EVALUATION OF BIT PRESERVATION STRATEGIES

Eld Zierau & Ulla Bøgvad Kejser
The Royal Library of Denmark
Dep. of Digital Preservation
P.O.BOX 2149
1016 Copenhagen K, Denmark

Hannes Kulovits
Vienna University of Technology
Inst. of SW Tech. & Interactive Sys.
Favoritenstraße 9-11/188/2
A-1040 Wien, Austria

ABSTRACT

This article describes a methodology which supports evaluation of bit preservation strategies for different digital materials. This includes evaluation of alternative bit preservation solutions. The methodology presented uses the preservation planning tool Plato for evaluations, and a BR-ReMS prototype to calculate measures for how well bit preservation requirements are met.

Planning storage of different types of data as part of preservation planning involves classification of the data with regard to requirements on confidentiality, bit safety, availability and costs. Selection of storage parameters is quite complex since e.g. more copies of data means better bit safety, but higher cost and higher risk of compromising confidentiality.

Based on a case study of a bit repository offering differentiated bit preservation solutions, the article will present results of using the methodology to make plans and choices of alternatives for different digital material with different requirements for bit integrity and confidentiality. This study shows that the methodology, including the tools used, is suitable for this purpose.

1 INTRODUCTION

This paper explores how bit preservation strategies can be evaluated against different bit repository solutions. A preservation strategy presents the chosen solution for bit preservation. The bit preservation strategy must ensure that the actual bits remain intact and accessible at all times, and is the starting point for further preservation actions. Functional (logical) preservation, which assures that the data remains understandable through further preservation actions are *not* part of bit preservation.

The research question we want to investigate is how we can evaluate requirements for a bit repository. This concerns e.g. bit safety, confidentiality and cost for alternative bit preservation solutions.

Requirements for bit preservation can be hard to express on the general level. As Rosenthal et al. notes it is a question of risk analysis [5]. We will in this article take an approach where requirements are defined in terms of importance of risk preventions. Formulation of the requirements is primarily based on the ISO 27000 series [2], complimented with analysis of bit safety [4], and own experiences.

Bit preservation implementation is hard in itself, and a lot of the technical and organisation details on the final bit preservation solution can be crucial for how well it fulfils requirements for risk prevention as explained in [6]. The challenge here is to express how different combinations of ways to store and check data copies will meet requirements.

The article presents a methodology which can help in evaluation of bit preservation strategies against choice of bit preservation alternatives. The methodology seeks to separate evaluation of requirements from the complexity of bit preservation in order to make an evaluation more clear and understandable. This is done using a tool which we call: Bit Repository – Requirement Measuring System (BR-ReMS). It is a prototype, which contains the details separated from the requirements. The BR-ReMS results are scores on how well a bit preservation solution prevents different risks.

The methodology uses the preservation planning tool Plato to evaluate how well potential bit preservation strategies meet bit preservation requirements (as a result of the BR-ReMS). Plato is a Planets tool for specification of preservation plans, primarily on logical preservation strategies [1]. In this article we will use it for evaluation of bit preservation strategies only.

In order to investigate the soundness of the methodology, we include three cases of digital material with different requirements for confidentiality and bit safety.

2 METHODOLOGY

The methodology behind our evaluation of bit preservation strategies is based on assumptions on how we can express bit preservation strategies and include requirements, assumptions on parts in bit preservation solutions, and which tools we use for the evaluation.

2.1 Assumptions on Bit Preservation Strategies

We will assume that we can evaluate a bit preservation strategy in terms of evaluating requirements against solutions. This conforms to the definition of requirements that document constraints and influence factors on potential preservation strategies in Plato.

In our case study, the bit repository requirements are assumed to be best formulated in terms of risk prevention. There are many other ways to formulate the requirements, for example at a much more detailed technical and organisation level. It should be noted that the methodology would also apply if another approach were chosen for requirements. The change would only have to be made in the set-up of the BR-ReMS and Plato tools used.

2.2 Assumptions on Bit Preservation Solutions

We will assume that bit preservation solutions can be represented as a solution offered by a conceptual bit repository (BR). A BR is a repository with a technical system managed within organisations with all aspects of an OAIS[1] system as defined in [6].

We will need to make assumptions on how bits are preserved. The assumption is that data must be kept in more copies represented as replicas. Each replica is a copy of the data stored in a pillar. A pillar is defined as a unit, which can be seen and analysed as an individual unit at the abstract level.

Replicas located on different pillars must be coordinated and possibly checked at a general BR system level. This architecture is illustrated in Figure 1. The assumption is only made on the conceptual level. This means that this architecture applies for a Danish National BR under implementation [6], or on a LOCKSS[2] system, or a SAN[3] system with backup.

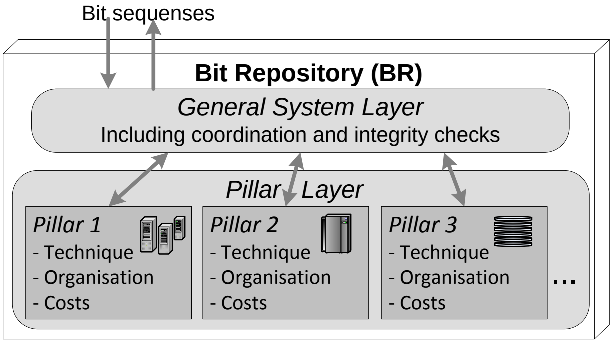

Figure 1. Bit repository with pillars.

Each pillar has different characteristics e.g. the type of media on the technical side, the physical location and procedures for operation on the organisational side, and the costs of using a pillar as basis for a replica. Similarly the general system layer has different characteristics e.g. communication protocol, speed, and bit audit frequency.

For simplicity we assume that bit integrity checks are made on a voting system based on checksums. For example, three replicas participate in a voting, where two replicas agree on a checksum, but the third does not. In this case the third replica will be reported as the faulty one. Voting is based on checksums instead of full comparisons for efficiency reasons.

An additional assumption is that a replica can be a derived replica in form of a checksum. We will call this a checksum replica instead of a full replica which contains a full copy of the data. Checksum replicas are included, since choice of having checksum replicas can increase bit safety at a low cost, but the risk analysis will e.g. depend on its physical location. This is based on Danish experiences explained in [6].

2.3 Using the BR-ReMS and Plato

At the start of this study we intended only to use the Plato tool for evaluation of bit preservation evaluation. However, it quickly became obvious that the specification of a bit preservation strategy and the influence of changing a single characteristic on a pillar were too complex to express directly in Plato.

This lead to the development of the BR-ReMS prototype, which is used to encapsulate the details on different characteristics for parts of the BR, and how they in combination change the measured levels of e.g. bit safety and confidentiality risks. The BR-ReMS produces the results which can be used in evaluation of a bit preservation strategy defined in Plato. This is illustrated in Figure 2.

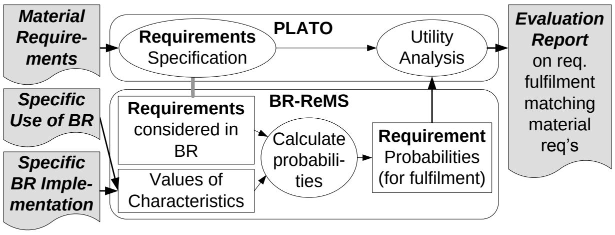
Material Require-ments
PLATO
Requirements Specification
Utility Analysis
Evaluation Report on req. fulfilment matching material req's
Specific Use of BR
Specific BR Imple-mentation
BR-ReMS
Requirements considered in BR
Values of Characteristics
Calculate probabili-ties
Requirement Probabilities (for fulfilment)

Figure 2. BR-ReMS and Plato.

The white square represents specified data whereas the grey squares represent actual input and output. The circles are processes where the arrows give directions of the information flow. The thick grey line indicates that requirements considered are the same.

3 SETUP OF REQUIREMENTS AND TOOLS

In order to understand how the methodology works, we here give a description of the set-up of the tools, as well as the choices made in definition of the requirements.

3.1 Plato

Plato is a preservation planning tool developed within the Planets[4] project and is available to the public in an open source version[5]. It has been developed in order to provide a systematic approach for evaluating potential alternatives for preservation actions and building thoroughly defined, accountable preservation plans for keeping digital content alive over time. The method follows a variation of utility analysis to support multi-criteria decision-making procedures in digital preservation planning. The selection procedure leads to well-documented and transparent decisions.

The applicability and usefulness of the tool has been validated in a series of case studies involving different organisations and digital content such as described in [3]. However, instead of evaluating migration tool with respect to the requirements, we here use the approach to analyse the results of the BR-ReMS for alternative bit preservation solutions. The results of the BR-ReMS are analysed and aggregated, corresponding to evaluation of the bit preservation strategy. Further details on this process can be found in [1,3].

[1] OAIS (Open Archival Information System). 2002. ISO 14721:2003.

[2] See http://lockss.stanford.edu/lockss/Home

[3] See http://en.wikipedia.org/wiki/Storage_area_network

[4] Preservation and Long-term Access through NETworked Services (Planets). See http://www.planets-project.eu/

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

3.2 The BR-ReMS Prototype

The BR-ReMS prototype is developed using Microsoft Access 2003. The set-up for specific cases is based on requirement definitions and definitions of different characteristics. A requirement definition includes definition of a function which calculates to which degree the requirement is met for different BR solutions. The calculations are based on the specified characteristics. This is exemplified in Figure 3, where the boxes with dashed lines are templates, and their use is indicated by thick grey lines.

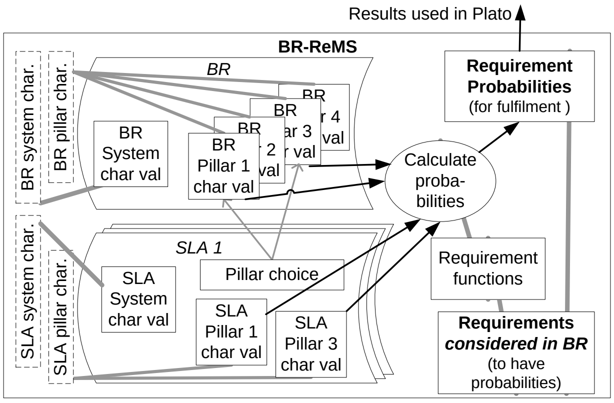

Figure 3. The BR-ReMS prototype.

As illustrated in Figure 3 there are different types of characteristics. There are the BR characteristics which are predefined by the actual 'BR implementation' (see Figure 2). And there are the service level agreement (SLA) characteristics, which are defined by individual SLAs for 'specific use of the BR' (see Figure 2). A SLA is defined as the agreement of level of service between the unit responsible for the BR and a user preserving bits in the BR, e.g. on which pillars the replicas are placed, and for each pillar, whether it is checksum or full replica. Note that we only talk about a conceptual SLA for a conceptual BR, i.e. there are no requirements to degree of formality and whether the SLA is involves several organisations operating different parts of the BR.

The BR characteristics are divided into BR general system characteristics (e.g. for transmission of data or coordination ensuring hardware/media migrations are not performed at the same time), and BR pillar characteristics for the individual pillars (e.g. hardware type, or characteristics related to natural disasters). In the same way the SLA characteristics are divided into SLA general system characteristics (e.g. bit audit frequency) and SLA pillar characteristics (e.g. digital objects are checksum replicas or full replicas).

The characteristics are defined in two steps. Firstly, the characteristic itself is defined. Secondly, the value(s) of the characteristic are defined for the different parts of the BR and individual SLAs.

Requirements are defined along with their functions. These functions can be quite complex and depend on different types of characteristics. In order to ease the calculation general functions are introduced for each pillar characteristic (both BR and SLA pillar characteristic) to be calculated across the pillars selected in a SLA. Some sub-functions also go across pillar characteristics and general system characteristics, as for example comparing frequency of bit audits with Mean-Time-To-Failure on the different media. For such purposes intermediate result characteristics are introduced which can be used in more complex calculations. Note that calculation over more pillars will work differently depending on the requirement it belongs to. For example, in calculation of bit safety requirements, adding a replica will always lower the risk of loosing bits. On the other hand in calculation of confidentiality requirements, the general rule is that adding an extra full replica will mean higher risk for lack of confidentiality.

The setup of the functions is still on a prototype level at this stage. The functions could be better described and tuned by use of more complex calculations e.g. using statistically models for error occurrence etc.

Since the details on calculation of how requirements are met are important, the BR-ReMS also offers reporting on definition of values of characteristics and definition of function used. Such reports would be input for a thorough evaluation of a bit preservation strategy or to audit actual implementation of BR parts.

3.3 Requirements used in Plato and in BR-ReMS

The definition of the requirements represented in the SLA will express the bit preservation strategy to be evaluated in Plato, as well as the requirements that the BR-ReMS produces results for. That means the requirements must be specified in both the BR-ReMS and Plato.

We will here base requirements on the ISO 27000 series [2], as far as possible. The reason for this choice is that the ISO standard is a commonly used standard in repositories. It includes confidentiality (ensuring that information is accessible only to those authorised to have access) and integrity (safeguarding the accuracy and completeness of information and processing methods) as some of the main risk areas for information security. These are also the aspects that we have chosen to focus on in this article. This choice is mainly made in order to narrow the scope, but also because of the way that adding a full replica influences fulfilling these requirements in different ways. The availability aspect, as well as organisational aspects and cost, are just as important and can be included at a later stage using the same technique as for bit integrity and confidentiality. The organisational aspects could also use the criteria from the Trustworthy Repositories Audit & Certification (TRAC)[1] for disposition of requirements and relevant BR characteristics.

Looking closer at integrity, we find that authenticity is not relevant in connection with a BR which only is concerned with bits, and rendering and transformation is also out of scope. Neither the ISO 27000 series nor

[1] See http://www.dcc.ac.uk/tools/trustworthy-repositories/

TRAC is specific in expressing integrity in terms of bit preservation, although DS/ISO/IEC 27005 annex C has a useful list with examples of typical threats. These are partly included in our list of requirements. However, the risks prevention based on ensuring bit preservation (number of copies, integrity check frequency and independence between copies as described in [4]) needs to be taken into account as well. This gives us the requirements tree as illustrated in Figure 4. It is drawn using the open source mind map tool Freemind.

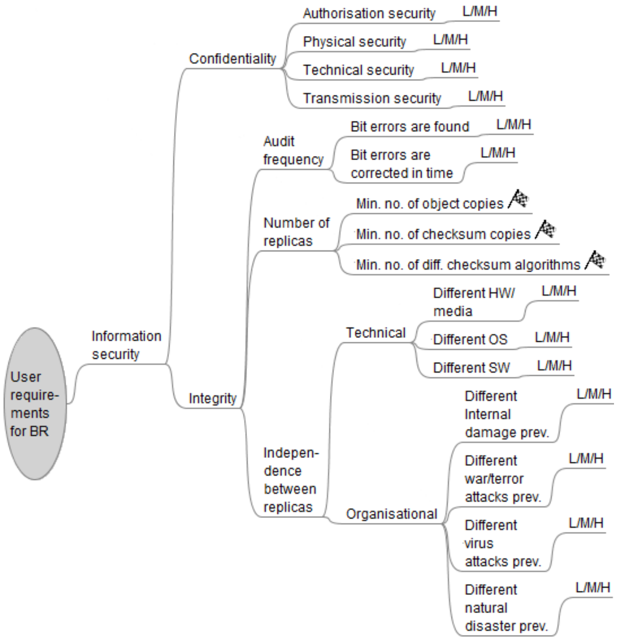
User require-ments for BR
Information security
Confidentiality
Authorisation security L/M/H
Physical security L/M/H
Technical security L/M/H
Transmission security L/M/H
Integrity
Audit frequency
Bit errors are found L/M/H
Bit errors are corrected in time L/M/H
Number of replicas
Min. no. of object copies
Min. no. of checksum copies
Min. no. of diff. checksum algorithms
Indepen-dence between replicas
Technical
Different HW/ media L/M/H
Different OS L/M/H
Different SW L/M/H
Organisational
Different Internal damage prev. L/M/H
Different war/terror attacks prev. L/M/H
Different virus attacks prev. L/M/H
Different natural disaster prev. L/M/H

Figure 4. Requirements for a BR (in a mind map).

The branches indicated by a flag symbol are only indirectly included here, in the sense that they are specified as part of the SLA cases which we will define later. The rest of the branches represent importance of requirements which can be measured using an ordinal scale Low/Medium/High. In the following the requirements from the different branches in Figure 4 are explained. For later reference, each requirement is prefixed with an abbreviation number.

According to the ISO standard the *confidentiality* related requirements should be specified to how data is classification in terms of value, legal requirements, sensitivity and criticality. This leads to requirements of preventing the following risks.

C1: Authorisation security violation, which concerns authorisation in all parts of the BR.

C2: Technical security violation which includes e.g. spying via technical means

C3: Physical security violation which concerns e.g. physical access and theft.

C4: Transmission security violation which particularly looks at transmission issues

The *audit frequency* to ensure *integrity* addresses frequency and timely data restoration. This leads to requirements of preventing the following risks:

A1: Bit errors are found which depends on algorithms for detecting errors and timely appliance.

A2: Bit errors are corrected in time which depends on e.g. when corrective actions take place, and how often audit checks are performed held up against mean-time-to-failure for the individual replicas.

The *independence* between replicas is to ensure that *integrity* is not compromised due to similar errors which can corrupt the data in similar ways.

Risks to be prevented by differences on the *technical* level are:

IT1: Different hardware/media which concerns both the type of media and vendors of hardware.

IT2: Different operating system which concerns the origin of the operating system, the type, and the vendor.

IT3: Different software ensures that the same error will not occur for several copies due to same error in the software installed, e.g. language interpreter or software for BR application.

Risks to be prevented by differences on the *organisation* level are:

IO1: Different internal damage preventions which concerns internal damage e.g. caused by an operator. For simplicity we have also included errors caused by faults in power supply under this category.

IO2: Different war/terror attacks preventions which e.g. relates to the geographical location.

IO3: Different virus, worms attacks preventions which related to how such attacks are prevented.

IO4: Different natural disaster preventions, where natural disaster can be anything from flood to volcanic activity. For simplicity we have also included errors caused by magnetism or radiation here.

4 EXPERIMENT CASES

To make the final cases for evaluation of bit preservation strategies, we need to define cases for; firstly, the digital material to which we want to make a bit preservation strategy along with the levels of risk prevention that we require. Secondly, a case of a BR implementation which offers different bit preservation solutions along with cases of SLAs defining how the services can be used for the digital material.

4.1 Material Cases

The material cases cover different data material that require different confidentiality and bit integrity levels. In Figure 2 this is the ‘material requirements’ which are expressed as importance of preventing the risks expressed in the requirements tree (see Figure 4). Each material case is prefixed with an abbreviation number, which will be used as reference in later tables.

M1: Digital born diaries which are highly confidential, and irreproducible.

M2: Digital born images which are open to the public and irreproducible

M3: Digitised books that are open to the public, and reproducible through re-digitisation.

Table 1 shows the requirement which we have estimated for the different material cases. The importance of preventing the risks is L=Low, M=Medium or H=High.

	Material case		
Requirement	M1	M2	M3
Confidentiality			
C1 (author.)	H	L	L
C2 (phys.)	H	L	L
C3 (tech.)	H	L	L
C4 (trans.)	H	L	L
Integrity			
A1 (found)	H	H	M
A2 (corrected)	H	H	H
IT1 (HW)	H	H	L
IT2 (OS)	H	H	M
IT3 (SW)	H	H	M
IO1 (internal)	H	H	M
IO2 (war)	H	H	L
IO3 (virus)	H	H	H
IO4 (disaster)	H	H	M

Table 1. Requirements for digital material cases.

The Table 1 shows that for M3 (digitised material) it is of medium importance to find single errors, but of high importance to have errors corrected, if large volumes and thus investment of the original digitisation are in danger. Loss of data in a war or terror attack is however only viewed as of low importance.

4.2 BR Case

As a case of a 'specific BR implementation' (see Figures 1 and 2), we have selected different pillar implementations and defined characteristics and functions for calculation of requirements probabilities.

4.2.1 Selected Pillars

As basis for a concrete BR we have made examples of pillars used for Danish BR implementation, supplemented with a cloud pillar (e.g. DuraCloud[1]) and a pillar under different law. The pillars are listed in Table 2.

Pillar	Short description
DiCph	Distributed disk system with RAID in org. A in Copenhagen
DvCph	Off-line DVD in org. C in Copenhagen
TpAar	Tape station in org. B in Aarhus (app. 100 km from Copenhagen)
DiAar	Server optimized for robustness in organization B in Aarhus
Cloud	Cloud in unknown organisation
DiAus	Disk based system in org. in Austria

Table 2. Pillars in BR case.

The cloud pillar is interesting because clouds are emerging, and it would be relevant to see what impact a full replica in a cloud could have on bit integrity and confidentiality. A parameter for bit integrity is also the geographical placement, to determine distances between pillars and danger zones pillars are located in. Since Denmark is small which, we have chosen to add a pillar placed in another country. This choice can also affect confidentiality, because of legal issues.

A pillar has many characteristics and changing just one characteristic can mean a different outcome. The naming of the pillars should therefore only be taken as a short abbreviation for some of its characteristics.

4.2.2 Selected Characteristics

The system and pillar characteristics are many. Even in the prototype BR-ReMS the number is about 100. Therefore we will here only explain what they cover generally, illustrated with a few examples, and referencing where further relevant input can be found.

The characteristics included for this case study are partly based on details of the ISO 20005 Annex C on typical threats. More detailed characteristics could be made by adding relevant parts from the ISO 20005 Annex D on vulnerabilities and methods for vulnerability assessment. Note that Annex D is a specialisation of Annex C, or rather Annex C lists the threats that can cause the vulnerabilities.

The ISO standard takes another approach than the one described here, since its aim is not calculations. For calculations, we need parameters from the technical and the organisational perspective, as well as defining them in terms of facts of the implementation. For instance, concerning risk of flood, we need characteristics on if it is in a flood zone, and in this case what organisational and physical prevention procedures that exist.

Additionally, there are characteristics that are specific to active bit preservation (e.g. bit audit frequency, type of checksum algorithm) and the facts on technical details (e.g. on capacity, Mean-Time-To-Failure, expected hardware life time, media technology) and organisational data (e.g. physical location).

4.2.3 Selected Requirements Calculations

Because of the large number of characteristics and the complex interrelations, the calculations are made at varied levels of detail. For instance the IO1 (internal damage prevention) depends on 25 characteristics.

4.3 SLA Cases

The SLA cases represent cases of 'specific use of BR' (see Figure 2) and constitute the alternative solutions for bit preservation. These are therefore the alternatives to be specified and evaluated in Plato.

The SLA cases consist of a pillar combination for the replicas, as well as the type of replica (C=checksum, F=full) that is stored on the individual pillars. Table 3 lists the following SLA cases with choice of pillar combinations and replica types:

[1] See http://duraspace.org/duracloud.php

S1: As present in DK (except a checksum replica).
S2: Influence of exchange with checksum replica.
S3: Optimised confidentiality in organisation A.
S4: Influence on confidentiality with Cloud replica.
S5: Optimised bit integrity with two full replicas.
S6: Influence of an extra checksum.

	SLA case					
Pillar	S1	S2	S3	S4	S5	S6
DiCph	F	F	F	F	F	F
DiAar	F	C	C	C	C	C
TpAar	F	F				
DvCph			F			
Cloud				F		C
DiAus					F	F

Table 3. Service Level Agreement cases.

For the sake of simplicity we here leave out SLA details on e.g. frequency of bit audits, and we only use one type of checksum e.g. MD5.

5 RESULTS

We will now look at the results we can get from use of the methodology on the simplified case studies. We will firstly look at the results of the BR-ReMS prototype, before proceeding to the actual evaluation using Plato.

5.1 Prototype BR-ReMS Results

The BR-ReMS prototype found that the different requirements were met to L=Low, M=Medium or H=High degree for the different SLA cases. The results are listed in Table 4.

	SLA case					
Requirement	S1	S2	S3	S4	S5	S6
Confidentiality						
C1 (author.)	M	M	H	L	L	L
C2 (phys.)	M	M	H	L	M	M
C3 (tech.)	M	M	H	L	M	M
C4 (trans.)	M	M	H	L	M	M
Integrity						
A1 (found)	M	M	M	M	M	H
A2 (correctd)	M	L	L	L	M	M
IT1 (HW)	M	M	H	L	M	M
IT2 (OS)	H	H	H	L	H	H
IT3 (SW)	M	H	H	L	H	H
IO1 (internal)	M	M	M	L	M	M
IO2 (war)	M	M	L	L	H	H
IO3 (virus)	M	M	H	L	M	M
IO4 (disaster)	M	M	M	L	H	H

Table 4. BR-ReMS results of requirement fulfilment.

Note that the results given here are made independently of specific material cases. It can also be noticed that especially case S4 generally has a very low score on most requirements. The reason is that one full replica was placed in a cloud, where we do not know much about the pillar characteristics. Since the calculations need to account for worst case, we consequently get the value Low for many of the requirements. Note that if we had more precise knowledge of the cloud pillar characteristics then this picture would probably differ.

The difference between case S1 and S2 was that one full replica was exchanged with a checksum replica. This gives lower score on correction, but also higher score on different software. The reason is that difference in hardware only looks at variations for full replicas, which in this case are placed on the two pillars that differ in software.

The relatively high scores in case S3 are mainly a consequence of having one full replica on highly secured DVDs that are off-line and non-magnetic material. There is also a parameter that the other full replica is handled in house.

It is important to note that these results are only indications. The BR-ReMS is still only a prototype. More granularity and more specific functions are needed to give more precise measures.

5.2 Plato Results

Firstly we make a general evaluation of how well the different six SLA alternatives meet the requirements in general, i.e. not considering specific material cases. The results are given in table 5:

	SLA case					
Rank level	S1	S2	S3	S4	S5	S6
Confident.	1,5	1,5	2,5	0,5	1,3	1,3
Integrity	1,6	1,4	1,5	0,8	1,8	2,0
Total	3,1	2,9	4,0	1,3	3,0	3,3

Table 5. Plato results for SLA cases in general.

The results are found by transforming the BR-ReMS results to a uniform scale between 0 and 5 for each requirement (here using: Low=1, Medium=3, High=5), which Plato uses to give a ranked list of the alternatives. For simplicity only the totals for confidentiality and integrity requirements are included in the table.

The ranking in Table 5 shows that case S3, designed to ensure high confidentiality, has the top score both in total and on the confidentiality level. The case S6, with an extra checksum, is the top score on the integrity level. Finally, the S4 case including a full replica in a cloud is ranked with lowest score, due to the low score in the BR-ReMS.

Now we proceed with the evaluation for the three specific types of digital material. Here we scale the results by comparing the required level of importance with the resulting degree that the requirement is met. The schema for defining scales is given in Table 6. The zero value is based on a decision ***not*** to accept a result where the importance for of a requirement for a specific material case is High, but for a specific SLA case the resulting BR-ReMS probability value is Low.

Required value	BR-ReMS result L	M	H
L	5	5	5
M	3	5	5
H	0	3	5

Table 6. Transforming scheme to Plato scale.

5.2.1 *Plato Results for Case M1*

Table 7 gives the Plato results for *digital born diaries* (M1). There is only one alternative for the digital born diaries with a utility value greater than zero, leaving case S1 as the optimal solution. The reason that the other cases are eliminated is that there appear zeros for one or more of the requirements, thus the total performance value becomes 0 (stars indicates where such a requirement appeared in the table).

	SLA case					
Rank level	S1	S2	S3	S4	S5	S6
Conf.	2,4	-	-	-*	-*	-*
Integrity	0,6	-*	-*	-*	-	-
Total	3,0	0	0	0	0	0

Table 7. Plato results for SLA cases to M1.

In the case study we actually designed case S4 to fit this material, and case S4 did also have good scores on confidentiality in the general evaluation (see Table 5). Even taking the inaccuracies into account, this is therefore a bit surprising. The detailed reason is that case S4 got Low score for requirement A2 'bit errors are corrected in time' (see Table 4) while the requirement was to have a High score (see Table 1). The same applies for the requirement IO2 'Different war/terror attacks preventions'. The decision not to accept a Low value for a High requirement therefore has the result of eliminating case S4. This is quite reasonable, when we look at digital born material.

The reason for the Low score on requirement A2 is that one full replica is placed on a DVD in the DvCph pillar (see Table 2), which in our example is only properly checksum checked every 2 years. Even though a separate checksum is offered for voting, there is relatively high risk that the full replica on the DvCph pillar may also be damaged, in cases where the full replica on the DiCph pillar is found to be with error. The reason for the Low score on requirement IO2 is that the two full replicas are placed only one kilometre apart.

If we had chosen only to give positive values in the scores (see Table 6), the result would have been different and case S4 would have been chosen. In a real life situation the choice of zero would be reasonable, and the result should therefore instead lead to a new evaluation, where e.g. a full TpAar replica was added to the SLA. Note, that in some cases, only minor changes in a SLA, e.g. frequency of integrity check on a specific pillar, could make a difference for the result.

5.2.2 *Plato Results for Case M2*

Table 8 gives the Plato results for *digital born images* (M2). The winning alternative for digital born images is case S6. Cases S2, S3, and S4 are eliminated for the same reasons as for the M1 (High requirement value for A2). This leaves the cases S1, S5 and S6.

	SLA case					
Rank level	S1	S2	S3	S4	S5	S6
Conf.	1,0	-	-	-	1,0	1,0
Integrity	2,5	-*	-*	-*	2,9	3,3
Total	3,5	0	0	0	3,9	4,3

Table 8. Plato results for SLA cases to M2.

It is quite reasonable that case S6 wins over case S5, since case S6 contains the same pillars as case S5, but added with an extra checksum. On the other hand it is not obvious why case S6 wins over case S1, since case S1 has three full replicas, while case S6 has only two full replicas and two checksum replicas. The reason is that case S6 is better protected against war and natural disasters by having a full replica abroad (pillar DiAus). Details in the result also show that case S6, because of the extra voter, has a better score than case S1 on requirement A1 'Bit errors are found'. However, because of the inaccuracies in this study, this should *not* lead to a conclusion that an extra checksum is better than having three full replicas.

5.2.3 *Plato Results for Case M3*

Table 9 gives the Plato results for *digitised books*. All three alternatives S1, S5, S6 are winners as equally good alternatives for digitised books.

	SLA case					
Rank level	S1	S2	S3	S4	S5	S6
Conf.	2,5	-	-	-	2,5	2,5
Integrity	2,2	-*	-*	-*	2,2	2,2
Total	4,7	0	0	0	4,7	4,7

Table 9. Plato results for SLA cases to M3.

Cases S2, S3, and S4 are eliminated since we also here required high score for A2 'bit errors are corrected in time'. It can here be noted that case S3 would win in the case of M3, if the score for A2 had not been zero. All other requirements would then have score 5.

A more highly evolved BR-ReMS, with more granularities and details, would likely produce different results, which could lead to choice of a case. Adding requirements on cost and availability, will also change the similar performance values. The reason is that digitised material available for the public, most probably will have requirements of relatively low costs and fast access to material e.g. via a pillar with distributed architecture with high CPU power per data volume.

6 DISCUSSION

As pointed out several times, it is the methodology that is the result of this article. The results of the case studies only illustrate the use of the methodology, rather than giving real life trustworthy results. In order to get better results, there still is work to be done on the requirements aspects such as costs, detail and coverage of pillar characteristics, better BR-ReMS functions for calculating fulfilment of requirements, and more extended use of facilities in Plato.

Requirements could be further developed using the ISO 27000 standard, but could also be based on TRAC including organisational trust, or other models. It should be noted that the methodology does not try to be a substitution for audits following such standards. The calculations made in the BR-ReMS can only give approximations, no matter how detailed it gets. It is meant as a support in evaluation of a bit preservation strategy. Audits of whether pillar characteristics hold should be supplements possibly required in a SLA.

Additional refinement, both on requirement level and pillar characteristics, could be made for issues like encryption, compression, checksum checks using different checksum types etc. Note that these could also be added on the requirements level, if for example an organisation has a policy that *no* digital born material may be encrypted. Use of Plato could also be much more advanced for such cases, e.g. weighting the non-encryption requirement high compared to other requirements. Furthermore, granularity of values for requirements and results could be enhanced to give more nuanced analysis.

Refinement of the functions for requirement fulfilment will be a subject for discussion. Firstly, detailed and possibly automated calculations can easily become too complex to audit, and too rigid to handle inclusion of new aspects. Secondly, different approaches to calculate whether bit audits are done as frequently as needed may give a different outcome. The calculation will probably be based on measures like Mean-Time-To-Failure where it can be debateable how much we can trust such measures.

The level of refinement should also take e.g. hardware/media migrations and upgrades of software into account. If the level of details for characteristics and requirements are too high, it will be hard to make e.g. migrations without re-negotiating all SLAs using the pillar in question. The best solution would be, if a migration plan could be based on re-calculations of characteristics to see whether it would have any negative affect on them. In this case the migration could take place without any re-negotiations.

7 CONCLUSION

The presented methodology has been shown to be useful as an aid to evaluation of alternatives for a bit preservation strategy. Even for the simple case study, with little granularity in requirements and results, and with a BR-ReMS prototype with little refinement, we could produce results that pointed out weaknesses in the SLA cases covering different pillars and characteristics.

The planning tool Plato helps in the analysis of the results. Without Plato, it would have been much more difficult to analyse the results of the BR-ReMS.

The BR-ReMS has also proven useful, at least in the way it structures characteristics for a BR. There may be other approaches to define requirements which the BR-ReMS also can support.

Even though the methodology has been shown to work, there is still a lot of work to do on requirement specification including standards like TRAC, ISO, DRAMBORA[1], and work on detailing the BR-ReMS on characteristics and calculations on requirements specification. Furthermore development of more detailed requirements in Plato will enhance the outcome of using the methodology.

Further work will also study how the methodology can assist consumers in choice of bit preservation strategy and formulation of SLAs, as well as how it can assist service providers in long term operation of parts of a bit repository fulfilling SLAs.

8 ACKNOWLEDGEMENTS

Part of this work was supported by the European Union in the 6th Framework Program, IST, through the PLANETS project, contract 033789.

9 REFERENCES

[1] Becker, C., Kulovits H., Rauber A., Hofman H. "Plato: A service oriented decision support system for preservation planning", *Proceedings of the Joint Conference on Digital Libraries*, Pittsburgh, USA, 2008.

[2] DS/ISO/IEC 27000-27007, first edition, 2009.

[3] Kulovits H., Rauber A., Kugler A., Brantl M., Beinert T., Schoger A. "From TIFF to JPEG 2000? Preservation Planning at the Bavarian State Library Using a Collection of Digitized 16th Century Printings", *D-Lib Magazine Vol. 15 No. 11/12*, 2009.

[4] Rosenthal, D.S.H. "Bit Preservation: A Solved Problem?" *Proceedings of the International Conference on Preservation of Digital Objects*, London, United Kingdom, 2008.

[5] Rosenthal, D.S.H., Robertson, T., Lipkis, T., Reich, V., Morabito, S. "Requirements for Digital Preservation Systems, A Bottom-Up Approach", *D-Lib Magazine Vol. 11 No. 11*, 2005.

[6] Zierau, E., Kejser, U.B. "Cross Institutional Cooperation on a Shared Bit Repository". *Proceedings of the International Conference on Digital Libraries*, New Delhi, India, 2010.

[1] See http://www.repositoryaudit.eu/