure 4.8: Criteria Hierarchies Tool Hierarchy Comparison

Users are not only interested in comparing impact factors within one criteria hierarchy, they also want to compare criteria hierarchies against each other. For this reason the criteria summary was introduced which is shown in figure 4.8. It lists all defined criteria hierarchies with the number of included criteria and all impact factors in a clear table structure for easy comparison.

Architecture

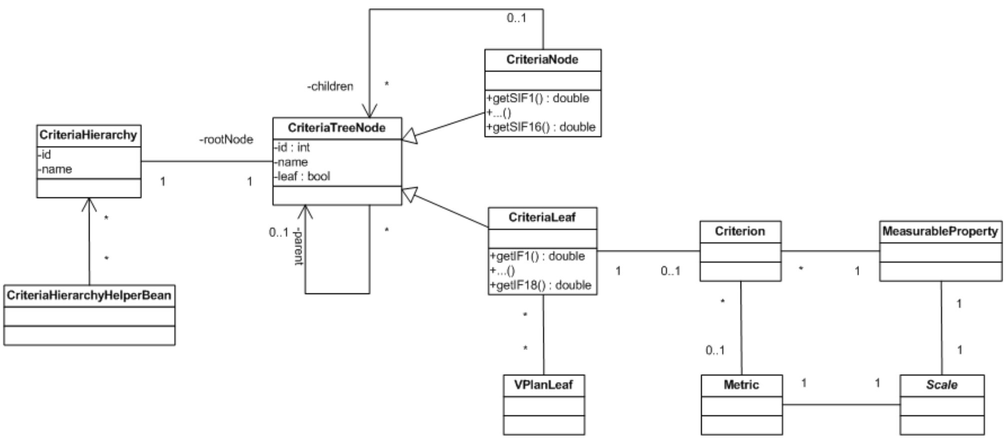


Figure 4.9: Criteria Hierarchies Tool class diagram

The model used for all of the presented tasks is showed in figure 4.9. It is used for creating and evaluating hierarchical criteria structures. The criteria tree showed in figure 4.7 is based on this model. A Criteria consist of CriteriaTreeNodes which itself consist of CriteriaNodes and CriteriaLeaves. A CriteriaLeaf requires an assigned Criterion to make it usable for impact assessment. At attaching a specific criterion to a CriteriaLeaf all VPlanLeaves matching this criterion are assigned to it. These VPlanLeaves are then used to directly calculate the single criteria impact factors. CriteriaNodes are able to calculate their criteria set impact factors because based on the given tree structure they know their underlying CriteraLeaves and therefore their associated VPlanLeaves. This data is finally aggregated resulting in criteria set impact factors. Single criteria and criteria set impact factors are calculated according the rules presented in chapter 3.3.

CriteriaHierarchyHelperBean is the backing bean for all three Criteria Hiearachy relevant views and responsible to support them at fulfilling their duties.

CHAPTER 5

Results and Discussion

5.1 Data to analyse

All data relevant for our analysis is available and enriched with semantic to be automatically processable. Tools were created to analyse this data based on the impact factors defined in chapter 3.3. This section presents the results of this impact factor analysis and discusses its outcome.

Data relevant for our analysis consist of 14 relevant Preservation Plans presented in table 5.1. These plans were created by trusted institutions and span several different preservation scenarios. Overall these plans contain 631 decision criteria whereas 583 were mapped to uniquely identified properties and therefore can be analysed in detail. The other 48 not mapped criteria only occur in one plan, have a rejection potential of 0 and are therefore not further considered. 367 different properties were used to map all these criteria.

No	Object Type	Object Format	Institution Type
1	Databases	MS Access	Archive
2	Documents	Word Perfect	Archive
3	Documents	PDF	Library
4	Documents	PDF	Library
5	Documents	PDF	Research
6	Documents	PDF	Archive
7	Images	TIFF-6	Library
8	Images	TIFF-6	Library
9	Images	TIFF-5	Library
10	Images	NEF	Archive
11	Images	different raw image file formats	Research
12	Images	GIF	Library
13	Video Games	ROMs of SNES video games	Research
14	Video Games	Media images of floppies and CD-ROMs	Research

Table 5.1: Relevant Case-Study data

5.2 Gain insight into preservation planning criteria

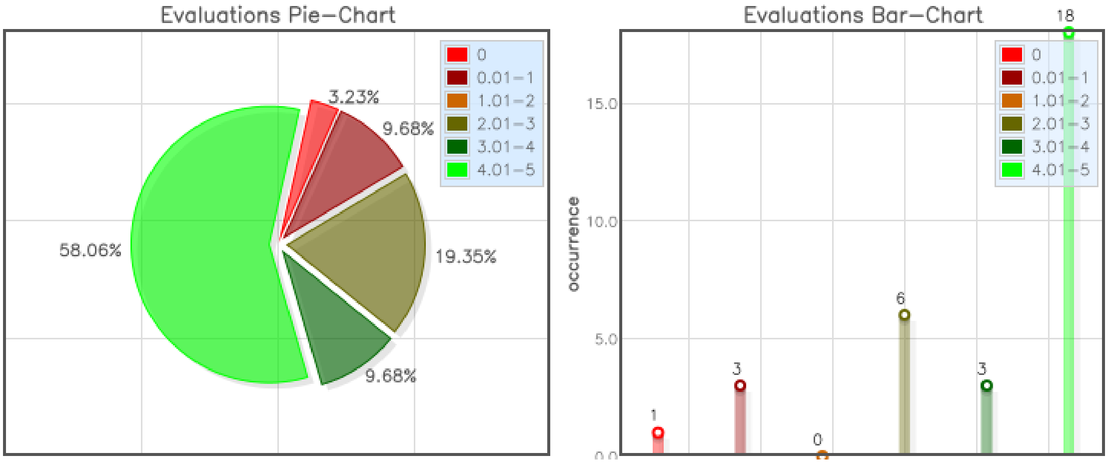
In this section we take a look at the analysing results of the above presented data and discuss what new insight into Preservation Planning criteria can be gained based on this. Analysing this data across all different scenarios does not make sense. A scenario can be described in general as projected course of action. In preservation planning especially similar courses of action are interesting for aggregation concerns. Take for example a responsible preservation planner who wants to create a preservation plan for his collection of digital images. To define the decision criteria for this plan he is for guidance reasons interested in the criteria which had the most impact in similar scenarios. He is not interested in criteria of other plans relating to documents or video games or even all criteria in common. He is only interested in criteria which correspond to his specific scenario. At the moment a scenario is only defined by the content type to preserve because this factor constrains our present data in a sufficient way. As more data evolves, one could think about more constraining factors to be able to define the projected course of action more precisely. The results presented in this chapter relate to the image preservation scenario. This scenario is selected because it is based on the most representative set of data available. It is based on 6 plans consisting of 239 criteria whereas 210 were mapped to uniquely identified properties. 129 different properties were used to map these criteria.

Single criteria impact

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 5.1: *Format Standardization* impact factors

When focusing on single criteria our previously presented tool Knowledge Browser enables us to analyse single criteria impact. Thus this tool allows us to select a desired criterion and show selective statistics about it. Figure 5.1 shows the impact factors of the outcome format criterion standardization. Format standardization is evaluated on an ordinal scale with the possible values *none*, *defacto standard* or *international standard*. Looking at the presented impact factors several observation can be made. It has a spread of 100 percent. It has a weight of 0.0238 which is identical to the discounted weight because this criterion is used in every plan. Its Potential and its Range are nearly the same indicating a high variation of this criterion which can be seen also in its variation of 71.3 percent. In 2 plans (which is one third of all plans) this criterion has potential to knock out alternatives but only produces concrete rejection in one plan. It also only rejects one alternative.



Ordinal Transformer	0.0	0.5	1.0	3.0	4.0	5.0
mapping				none	de facto standard	international standard
mapping			none		de facto standard	international standard
mapping		none			de facto standard	international standard
mapping	none				de facto standard	international standard
mapping			none	de facto standard		international standard
mapping	none				de facto standard	international standard

Figure 5.2: *Format Standardization* visualization

Figure 5.2 shows the transformer mappings of *format standardization* and its resulting evaluations. The frequency distribution of the criterion evaluation shows that in more than 50 percent of all evaluations result in the maximum possible score. The other evaluations are uniformly distributed over all other possible values. In particular interesting are the actual rejections which result in 3.23 percent of all values. Taking a look at the transformer table it can be seen that in two plans rejections are possible - this is indicated by the mapping to 0. It can be also seen that the three possible values *none*, *defacto standard* and *international standard* follow a strict ordering. *None* is in all cases the worst value and even allows rejections. *de facto standard* is always some compromising value between and finally *international standard* is always the best value and even always connected to the highest possible score.

This gained knowledge about single criteria sheds light into decision criteria usage in digital preservation. This knowledge can be further used to improve the decision making process, by for example supporting the responsible planner at criteria definition stage. This support can be for example to support proactive recommendation of utility settings, weights, etc.

This single criteria analysis done for every used criteria results in the impact factors table presented in figure 5.3. Only criteria which are used in at least one third of the plans are listed. Taking a look at this table does not reveal insight into the overall impact of all used criteria because too much data is present whose meaning cannot be seen directly. All of these factors focus on different impact aspects so I have to know what I want to know to focus on the right factors. Take for example the outcome effect criterion archival storage which assesses the effecting archival storage costs after applying a preservation action. It has an enourmous large actual output range of 0.665 which is three times as large as the next bigger Range. When looking at the impact on the final score of an alternative it is clear that this criterion has a huge impact rate. But no rejects are possible. Concerning the question about evaluation criticality of a criterion - meaning how critical is it if this criterion is evaluated wrong and therefore results in a wrong score - Range is not the only thing to care about. For example the outcome format criterion *nr of free tools (not open source)* has in comparison only a tiny Range of 0.031. But because it can cause rejections a wrong evaluation can cause even worse results than archival storage. Thus it can be seen that different questions have to consider different factors for answering them. Therefore based on these presented impact factors conclusions cannot be drawn directly. We first have to think about the questions we have, then consider which factors are indicators for this questions and then we can start detailed analysing. The approach of tackling concrete questions based on our presented impact factors is discussed in chapter 5.3.

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 5.3: Single criteria impact factors table [25]

Criteria-set Impact

Beside the impact of single criteria we are also interested in identifying criteria set impact. This can be used for forming criteria groups of interest like criteria categories or somehow coherent criteria, and analysing their accumulated impact. Accumulated impact is of interest because we are interested in what impact all action criteria have together, in comparison to all format criteria. The Criteria Hierarchy tool provides us with this functionality. Figure 5.4 shows all defined criteria sets and their impact factors. Criteria sets have been defined for all criteria categories *outcome format*, *outcome object*, *outcome effect* and *action*. Outcome Object is represented by its subclasses *Representation Instance Criteria*, *Information Criteria* and *Transformation Information Criteria*. Because image similarity criteria are of high interest at preserving images an own category was created for them. Action is further separated in its main components.

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	0,451	0,956	0,336	0,428	17	36,17%	6	12,77%	50%	10	4,03%
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	0,028	0,129	0,188	0,188	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	0,641	1,25	0,179	0,256	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	1,92%
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%	2	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%	3	0,58%
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	2,94%
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	16%
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%	6	1%

Figure 5.4: Criteria set impact factors table [25]

In the following we discuss the observations which can be made based on the presented data set. **format criteria** are used in every plan and have a spread of 25.27%. Their compound weight is 18.3%, a potential of 0.812 and a Range of 0.451. 36.17% of all format criteria have the potential to knock out alternatives but only 12.77% actually reject alternatives. What is especiall interestin to see here is that every second plan is affected by rejections caused by format criteria.

When looking at **Action Criteria** *maintainability*, *usability* and *reliability* factors have by far the lowest Potential and the lowest Range. They nearly have no variation and do not even have the potential to reject alternatives. Those are all indicators for a low impact on the final decision. *Performance efficiency* and *functional completeness* action criteria have the potential to reject alternatives which in the end never actually happened. Especially performance efficiency criteria are considered critical by the responsible planners because 80% of all these criteria have rejection potential. The resulting actual rejects of 0 show that the performance of no preservation tool was unacceptable. However tool performances were diverging indicated by an avaerge Variation of 0.228. A high plan coverage of 80% shows that performance and functionality action concerns were considered in nearly all plans. *Action portability criteria* are conidered in every plan. *Business factors* are have by far the highest Potential and Range indicating a high effect on the final score and therefore decision. Additionally 85% of these criteria have the potential to reject values indicating a high criticallity. As seen before similar for other action critiera, business concerns also do not produces no single reject. *All action criteria* together have a weight of 0.314, a Potential of 1.415 and the Range of 0.641. Those are the highes values mesured for all investigated criteria sets and indicate a high influence on the final score.

Additionally 50% of all action criteria have the potential to reject alternatives. Despite this fact not even one reject was caused by action criteria.

Outcome Object Criteria can be separated into *Representation Instance Criteria*, *Information Criteria* and *Transformation Information Criteria*. *Information Criteria* are used sparsely indicated by a Coverage of 33.33% and a Spread of only 1.17%. Despite this fact the Range of 0.152 produced by these critera is unexpeted high. This indicates that those criteria are not used frequently but if they are used in a plan they have a relatively high impact on the final score. Additionally those criteria rejected alternatives in 50% of their use cases also indicating their relevance when they are used. *Representation Instance Criteria* are used frequently indicated by a Spread of 18.06% and a Coverage of 100%. Those criteria have a higher potential than *Information Criteria* which can be explained by the fact that those criteria mainly check if the representation of the data is is accurate. Faults in data representation of course have a huge negative impact on the final decision. This can also be seen by a relatively high Potential Rejection Rrate of 38.47%. On the other hand the low variation of 0.049 and the the Rejection Count of only 1 show that representation requirements were met in most of our case studies. *Transformation Information Criteria* are the outcome object criteria producing the highest influence on the final score. This can be seen by a Potential of 0.817 and a Range of 0.363. Such high scores are not a surprising fact because those criteria test the fulfilment of the main goal of digital preservation, which is to sustain the authenticity of the original object to preserve. Additional significance factors are a high Weight of 18.8% and a high variation of 0.58. Every third plan was affected by a reject out of those criteria indicating that in every third plan at least one preservation action did not preserve the original object accurately. *All Outcome Object criteria together* have a weight of 27.4%, a potential of 1.205 and a Range of 0.578. These values are the second largest and are slightly below the ones of all action criteria together. The big difference between the impact of those two criteria sets is that action criteria never produce rejects and outcome object criteria rejects affect one third of all plans. This fact supposes that althought action criteria have a higher Range and a higher Potential, outcome object criteria have a larger impact on final decisions.

One set of criteria we are especially interested in this image preservation scenario are **Image Similarity Criteria** which focus on different approaches to assess image similarity. Even though these set contains only 12 criteria it can be seen that the assessment of image similarity plays an important role in preserving digital images. This conclusion can be drawn by looking at the range of 0.13 which is one third of the overall *Transformation Information Criteria* range. The highest Rejection Potential Rate of 58.33% and a hight Rejection Rate of 16.76% support this conclusion.

Outcome Effect criteria are only used by 50% of all case-study plans. Despite this fact they have a extremely high Potential and Range of 0.395 and 0.309. The high Potential indicates that if these criteria are used they have a high influence on the final score. This high influence on the final score is supported by the fact that Potential and Range lie close together. This indicates nothing else than a high Variation. It has a value of 0.583 and is the highest for all investigated criteria sets. The hight relevance of *Outcome Effect Criteria* is also be seen by the highest discovered Reject Rate of 16.67% showing that a large percentage of alternatives were rejected by those criteria.

Summing this up we have discussed the impact factors for all criteria sets. All criteria sets based on the main criteria categories *outcome format*, *outcome object*, *outcome effect* and *action* were identified as fundamental for finding the best decision to take. This conclusion supports the current taxonomy used to structure criteria and its relating properties. Even the based on its few criteria count often ridiculed *outcome object* category has been identified as significant. We further analysed outcome object and action criteria in more detail to gain additional insight into their containing criteria. For action criteria business related criteria are especially of interest. Maintainability, usability and reliability concerns seem to be not so interesting at all. For Outcome Object criteria, transformation information criteria seem to be the main decision driver. Especially in this scenario criteria trying to prove image similarity have been identifies as significant.

5.3 Answering concrete questions

As stated in the previous chapter, our defined impact factors help us to gain insight into criteria usage. Taking those numerous impact factor results into account (see figure 5.3) it is nonetheless not possible to directly answer questions like what is the minimal representative set of criteria for a given scenario, what criteria are the most critical, etc. The available amount of data is to overwhelming and covers too many impact aspects that direct conclusions can be drawn intuitively. In addition different impact factors are relevant for answering different questions. Thus in this chapter we try to identify the impact factors of relevance to answer several concrete questions. We follow the Goal Question Metric (GQM) paradigm to do this. GQM is is a common method mainly used in Software Engineering to define a measurement model. This measurement model is for better understandability defined on three levels: The conceptual level (goal), the operational level (question) and the quantitative level (metric). The abstract defined goal is splitted down into questions which are finally quantified by concrete measurable metrics. In our case our impact factors serve as metrics. More details about Goal Question Metric can be found under [32].

Minimal representative set

As figure 5.5 reveals the first goal we are interested in is to **identify the minimal representative set of criteria for a given Scenario S**. Achieving this goal can be useful for several reasons. First of all the minimal representative set of criteria can be used to support the user at criteria creation. By making sure the user does not forget representative criteria maximizes trustworthy. Additionally efficiency at evaluation stage can be improved by suggesting the user to leave out not representative criteria. On the other hand this representative set can be used to improve automation of the preservation planning process. Criteria definition is at the moment a 100 percent manual process and therefore forms a huge bottleneck. Having a representative set of criteria for several scenarios would enable us to partially automate this step.

To identify the minimal representative set of criteria two main questions have to be answered: *Should we add criterion X in scenario S?* and *Should we neglect criterion X in scenario S?*. These two questions represent the basic add and remove operations required to build a set.

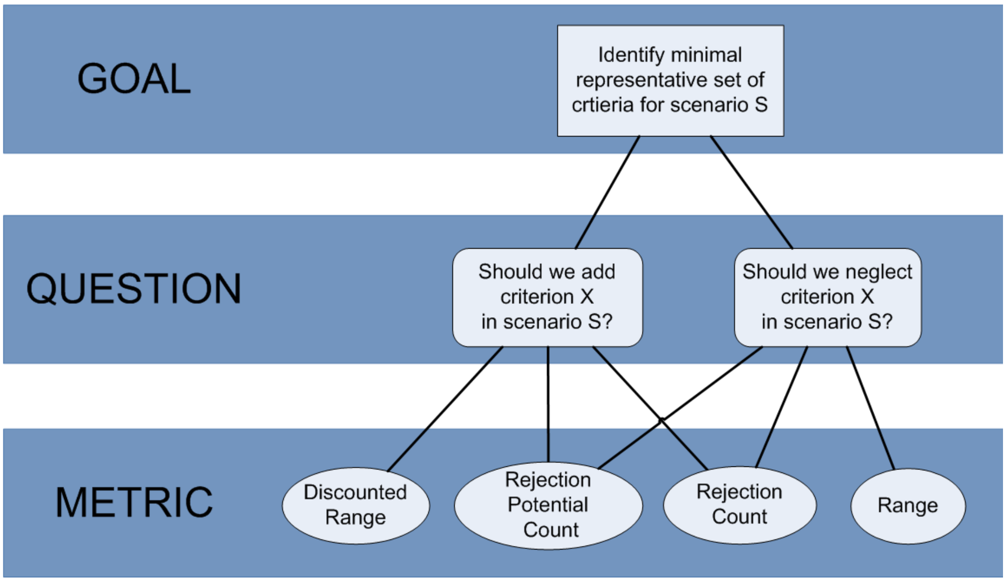


Figure 5.5: GQM Goal: Minimal representative set

To answer these questions the following impact factors were identified as appropriate metrics. *Discounted Range* seems important to answer the question if a criterion should be added. This factor is is a good trade-off between Spread and Range. The produced Range of a criterion is of relevance because it indicates its quota on the final score. But if a criterion has a hight Range but is used only very infrequently it seems to be not so important at all. Therefore the less often a criterion is used the more discounting the Range should experience. Exactly this is represented by Discounted Range. *Range* seems important to answer the question if a criterion should be neglected. Range is used instead of Discounted Range here because at question time this criterion in already in the minimal set and now we have to decide if it should be neglected or not. Thus we are simply interested in the Range this criterion produces when it was used and do not want a discounting based on its usage frequency. *Potential Rejection Count* is considered because it is important to know how often this criterion had the potential to reject alternatives. This is some criticality factor given by the responsible planner. *Rejection Count* is used to see how many rejects the criterion finally produced regarding its potential rejects. It can be seen as indicator how critical the criterion finally was. Thus the relation between Potential rejections and actually happened rejections is especially interesting. Taking these four factors into account it should be possible to identify the minimal representative set of criteria for a given scenario.

Criterion criticality

Another interesting goal to achieve is to **identify critical criteria in a given scenario**. The GQM tree of this goal is shown in figure 5.6. Critical criteria are criteria we especially have to be aware of because they have a high potential to change the final decision. Thus those criteria have to be handled with care at utility function definition, at aggregation mode selection and at evaluation.

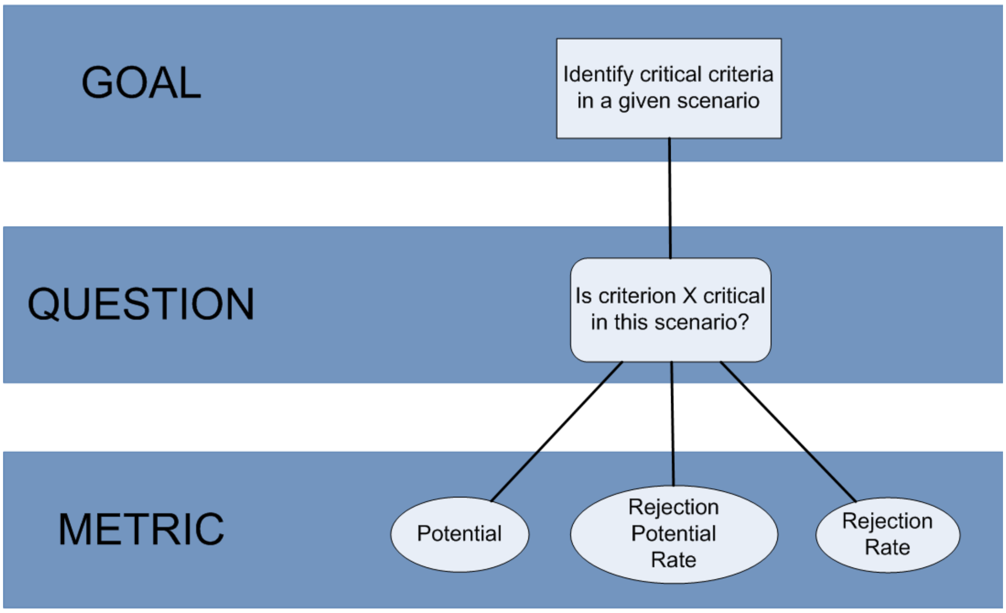


Figure 5.6: GQM Goal: Criterion criticality

Especially critical criteria have to be evaluated with care because an inaccurate evaluation has the potential to be responsible for a complete different final decision. Identify these criteria enables us to advice responsible planners to be especially careful at using those criteria. To answer the relating question *is criterion X critical in this scenario* the following relevant metrics have been identified. In this case we do not consider the Range of a criterion because we are not interested in the actual output range of a criterion but in the potential output range a criterion can produce. This identifies its criticality. Therefore *Potential* is used for answering this questions. We additionally do not use Discounted Potential here because the criticality of a criterion does not depend on its usage frequency. Using a criterion more often does not mean it gets more or less critical in a scenario. Another factor to consider is *Rejection Potential Rate*. The more often this criterion was defined with rejection potential, the more often it was seen as discerning by the user. Thus we should carefully consider this factor too. Finally *reject rate* is a indispensable factor because it indicates the actual discerning of this criteria. This is done by identifying how often alternatives were actually rejected and therefore final decisions were changed.

Automate measurements

Finally *identifying automated measurements to implement next* can help us to improve the evaluation stage of the preservation planning process. There already exist automated measurements for several criteria like *image width equal* or *performance efficiency: time per object*. Nevertheless most of the present criteria are not automatically measurable. Therefore manual evaluating has to be performed which is resource and time intensive. Increasing the number of automatically measurable criteria would improve the performance of the preservation planning process

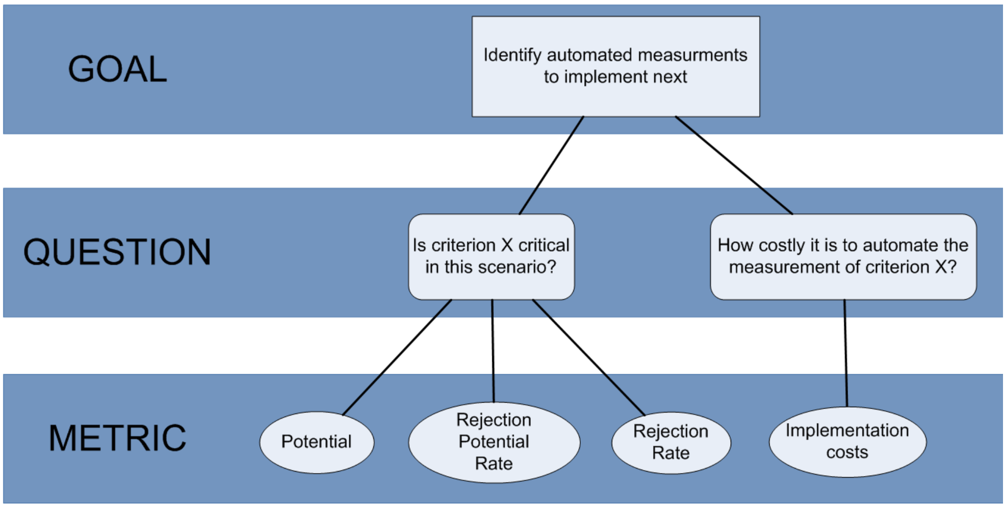


Figure 5.7: GQM Goal: Automate measurements

in and additionally advance its automation because of the less human interaction needed. To achieve the greatest progress in the shortest amount of time we want to identify the criteria which have to be automated next. Consequently we have to answer the questions *How costly it is to automate the measurement of criterion X?* and *How critical is criterion X?*. The more critical a criterion is the more recommendable it is to have an automated measurement for it. Automated measurement should prevent evaluation faults and therefore reduce the risks of wrong evaluation input given by the user. Metrics to assess this question were discussed in the chapter before and are thus not further mentioned here. The other question about automation costs cannot be answered by our given impact factors and thus has to be evaluated by human beings for each criteria. For example measurement like the width of an image are quite cheap because several free tools exist which are able to carry out this measurement. On the other hand measurements like the determining of figure positions in word documents are a much more sophisticated to carry out and therefore have to be rated as very expensive. The automated measurements to be implemented next should be the ones which are very critical and very cheap to implement.

Of course new goals and questions will arise in the future which are not addressed in this chapter. Following the here presented approach it should be possible identify the relevant metrics or impact factors for each arising questions.

CHAPTER 6

Summary and Outlook

6.1 Summary

We introduced an approach to quantitatively assess the impact of preservation planning criteria. This approach is based on analysing significant preservation plans and their criteria collected over the last years. To be able to analyse these plan criteria in an automated way we had to enrich them with semantic. This association of semantic is done by assigning criteria to predefined properties with predefined meanings. A set of 473 properties had to be defined to cover nearly all present criteria with properties. Additionally an appropriate property model had to be constructed for structuring these numerous properties in a reasonable way. For property definition and property model restructuring several standard models like ISO 25010, Library of Congress format evaluation factors, The National Archives file format evaluation guide, PRONOM information model, etc. were consulted to achieve meaningful and enduring results.

After we made the present data ready for automatic processing we defined a set of quantitative impact factors to be able to analyse this data. The two tools Knowledge Browser and Criteria Hierarchy Tool were created to support us at data analysing. The emerging analysis results revealed us new insights into the usage of single and sets of preservation planning criteria and let us answer questions like *which criteria are used most frequently?*, *which criteria have the highest weight assigned?*, *which criteria produce the highest output range?* or *which criteria cause the most rejections or even have the potential to do so?*. In addition we identified the impact factors which can be used to answer concrete questions like *Which criteria form the minimal representative set?* or *Which criteria are critical?*.

6.2 Future Work

During the work on this thesis several new questions showed up which may be topic of further research. Those will be introduced in the following.

In section 5.3 we raised several questions of interest based on our given impact factors like *what is the minimal representative set of criteria for a given scenario?*. Further we identified the impact factors necessary to answer those relating questions. Knowing the impact factors which have be taken into account is a good starting point but too abstract to be able to answer these question in detail. Thus in further research we have to identify how to apply these impact factors to get to concrete answers. Taking as example the minimal representative set problem, we have to define exactly under which conditions a criterion has to be added to this set and under which conditions a criterion has to be removed from this set. Consequently we have to define some kind of applicable rules which answer these questions in detail. Applying these rules to a given set of criteria impact factors should result in concrete answers like the minimal representative set of these criteria.

The three questions stated in section 5.3 are based on three ways to improve the preservation planning process. The questions *identifying the minimal representative set of criteria* and *identifying the critical criteria* in a scenario are aimed to improve the criteria definition stage of the process in terms of automation and user support. The question *what criterion measurements to automate next* is aimed to improve the evaluation stage of the process in terms of automation. Taking the whole preservation planning process into account additional ways to improve this process should be identified based on the gained criteria impact.

All impact factor values presented in this thesis are based on a current and solid base of relevant case-study data. The process of impact assessment is dynamic and extensible and is able to analyse any set of plans as long as their relating criteria are enriched with semantic through property assignment. Thus it makes sense to redo the impact assessment every time new relevant plans arise because the more data is available for analysis the more accurate results emerge.

Bibliography

[1] Apache myfaces. http://myfaces.apache.org, August 2011.

[2] Dbpedia. http://dbpedia.org, August 2011.

[3] Jbosss profiler. http://www.jboss.org/jbossprofiler, September 2011.

[4] jsflot - a javaserver faces jsf chart library. http://code.google.com/p/jsflot, September 2011.

[5] Plato preservation planning tool. http://www.ifs.tuwien.ac.at/dp/plato, February 2011.

[6] Preserv2 - file format registry. http://p2-registry.ecs.soton.ac.uk, May 2011.

[7] Primefaces. http://www.primefaces.org, August 2011.

[8] Richfaces - jsf component framework by jboss. http://www.jboss.org/richfaces, August 2011.

[9] The National Archives. Evaluating your file formats. Technical report, The National Archives, 2011. http://www.nationalarchives.gov.uk/documents/information-management/evaluating-file-formats.pdf.

[10] The National Archives. The technical registry pronom. http://nationalarchives.gov.uk/PRONOM/Default.aspx, May 2011.

[11] Christoph Becker, Guenther Kolar, Josef Kueng, and Andreas Rauber. Preserving interactive multimedia art: A case study in preservation planning. In *Proceedings of the Tenth Conference on Asian Digital Libraries (ICADL'07)*, volume Springer Lecture Notes in Computer Science 4822, pages 257–266. Springer Berlin / Heidelberg, 2007.

[12] Christoph Becker, Hannes Kulovits, Mark Guttenbrunner, Stephan Strodl, Andreas Rauber, and Hans Hofman. Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans. *International Journal on Digital Libraries (IJDL)*, 10(4):133–157, December 2009. http://dx.doi.org/10.1007/s00799-009-0057-1.

[13] Christoph Becker, Hannes Kulovits, and Andreas Rauber. Trustworthy preservation planning with plato. *ERCIM News*, 80:24–25, January 2010.

[14] Christoph Becker and Andreas Rauber. Improving component selection and monitoring with controlled experimentation and automated measurements. *Information and Software Technology*, 52(6):641–655, June 2010.

[15] Christoph Becker and Andreas 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. http://dx.doi.org/10.1002/asi.21527.

[16] Christoph Becker and Andreas Rauber. Preservation decisions: Terms and conditions apply. In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, pages 67–76. ACM, June 2011.

[17] Christoph Becker, Stephan Strodl, Robert Neumayer, Andreas Rauber, Eleonora Nicchiarelli Bettelli, and Max Kaiser. Long-term preservation of electronic theses and dissertations: A case study in preservation planning. In *Proceedings of the 9th Russian Conference on Digital Libraries (RCDL 2007)*, Pereslavl, Russia, October 2007. http://rcdl2007.pereslavl.ru/en/program.shtml.

[18] Valerie Belton and Theodor J. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer, 2002.

[19] Adrian Brown. Pronom 4 information model. Technical Report 1, The National Archives, January 2005. http://www.nationalarchives.gov.uk/aboutapps/fileformat/pdf/pronom_4_info_model.pdf.

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

[21] José Figueira, Salvatore Greco, and Matthias Ehrgott. *Multiple criteria decision analysis: state of the art surveys*. Springer Verlag, 2005.

[22] Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System (OAIS)*, volume 1. CCSDS Secretariat, Program Integration Division (Code M-3), National Aeronautics and Space Administration, Washington, DC 20546, USA, blue book edition, January 2002. http://public.ccsds.org/publications/archive/650x0b1.PDF.

[23] Xavier Franch and Juan Pablo Carvallo. Using quality models in software package selection. *IEEE Software*, 20(1):34–41, January/February 2003.

[24] Salvatore Greco, Jose Rui Figueira, and Matthias Ehrgott. *Trends in Multiple Criteria Decision Analysis*. International Series in Operations Research & Management Science. Springer, 2010.

[25] Markus Hamm and Christoph Becker. Impact assessment of decision criteria in preservation planning. In *Proceedings of the International Conference on Preservation of Digital Objects 2011 (iPres 2011)*, November 2011.

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

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

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

[29] Ralph L. Keeney and Howard Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press, 1993.

[30] Library of Congress. Sustainability of digital formats - planning for library of conress collections. Technical report, Library of Congress, 2011. `http://www.digitalpreservation.gov/formats/`.

[31] Preserv2. Risk analysis - http://nationalarchives.gov.uk/pronom/format/617 (default profile). Technical report, Preserv2, August 2009. `http://p2-registry.ecs.soton.ac.uk/risk_summary.php?format=http://nationalarchives.gov.uk/pronom/Format/617`.

[32] Egon Berghout Rini Van Solingen. *Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development*. McGraw-Hill Inc.,US, 1999.

[33] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis - The Primer*. John Wiley & Sons, Ltd, 2008.

[34] Andrea Saltelli, Stefano Tarantola, Frencesca Campolongo, and Marco Ratto. *Sensitivity Analysis in Practice - A Guide to Assessing Scientific Models*. John Wiley & Sons, Ltd, 2004.

[35] David Tarrant, Steve Hitchcock, and Les Carr. Where the semantic web and web 2.0 meet format risk management: P2 registry. *International Journal of Digital Curation*, 6(1):165–182, 2011.

[36] Evangelos Triantaphyllou and Alfonso Sánchez. A sensitivity analysis approach for some deterministic multi-criteria decision-making methods. *Decision Sciences*, 28(1):151–194, 1997.

[37] UNESCO. UNESCO charter on the preservation of digital heritage. Adopted at the 32nd session of the General Conference of UNESCO, October 17, 2003. `http://portal.unesco.org/ci/en/files/13367/10700115911Charter_en.pdf/Charter_en.pdf`.

[38] W.T.M. Wolters and B. Mareschal. Novel types of sensitivity analysis for additive mcdm methods. *European Journal of Operational Research*, 81(2):281–290, 1995.