The Rescue of the Danish Bits

A case study of the rescue of bits and how the digital preservation community supported it

Eld Zierau
Royal Danish Library
Søren Kierkegaards Plads 1
1221 Copenhagen K
Denmark
elzi@kb.dk

ABSTRACT

The aim of this paper is to illustrate how vulnerable bit preservation is, regardless of how well prepared you think you are, and hopefully to inspire other institutions that may face similar challenges at some point.

At the same time, this paper illustrates the importance of research in digital preservation and the need for conferences like iPRES to enable communication about relevant practical experiences and research results.

The paper includes an illustrative story about how parts of the Danish digital cultural heritage was rescued after being endangered by accumulating political, environmental and organizational events, and where the final solution involves outsourcing parts of the bit preservation solution.

The story includes a description of how the bits were rescued by different means. Firstly, by using research results to convince political management that bit preservation is much more than a storage solution that can be outsourced. Secondly, by initiating various actions to re-establish the wanted level of bit safety. And thirdly, by establishing a contractual and procedural basis for outsourcing parts of the bit preservation.

KEYWORDS

Bit preservation, risk management, outsourcing, organization, community, cooperation

1 INTRODUCTION

The purpose of this paper is many-sided. It aims at contributing to problem solving for different risk scenarios for bit safety and include outsourcing as part of the solution. The paper also illustrates a case where the digital preservation community (e.g. represented by iPRES) has had an obvious impact on digital preservation in practice.

The ultimate goal (and hope) with this paper is to contribute to the discussion of solutions for bit preservation and what parts of bit preservation that can be outsourced. It may also serve as a case study for the "Preservation Storage Criteria" which are currently under development, and aimed at support of decision making on preservation storage [1].

The case describes how Danish digital cultural heritage was rescued after being endangered by two waves of accumulating events. It will illustrate the vulnerability of bit preservation, - even when a lot of effort has been put into being on top of the risks involved. The paper will include examples of the solutions in the form of a description of how the bits were rescued by:

- *Use of research results in an expert statement* (included) to convince political management that bit preservation is far from just a technical question, - and that it is necessary to conduct continued ongoing risk management for the bit preservation.
- *Risk assessment and mitigation actions* that have been initiated in relation to the different threats to the bit safety.
- *Description of elements in outsourcing agreement* which represents the – to us – most important issues in settling on the final contract with associated procedure descriptions for parts of the bit preservation solution.

To introduce the challenges in a non-technical way, the first wave of threatening events is described in a fairy tale style below:

Once upon a time in the small kingdom of Denmark, there were two national libraries. Each library was responsible for preserving different parts of the Danish digital cultural heritage. Their mission was to ensure that they had at least one healthy "clone"of their material alive 'forever after'. The libraries helped each other fulfil this mission by independently taking care of each other's clones, and treating them as if they had been their own. This way, they ensured that any damaged clone was quickly replaced by a new healthy one.

One day the government decided that the two libraries had to merge. The libraries were determined to make it a 'happily ever after' 'marriage of convenience'. However, the marriage became a threat to the clones, since they were now in one household, and thus there could be threats caused by one 'common' (not two independent) household. The new merged library did its best to imitate separate households, until a more sustainable solution was found.

Then the government issued a new order: All the clones had to be placed in a separate IT household, while the merged library still had the responsibility of the clones' survival. This was even worse, since it meant that the clones would not only be in one household, but in a foreign household which had no prior knowledge of how the clones should be treated. Furthermore, the

IT household was unwilling to let the library follow and adjust the treatment of the clones.

The library fought for the well-being of its clones and managed to convince the government that only one of each clone could be moved to the IT household. Furthermore, the IT household eventually agreed to allow the library to follow and adjust the treatment of the out-of-home clones, when needed. All were now happy, since this also solved the problem of a common household.

… In traditional fairy tales the story ends here, but in more modern 'fairy tales' like *Star Wars*, there is a life after the happy ending which may not be "happily ever after". For the new Royal Danish Library, the marriage is still going strong, but the circumstances for bit preservation have run into further complications represented by the second wave of events, described later.

To understand the more serious real-life part, the next section will provide an introduction to basic theories for bit preservation and related concepts used in this paper. On this basis, a more detailed description of the full risk scenarios and how they were handled is provided. This is succeeded by the translated expert statement and a section about the elements in the final setup of an outsourcing agreement. Finally, there is a description of the second wave of events before the ending discussion and conclusion.

2 BIT PRESERVATION

This section serves as a basic introduction to bit preservation as a discipline. The purpose is to enable a better understanding of the terms used in the expert statement. The expert statement will provide additional argumentation.

2.1 Bit Preservation Definition and History

Bit preservation is defined as the required activities to ensure that the bit-streams remain intact and readable. In other words, to make sure that we do not lose bits or the ability to read them.

Bit preservation has been acknowledged as a discipline over the last two decades. Although, there are still voices saying that bit preservation is a solved problem, there is growing evidence that bit preservation involves many facets, where some problems are still unsolved.

Particularly literature from the 1990s contains numerous examples of loss of digital materials due to lack of bit preservation. However, the loss of data has been a sensitive topic, which few will admit to have suffered [2]. There is still loss of data and most likely not all cases have been made public.

In the late 1990s, the actual handling of physical storage was seen as the least of the worries as well-established techniques (such as checksums) were trusted to ensure bit safety [3]. In the past two decades, there have been great technical enhancements of storage technology, which have contributed to minimize the lack of focus on bit preservation. For a long time, bit preservation was regarded as a question of deciding on the right storage media (similar to the focus for physical preservation, e.g. the paper in a book). Examples are digital media with high life expectancy (e.g. using microfilm as a media [4]) and self-repairing software like Redundant Array of Inexpensive Disks (RAID) [5]. However, neither long lasting media nor RAID is sufficient to ensure bit preservation [6].

Over time, there has also been a common misunderstanding that bit preservation is a question of replication in the form of backup [7]. The problem with backup is that the copies are not equally worthy, and they are never checked for errors. Consequently, in case the original is lost, there is no guarantee that the backup copy has not been lost or damaged in the meantime.

During this century, there has been a growing awareness of the necessity to involve risk management of different threats, other than hardware threats, to the bit safety [8,9], especially expressed in David Rosenthal's paper: "Bit Preservation: A Solved Problem?" [10].

2.2 Bit Preservation Implementation

A basic general view of a bit preservation implementation is illustrated in Fig. 1.

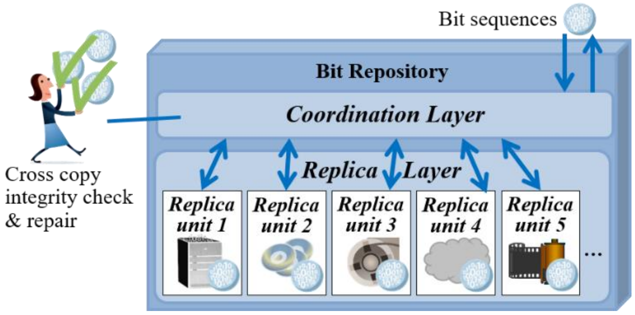


Figure 1: Components in bit preservation.

Fig. 1 reflects the three main ingredients for establishment of bit preservation, which are [10]:

- Several copies of data
- Independence between the copies
- Frequency of checks of whether the copies are identical

It is important to note that a bit repository is both organization and technique, i.e. much more than just technology. Organization and technique are both needed at the coordination layer and in each of the replica units. A replica unit provides the foundation for secure storage of one copy of the data.

Independence between the copies is illustrated by the different media used for each of the replica units. Another typically used independency factor is the geographical location, but there are many other independency factors, like software, human intervention, operating system and media manufacturer.

Checks of whether the copies are identical are often based on the checksums of the files. This method is used because direct comparison of files is very expensive. A checksum-based comparison can be justified, since the risk of having the same checksum for a corrupted file and for its original is negligibly small. Integrity checks through comparison across all the copies are also illustrated in Fig. 1, as part of the coordination layer. Furthermore, this layer covers repairing of faulty copies as well as coordination of ingest and access to the copies. In order to enable identification of a copy with errors, there has to be at least three "votes". As it can be expensive to have full copies as the basis for a "vote", there can be additional, separate and cheap copies of checksums, solely to act as votes when doing cross-integrity checks.

Risk management combined with ***cost benefit analysis*** must be used in order to decide on the number of copies, independency factors and frequency of integrity checks. Nobody has unlimited budgets, therefore you will always aim to have the optimal implementation of bit preservation from what the budget can provide. However, bit preservation is about mitigating risks of losing data, thus the questions left are how large the budget is and what risks you are willing to take.

The risk management must also take into account the different ***information security*** issues covered by the ISO 27000 series [11], which apart from integrity (bit safety) also covers confidentiality and availability. There are various examples of such additional requirements, e.g. accessibility of bit preserved material [12] and requirements to confidentiality [13].

Sustainable bit preservation requires continuous risk management. This is emphasized in the ISO 16363 standard "Audit and certification of trustworthy digital repositories" [14] p. 19:

> "A trustworthy digital repository will understand threats to and risks within its systems."

Some level of auditing is therefore required to keep the risk analysis up to date. This standard is based on the ISO 14721 standard "The OAIS Reference Model" [15], which is not very explicit on bit preservation. A model building on OAIS that is more explicit on bit preservation is the Outer OAIS – Inner OAIS model (OO-IO model) [16], which is an outcome of an international research project "Framework for Applying the OAIS Reference Model to Distributed Digital Preservation" [17] in the early 2010's.

3 THE REAL STORY

The merger of The Royal Library in Copenhagen (KB-Cph) and the State and University Library in Aarhus (KB-Aarhus) was announced in the fall of 2016. By then, both libraries already used the same software/framework for their bit preservation solutions. This software framework was developed previously in a joint project between the two libraries and The Danish National Archives [18]. A lot of work had been invested in establishment of solid bit preservation in Denmark, and we considered ourselves to be in control of the bit preservation challenges. The two libraries used different operational implementations of the framework. Initially, the merger did not influence daily operations of bit preservation notably. This is why nothing had been done to mitigate the risk of having a merged organization when we reached the fall of 2017.

Before iPRES 2017, the Danish government had already stated that all IT operation should be moved from the newly merged Royal Danish Library to an organization called the Agency for Governmental IT Services (here abbreviated State-IT). The State-IT already managed IT hard- and software for other state departments. However, at that stage, bit preservation was seen as a special case with separate negotiations between the government and the library. Therefore, there seemed to be an understanding that bit preservation should be treated differently from normal storage operations.

On the last day of iPRES 2017, I received a request to give as many arguments as possible for NOT moving all bit preservation to the State-IT. This request came because the negotiations had turned out in a way that left no opening for the library to keep any copies of their data nor to conduct auditing for the outsourced copies. I was aware that the library staff being at the negotiation table were fully experienced in bit preservation. Therefore, they had probably already put forward all the right arguments. This is the reason why I suggested to make an expert statement, in which I could use my PhD degree in digital preservation to emphasize research results and case studies from the community to push forward documented research-based arguments. This expert statement is provided in the next section of this paper translated into English.

I cannot tell what role the expert statement has played in the final outcome. However, I can say that I was enrolled in the negotiations just after writing the statement, and it was my experience that at this stage there were openings, both regarding keeping copies and conducting audits.

The library already had a task of mitigating the risks associated with the fact that all copies were operated within the same (merged) organization. It turned out that outsourcing some of the copies would be the solution for this task, since it would reintroduce the organizational independence between copies. Furthermore, it was possible to ask for a different geographical location for the outsourced replica units. This provides geographical independence, especially for data which were only stored on tapes (all tape replicas were placed in the same city, but in different locations).

The negotiation resulted in an acceptable proposal, and this is where the fairy tale ends. In this case, the biggest obstacle was to convince top level management (above the library) that bit preservation is much more than just IT storage. Before describing the next threats, the translated expert statement is provided.

4 THE EXPERT STATEMENT

The following is a translated version of the expert statement that I provided. It is titled: "Expert statement about requirements to bit preservation of the Danish digital heritage".

The contents and the message are the same, but there have been slight modifications in the translation:

- Internal references to institutions have been adjusted to be understandable for foreigners
- Rephrasing has been made to make it clearer in English
- The references have been merged with references for the rest of this paper

Please excuse my self-referencing. I hope the reasons for this are made clear by the contents of the statement and the previous description of why the statement was written.

4.1 Introduction [of Statement]

The purpose of this document is to describe the needed requirements for bit preservation of the part of the Danish digital cultural heritage which the Royal Danish Library is responsible for. It is the library's responsibility to ensure as high a probability as possible that this digital cultural heritage will still exist for future generations.

The Danish digital heritage covers an increasing part of the Danish cultural material for both present and past time. It covers born-digital material, e.g. information from the internet, e-books, e-mails from cultural personalities, computer games with relevance for Danish heritage, radio/TV, online newspapers and much more. Furthermore, the digital heritage covers substitution digitization in cases where the physical original no longer exists, e.g. digitization of film negatives that are deteriorating as a result of their chemical composition.

The author of this expert statement holds a PhD degree in digital preservation building on a Master's degree in computer science. The PhD dissertation was titled "A Holistic Approach to Bit Preservation" [19]. She is the only person in Denmark with a computer science based PhD in digital preservation, and as the title of the dissertation indicates the topic is bit preservation. Consequently, she is the person with most knowledge of this field in Denmark. She is also one of the leading experts internationally, and she is recognized for her work with bit preservation.

The primary purpose of the descriptions in this expert statement is to make clear which elements are necessary in bit preservation, as well as demonstrating the complexities related to bit preservation. The expert knowledge used for the descriptions here was also used for decisions on the current digital preservation strategy for the Copenhagen part of the library (please refer to [20]). The expert statement focuses on aspects of importance to a possible outsourcing of parts of the bit repository, - and to what extend outsourcing is possible without compromising the bit safety.

In order to ease the readability, some examples from the library will be provided. It must be emphasized, that these are only meant as examples, and thus, they do not represent an exhaustive list of issues that must be addressed in bit preservation.

4.2 Executive Summary [of Statement]

Today, the Royal Danish Library is responsible for the preservation of a large part of the Danish digital heritage. The very basis of preservation is the bit preservation – without the right bits, there is no chance to interpret data (as part of logical preservation). In other words, a failed bit preservation will mean irreversible loss of Danish digital cultural heritage.

Bit preservation is much more complex than backup, and it requires continuous risk management to ensure that data does not get lost and that other information securities are maintained (confidentiality and availability). Basically, bit preservation consists of ensuring that a number of copies of data is stored on independent technologies in independent organizations, where the copies are regularly checked for integrity and are repaired when needed. The number of copies, independence and frequency of integrity checks are determined by risk assessment. Subsequently continuous risk management must ensure that changes of conditions do not result in the risk of losing data.

Risk management is highly associated with the actual responsibility of bit preservation. Only the responsible organization can assess which risks are acceptable. Besides bit safety risk, this includes risks related to the different requirements of availability and confidentiality. Taking action on mitigating risks requires control to a degree where elements in bit preservation can be changed. Therefore, there are limits to which parts of bit preservation that can be outsourced to other organizations:

Firstly, if a single copy is outsourced, it is necessary to maintain control of the elements that ensure the copy to be independent from other copies, - otherwise the assessment of the number of copies needed may not hold. This means that frequent audits of the outsourced copy must be performed. Furthermore, it is necessary to have access to information about the conditions of the copy.

Secondly, there are limitations to how many copies you can outsource to the same storage provider. The reason is that changed conditions at the storage provider will have consequences for all the copies placed there. This poses an unacceptably large risk of the data being lost. The same is the case when a majority of votes in integrity checking is in the control of one organization.

On the other hand, outsourcing one copy of data can help to obtain organizational independence. However, this does require that the Royal Danish Library is in control of the total solution. This means that it must be possible for the library to change storage provider at a later stage, in case another storage provider can offer better opportunities to fulfil requirements to bit safety, confidentiality, availability or economy.

4.3 Basic Requirements to Bit Preservation

The basic principles of bit preservation were formulated by David Rosenthal from Stanford University (now retired). In short, the essence is that bit preservation is ensured by having a number of independent copies of data, where the integrity of the copies is checked regularly and corrected in case errors are discovered [10].

Number of copies: It is not possible to determine the number of copies needed in general for any organization. The reason is that a suitable number will depend on the infrastructure in which the copies are placed – and most importantly how independent the copies are. A simple example is that it will always be preferable to have three copies of data in different geographical locations, rather than a hundred copies placed in one basement where all copies could be destroyed by a fire or a cloudburst.

David Rosenthal represented the LOCKSS system for several years. He stated that there should be at least seven copies of data to ensure the bit safety in such systems, which primarily supports geographical and organizational independence[1]. However, in my experience, seven copies may not be enough, based on findings in one of the case studies in the DDP project.[2] In this case study, they were ten copies of data, but the organization experienced that seven of the ten copies were destroyed because of the same hardware error. This is yet another example showing that determination of the number of copies must rely on evaluation of their independence (and frequency of integrity check).

At the Royal Danish Library, we have three copies of data for digital materials regarded as important. This is a relatively small number of copies. The small number is justified by the large independence between the copies and the additional independent checksum copies supporting integrity checks (to point out the right copy in case two out of three copies have errors).

Independence: Independence between the different copies must ensure that the same event cannot harm too many copies, and consequently lead to loss of data. The independence must cover technique, human intervention and disasters like fire, hacking, natural disasters, war, economical cut-downs etc.

To mitigate risks of loss of data caused by human intervention (intentional and non-intentional), the library requires independent development of software used in operation of the individual copies. Furthermore, it is a requirement that different organizations operate the different copies and checksum copies (contributing to the integrity checks). According to David Rosenthal, one of the most common sources of data loss is human error. Therefore, the organizational independence is particularly important, - although it cannot be the only independence parameter.

To mitigate risks of loss due to technical faults, we require that data copies are placed in environments with different operating systems, software and hardware/media.

To mitigate risks of loss due to disaster events, we require that copies must be geographically independent. Examples of disaster events are fire, flood or other natural disasters. Especially in regard to natural disasters, the geographical independence is important. Furthermore, a copy placed in another country can assist in mitigating risks of destruction in case Denmark is invaded by a hostile power. It has also been discussed to store copies on independent media, since magnetic storage media can be harmed by magnetic storms. However, so far, it has not been possible to establish an optic media platform. Instead, the Copenhagen part of the library has chosen to store a copy in a mountain in another Nordic country.

Integrity check: Unlike backup, bit preservation includes an essential integrity check across copies. All copies of data are equally worthy. The cross-copy integrity check is used to find out whether any of the copies have been damaged, and to have them fixed if such damage is encountered. It should be noted that the integrity cannot be ensured by having internal integrity checks for a single copy only. Locally, there may be human intervention or technical faults that result in errors where a file is corrupted, but the checksum matches the corrupted file. In such cases, the error is not discovered.

The frequency of both internal and cross-copy integrity checks must be settled from an analysis taking into account the number of copies. If there are only two copies of data (supplemented by independent checksums to have enough votes to determine which copy is the right one in case of errors), an error in one of the copies will put the data at greater risk of being lost than in cases with more copies. With one corrupted copy out of two, just a minor error in the right copy will cause absolute loss of the data. Therefore, less number of copies should be accompanied by a higher frequency of cross-copy checks to be able to ensure timely correction of errors.

4.4 Bit Preservation and Information Security

According to the ISO 27000 series of standards about information security, it is necessary to consider different aspects, namely: integrity, confidentiality and availability [9]. Economy should also be taken into account, since it can influence the other aspects as well.[3]

The library has more than one bit preservation solution according to different requirements to different materials. The difference in requirements are met by placing different number of copies on different replica units. For example, the requirements for confidentiality are associated with additional costs for the involved replica units. Another example is bit safety for digital materials that are created from an analogue material which is also preserved. In this case, bit preservation only serves to protect the economical investment in digitization rather than protection from absolute loss. Therefore, bit preservation for these kinds of materials does not require as many copies compared to the number of copies required for irreplaceable born-digital materials. This is also reflected in the strategy for bit preservation at the Copenhagen part of the library as given in [20].[4]

It should be noted that the requirements for different replica units can change over time. Likewise, the combination of replica units for bit preservation of a collection of materials may change, e.g. in case new requirements are put forward at a later stage.

Integrity: In bit preservation, the integrity is covered as part of the bit preservation.

[1] Description of LOCKSS as well as the recommendation of using at least 7 copies can be found in the paper" Distributed Digital Preservation: Lots of Copies Keep Stuff Safe" [21].

[2] The DDP project (Distributed Digital Preservation) was primarily an American project, where I was invited into the project and partly financed by funding from the Danish Ministry of Culture. The project participants represented several large American universities as well as organizations like Internet Archive and Chronopolis that represented examples on use of distributed digital preservation. The project is described in the paper "Creating a Framework for Applying OAIS to Distributed Digital Preservation" [17]

[3] Examples and discussion of how information security aspects and economical aspect can influence bit preservation can be found in the paper "Representation of Digital Material preserved in a Library Context" [22] and "Evaluation of Bit Preservation Strategies" [23].

[4] A more concrete example of the need for several bit preservation levels can be found in "Preservation of Digitised Books in a Library Context" [24] and "Representation of Digital Material preserved in a Library Context" [22].

Confidentiality: Confidentiality is mainly an issue for the materials from the Copenhagen part of the library. A rather large part of these confidential materials has high value for the Danish digital cultural heritage. One example is materials donated by Danish authors, not to be made public within the next 50 years. Such materials are also bit preserved and the securement of confidentiality for the copies in bit preservation is therefore essential.

Confidentiality is in conflict with bit safety since bit safety is best secured by having as many copies of data as possible, whereas confidentiality is best secured by having as few secured copies as possible. It is also conflicting if encryption is used, since encryption poses a risk of losing data in case the encryption key is lost. This is the reason why the library has not allowed encryption of data in bit preservation so far.

Securement of confidentiality has to be made at all levels, i.e. in transfers of data and access to data. The more parties involved in bit preservation, the bigger the risk of leaks of confidential materials. If such a leak happens, it will have high impact on the Royal Danish Library's reputation as one of the most trustworthy repositories in Denmark.

Availability: There are many different types of availability that must be serviced. For instance, the library has a specialized set-up for processing Danish web archive materials. In this set-up, it is ensured that there is no connection to the open internet while processing the confidential (personal sensitive) data. Furthermore, it is ensured that the data cannot leave the servers they are placed on and the servers are secured in a locked cage with limited net and physical access.

In the Copenhagen part of the library, there has not yet been a need for fast access to large data sets under bit preservation. However, at some point there might be cases where economic benefits from accessing bit preserved copies for dissemination or research purposes can be achieved. This is expected to become a requirement at some stage.

Economy: Bit preservation needs to be economically sustainable in order to secure data over time. Bit preservation is not without costs and therefore it requires continuous funding for maintenance, evaluation, auditing and adjustments. One of the biggest threats to sustainable bit preservation is therefore the economy. If the budget is cut for bit preservation, it will only be possible to bring down the costs by redoing a risk analysis and possibly accept a reduction in securement of confidentiality, availability and bit safety whatever consequences this will have for the organization and the data.

4.5 Parts of a System for Bit Preservation

In order to create a system (both organizational and technical) to support bit preservation, it is necessary to have the following:

So-called ***replica units*** that each consist of both organization and technique to maintain a copy of data independently from the other copies. The replica unit must meet requirements that are set in order to fulfil requirements in the full bit preservation solution, e.g. requirements to media, placement, access or confidentiality.

A ***coordination layer*** that consists of technique and organization for coordination of copies on the replica units. This covers coordination of access, writing, cross-copy integrity check, and repair operations in case errors are found. The coordination layer must also support other requirements, e.g. different information security aspects.

Continuous surveillance and coordination of activities that can endanger the bit safety must also take place at the coordination layer level. For example, to mitigate risks of data loss as consequence of having several major media migrations on several replica units at the same time. Furthermore, the overall technology watch must be carried out at this level, for example to evaluate if it is beneficial to include a new media (like DNA) e.g. to obtain increased bit safety or lower costs.

Continuous risk management: to ensure that the decided bit safety level is reached through risk assessment and mitigation actions to prevent risks. Risk management has to be placed at the overall level. The risk management is essential to sustain long-term preservation, since technology, organization and digital material will change over time. This is the reason why this topic is further described in a separate section in this [statement] document.

4.6 Overall Risk Management

To ensure sustainable bit preservation, it is necessary to make continuous adjustments to risk management according to changes in conditions for the bit preservation. Examples of such changed conditions are changes in technology, in organization, in economy, in legal framework or in requirements to different information security for the bit preserved material (availability, confidentiality, integrity/bit safety).[5]

It is important to note that a total risk management of bit preservation solutions must include risk assessment of how easy it is to exchange replica units in a bit preservation setup. An exchange of replica units will eventually be necessary to obtain the most optimal solution regarding the security information requirements and economy. However, when dealing with large amounts of data, the data will be at risk while the exchange is taking place.

Continuous performance of auditing the different parts of the bit repository solution is needed in order to keep track of possible changes in the setup. Audits can reveal undiscovered (and possibly unintended) deficiencies that endanger the bit safety. When discovered adjusting actions can be taken in order to maintain the required level of safety.

Risk assessment: The risk assessments must cover all aspects of importance to the library. The bit safety (integrity) has the highest priority, but other information security aspects have to be included as well. Changed conditions for any of the information security aspects must be taken into account, since all the aspects

[5] An example of how to evaluate bit preservation for materials with different characteristics (bit safety and confidentiality) can be found in the paper "Evaluation of Bit Preservation Strategies" [23].

have an influence on the possibilities for setting up bit preservation.

An example of changed conditions could be a request to change the media of one of the replica units to a new type of media (e.g. DNA). Such a request could arise as a result of watch of relevant technologies for bit preservation. This would be relevant if the technology can contribute to obtain better bit safety or lower the costs without compromising the safety requirements.

Another example is a case where a full copy of data is lost, e.g. as a result of a natural disaster or the close-down of a storage provider. Such an incident will put the bit safety at risk, and continuous risk management will be an important tool to quickly restore the bit safety level.

A third example is that conditions can change for a replica unit, e.g. that the operating system is changed to another type similar to the operating system used in another replica unit. In such cases a risk assessment can help clarifying whether an extra copy (on a new replica unit) should be created, - or whether the replica unit in question should be replaced by a different one, - or whether there are other risk parameters that can be brought into play to fulfil the required low risk level.

A fourth example is that an audit of replica units reveals circumstances that endanger the safety of the data, e.g. due to neglect of agreed obligations, - or because the contract was not precise in its description of conditions that has turned out to be important.

A fifth example is that the overall risk assessment reveals dependencies between replica units that have not been dealt with satisfactorily.

A sixth example is that the financial circumstances require lower costs. In this case, there may be a replica unit which has such high costs that it becomes impossible to keep it within the new budget. Therefore, it needs to be replaced with another replica unit to maintain an acceptable bit preservation solution.

Auditing: Auditing is an important contribution to bit preservation. Audits are to ensure that the different replica units and the coordination layer are fulfilling the criteria which forms the basis of risk assessment. Conducted audits at the library have revealed misunderstandings, misinterpretations and overlooked risks. In these cases, the audit has resulted in planning and execution of adjusting actions and thereby eliminating the risks.[6]

An audit of a replica unit can only take place, if the organization holding the replica unit is open and willing to accept and actively participate in the audit. This organization must also accept that the library has strict requirements about the platform of the replica unit (to make sure that it is independent from other replica units).[7] Furthermore, the organization must be flexible in regard to changes in the requirements and contracts. Audit findings may result in such change requirements in order to justify the safety level. So far, this has been possible with the organizations that are contributing to set-up of bit preservation solutions for the Copenhagen part of the library.

Placement of control and responsibility: As with everything else, the responsibility of bit preservation can only be placed with the organization that has the finances as well as the ability to control and carry out actions to ensure bit preservation. This also includes making sure that the right skills are present to support qualified decision making.

As can be seen from above, bit preservation is quite complex and includes much more than just hardware and software. Most importantly, it also includes organization, independence, control and continuous risk management which requires skills within bit preservation.

For sustainable bit preservation, the economy in conjunction with control is essential. Without this combination, it will be impossible to exchange replica units in cases where this is a need in order to make better economy or better fulfilment of information security requirements. An example is a cooperation agreement between the library and another Nordic library to operate each other's replica units. This would mean much better bit safety and in most cases at a lower cost than for an outsource replica unit.

4.7 Conclusions [of Statement]

At the time of writing, the Royal Danish Library is responsible for digital preservation of a large part of the Danish cultural heritage. The Royal Danish Library also has the staff and skills to actually do digital preservation. Thus, it has to be the library that has the finances to actually perform digital preservation as well as the means to control it.

To a great extent bit preservation is a question of risk management and adjustments of the three basic elements[8] to decrease the risk of losing data and meet requirements for other information security aspects. This kind of risk management only makes sense, if the organization performing it has both the control and the responsibility. This control and responsibility cannot be outsourced, since loss of data is irreversible.

Outsourcing of replica units to bit repositories outside the organization can be a good idea to obtain organizational independence. However, in such cases the control of the full bit preservation solution needs to remain at the Royal Danish Library. This means that the library must be able to control the replica unit through audits as well as to change the replica unit in case the security is at an unacceptable level, - or if the replica unit must be replaced by another setup to fulfil requirements better or to be more economical without being less secure.

It is important to note that the Royal Danish Library's choice of having just three copies of data can only be justified by having and maintaining a high degree of independence between the copies. If the independence is lowered, the consequence can be

[6] Examples of level of detail as well as results from such audits are described in the paper "OAIS and Distributed Digital Preservation in Practice" [16].

[7] The challenges related to have a copy of data in a cloud solution in relation to independencies are described in "Evaluation of Bit Preservation Strategies" [23] and in relation to e.g. economy is described in "Distributed Digital Preservation in the Cloud" [25].

[8] Number of copies, independence between copies and frequency of integrity checks.

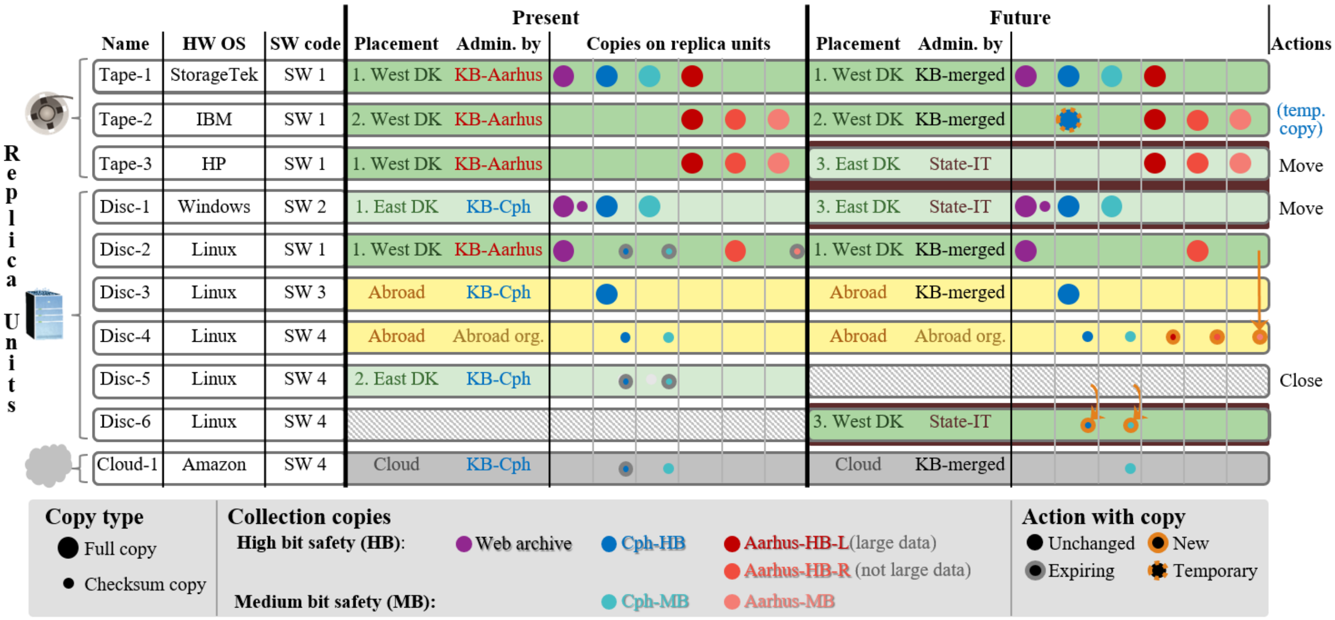
that an extra copy has to be added in order to minimize the risk of losing data.

In this risk perspective, it is also a natural conclusion that the Royal Danish Library cannot outsource the majority of replica units for bit preservation setup. Outsourcing a majority of replica units would imply an unacceptably high dependency on the storage provider, and thus an unacceptably large risk of losing data in case of disagreements or breach of contract.

Furthermore, there are many conditions that can change over time for the bit preserved data (e.g. related to confidentiality and availability). Such changes can require modifications in the setup of the replica units and coordination in the bit preservation solutions. In such cases, it will only be possible to make the necessary adjustments if the control and economy are placed at the library.

5 THE OUTSOURCING AGREEMENT

The decision to move replica units to the State-IT were made in December 2017. The plans was then adjusted with final decisions in the end of February 2018. The resulting plan is illustrated in Fig. 2 (apart from a new blue copy of data, which will be explained later in this paper).



Replica Units			Present			Future		
Name	HW OS	SW code	Placement	Admin. by	Copies on replica units	Placement	Admin. by	Actions
Tape-1	StorageTek	SW 1	1. West DK	KB-Aarhus		1. West DK	KB-merged	
Tape-2	IBM	SW 1	2. West DK	KB-Aarhus		2. West DK	KB-merged	(temp. copy)
Tape-3	HP	SW 1	1. West DK	KB-Aarhus		3. East DK	State-IT	Move
Disc-1	Windows	SW 2	1. East DK	KB-Cph		3. East DK	State-IT	Move
Disc-2	Linux	SW 1	1. West DK	KB-Aarhus		1. West DK	KB-merged	
Disc-3	Linux	SW 3	Abroad	KB-Cph		Abroad	KB-merged	
Disc-4	Linux	SW 4	Abroad	Abroad org.		Abroad	Abroad org.	
Disc-5	Linux	SW 4	2. East DK	KB-Cph		▨	▨	Close
Disc-6	Linux	SW 4	▨	▨		3. West DK	State-IT	
Cloud-1	Amazon	SW 4	Cloud	KB-Cph		Cloud	KB-merged	

Copy type
● Full copy
• Checksum copy

Collection copies
High bit safety (HB): ● Web archive ● Cph-HB ● Aarhus-HB-L(large data) ● Aarhus-HB-R (not large data)
Medium bit safety (MB): ● Cph-MB ● Aarhus-MB

Action with copy
● Unchanged ● New
● Expiring ● Temporary

Figure 2: Present and future set-up for replica units.

Luckily for us, the adjustment also resulted in the simplest solution since only a minimal number of replica units had to be moved, - and no full copies had to be moved from one replica unit to another. The two new State-IT replica units will be based on the existing technology from the library, i.e. the technical environment is physically moved to the State-ITs premises and organization. Fortunately, these two replica units already have exactly one full copy of all the data that required better organizational independence.

The two following major actions which will take place (as illustrated in Fig. 2):

- moving the hardware of replica unit **Tape-3**
  from a place in Aarhus administrated by KB-Aarhus
  to a place in Copenhagen administrated by State-IT
- moving the hardware of replica unit **Disc-1**
  from a place in Copenhagen administrated by KB-Cph
  to a place in Copenhagen administrated by State-IT.

Both actions will provide better organizational independence, - and in addition it will mean better geographical for the collections represented in replica unit **Tape-3**.

Besides this, there are some minor operations to move or create checksum copies. This is done to ensure that no single organization will have the majority of "votes" for any of the collections. Consequently, no single person can delete or change contents intentionally or unintentionally.

There is still work to be done making agreements and procedures to ensure that the library remains in control of risk management. In practice, this means that the agreements as a minimum must ensure that:

1: The library can obtain audit information for risk analysis covering all involved replica units.

2: The library can cancel or change the contract with the State-IT if this is required to obtain the best possible bit preservation

3: The library can require specific characteristics for the replica unit if needed to maintain independence with other replica units, e.g. regarding geographical location, organizational location, hardware producer, operating system etc.

4: The library can require that the State-IT needs prior acceptance by the library, before major upgrades or changes are executed, in order to mitigate risks of major simultaneous changes on several replica units.

5: It will be possible to use the data from the State-IT replica to restore other replicas and vice versa. At the time of writing, we are in the middle of the process of analyzing how to formulate the details for these requirements. We need to ensure that the requirements are implementable, cost effective and above all not colliding with other business requirements within the State-IT.

One of the biggest challenges is to find the balance between control of bit preservation risk management, and at the same time have a replica unit independently controlled by the State-IT. Some issues are related to the fact that the replica unit at the State-IT is part of a larger setup operated and developed by the library. Another interesting issue is where to place the skills about the used media. On one hand, the library needs the skills to evaluate differences compared to other similar media. On the other hand, the State-IT will be the ones actually working with the media, and therefore they should be able to offer advice about new technologies related to the media.

One of the severe points is to ensure that the library can obtain audit information. Previously, we did all auditing on all replica units ourselves, and these audits covered all aspects. The State-IT already has auditing procedures governed by another department of the government. These procedures cover the aspects of the ISO 27000 series standard, ensuring security and restriction of access, that procedures are followed etc. However, the existing standard audit cannot cover all information needed for risk management of bit preservation. We will need to add some extra reporting like thorough documentation for internal integrity checks. Most importantly, we need to inspect the replica unit setup on-site to identify if there are areas where the State-IT replica units are not independent from other replica units outside the State-IT. It is our experience that the previous audits conducted by the library have revealed such dependencies. Therefore, we do not believe that pre-defined reporting is sufficient, and this is why we have invented a new type of audit named ‘knowledge meetings’. The purpose of these meetings is to supply the library with additional information that cannot be covered by standard reporting or audit information.

There have been several issues regarding the definition of form and content of knowledge meetings. One example is that the library wants to ensure that no internal or external person can gain access to different replica units. This can be hard to control especially when the operator is external to the library. The library has previously experienced that the same person showed up for the purpose of doing maintenance for two different tape replica units. The reason was that this person was subcontracted by both consultancy companies each servicing one of the replica units. On the other hand, if the library requires all names of external personnel managing the daily operation of the replica units, this could easily compromise the State-IT’s interests and procedures.

There are of course also many other issues of a more technical character, like placement of test environment for the replica unit, defining a split of monitoring the system, ensuring capacity for access and upload time (e.g. daily deliveries from web archive harvesting and Danish radio and TV transmissions).

Although there are many obstacles, we are very optimistic and look forward to the final implementation of this improvement for the Danish bit preservation.

6 THE REST OF THE STORY – SO FAR

The ‘fairy tale’ ended on February 1st 2018, where everything seemed to result in a happy ending with additional independence in the bit preservation for several of the collections.

By ***March 6th***, a new challenge arose: we were informed by the organization abroad (holding parts of our replica units) that a planned event from May 2018 to the end of 2019 would endanger the data on replica units **Disc-3** and **Disc-4**. The organization plans to expand their premises next to the place where the servers are running. Since this place is inside a mountain, the expansion will include explosions near the servers. We were therefore advised to have the servers moved to another temporary location.

In general, data is not very safe on servers that are switched off and moved. Consequently, there is a potential risk of losing a large part of the data during the move. We started looking at risk scenarios, since the move of **Disc-1** and **Tape-3** to the State-IT could clash with moving **Disc-3** and **Disc-4** out of the mountain and back again when the expansion work ends. As shown in Fig. 2, the Cph-HB collection has (presently) two out of three full copies of data that can be at risk in case of clashing events, which is unacceptable. Although the timeframe for this case is quite large, it is relevant considerations, since timelines for construction work and move of equipment have a tendency to be changed. The decision was therefore to make an extra copy of the Cph-HB collection, and place it on **Tape-2** (the temporary blue copy in Fig. 2). This new full copy is not very independent from the full copy on **Tape-1**, since it is the same organization and only few kilometers apart. However, we consider this acceptable, since it only serves as a temporary additional full copy to mitigate the risk for the full copies on **Disc-1** and **Disc-3**.

At first, it seemed that there was plenty of time to establish the extra copy, since the move of **Disc-1** to the State-IT was postponed by several months.

By ***March 12th***, a new challenge arose: the Danish government gave notice of a possible lockout of all public employees starting April 10th. Although the lockout could be postponed, we had to deal with the risk that the lockout could take place at any time between beginning of April and the end of May. The lockout posed a threat to bit preservation, since very few staffers would be at work during this period, and the people at work would not include staff equipped to deal with emergencies in the event of damage to large parts of a full collection copy. Such a loss must be dealt with quickly to restore the safety level. In normal circumstances, the probability of a major loss during a lockout would be considered very low. However, in this situation the probability was very high, since the lockout period would be in direct collision with the timeframe for the move of the abroad replica unit **Disc-1** and **Disc-3**.

The risk analysis of this new situation pointed to the same solution as in the previous situation: the extra copy on **Tape-2**. However, there is a small risk associated with the creation of this copy, since the tapes on **Tape-1** would have to leave the safety box in order to be read and copied. Since this results in a scenario where two full copies are at risk, the best solution would be to make the copy before the start of the lockout. Copying terabytes of data from one set of tapes to another is not done overnight, - and the timeframe was very small, Easter holidays included.

By ***March 28th***, the lockout was postponed until April 28th. This means we can breathe a sigh of relief for now, as it seems possible to establish the extra copy by then. So, another happy ending so far, - but admittedly, I personally cross my fingers, that we do not have any more bit threatening events this year.

7 DISCUSSION

The case described here is probably different from most other bit preservation cases, e.g. due to legal framework, geographical conditions and risk assessment. However, there is a fair chance that some of the issues described will be quite similar in other bit preservation solutions.

Especially tolerated risks when outsourcing will differ. Therefore, the Danish case may only be used as inspiration, while the formerly mentioned preservation storage criteria are more suitable as a tool for analysis.

The most important message in the expert statement is that bit preservation is NOT solely a question of hardware and IT. It can therefore be used in its entirety, or parts of the scientifically based argumentation may be extracted to make this point.

An important point of this paper is that the expert statement could not have been made without use of sources from the digital preservation community (including iPRES). However, inspiration has come from many other places than just the referred papers, e.g. in the form of case studies and analysis on auditing, OAIS, policies and strategies. Furthermore, international community-based work like the DDP project and the storage preservation criteria are fostered by the networking taking place at such events.

8 CONCLUSIONS

This paper illustrates how vulnerable bit preservation is, especially if accumulating events threaten the bit safety.

In the rescue of the Danish bits, it was essential to make top-level management aware of the fact that bit preservation is much more than a question of hardware and IT. The expert statement has contributed to bring this awareness into play.

It has been illustrated why risk management is an important element in taking timely action to rescue the bits. It has also been argued that the ability to perform risk management must follow the organization that has the responsibilities. Finally, examples have been provided of the challenges of making contracts and procedures in relation to outsourcing parts of the bit preservation.

It is evident that not only the expert statement, but also a lot of the work carried out during this rescue, would not have been possible without using results from the digital preservation community, with iPRES being an important contributor of research results, case studies and setting the frame for networking.

REFERENCES

[1] Goethals, A., McGovern, N., Mandelbaum, J., Schaefer, S., Truman, G., Zierau, E. 2017. Digital Preservation Storage Workshop: Exploring Preservation Storage Criteria and Distributed Digital Preservation. In *Proceedings of the 14th International Conference on Preservation of Digital Objects*, Kyoto, Japan, 258-259.

[2] Manes, S. 1998. Time and Technology Threaten Digital Archives ..., The New York Times, April 1998.

[3] Waters, D., Garrett, J. 1996. Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. Web archive: archive.org, archive date: 2010-01-20 01:49:50 UTC, part, archived URI: http://www.oclc.org/research/activities/past/rlg/digpresstudy/final-report.pdf.

[4] Hofmann, A., Giel, D. M.: DANOK. 2008. Long Term Migration Free Storage of Digital Audio Data on Microfilm, In *Proceedings of the IS&T Archiving Conference*, Bern, Switzerland, 184-187.

[5] Chen, P. M., Lee, E. K., Gibson, G. A., Katz R. H., Patterson, D. A. 1994. RAID: High-Performance, Reliable Secondary Storage. In *ACM Computing Surveys, vol. 26, issue 2,* 145-185.

[6] Wright, R., Miller, A., Addis, M. 2009. The Significance of Storage in the "Cost of Risk" of Digital Preservation, In: *The International Journal of Digital Curation*, vol. 4, issue 3, 105-122.

[7] *The Murray Research Archive's policy for digital archiving*, web archive: archive.org, archive date: 2016-09-06 12:21:50 UTC, page, archived URI: http://murray.harvard.edu:80/policies.

[8] Baker, M., Keeton, K., Martin, S. 2005. Why Traditional Storage Systems Don't Help Us Save Stuff Forever, In *Proceedings of the 1st IEEE workshop on hot topics in system dependability*.

[9] Rosenthal, D. S. H., Robertson, T., Lipkis, T., Reich, V., Morabito, S. 2005. Requirements for Digital Preservation Systems, A Bottom-Up Approach. In *D-Lib Magazine, vol. 11, no. 11*.

[10] Rosenthal, D.S.H. 2008. Bit Preservation: A Solved Problem?. In Proceedings of the 5th International Conference on Preservation of Digital Objects.

[11] ISO/IEC 27000- 27007 (2011- 2016) on Information technology — Security techniques — Information security management systems.

[12] Clausen, L. 2006. Overview of the Netarkivet web archiving system. In *Proceeding of the 6th International Web Archiving Workshop*, Alicante, Spain, 11-24.

[13] Norcen, R., Podesser, M., Pommer, A., Schmidt, H. P., Uhl, A. 2003. Confidential storage and transmission of medical image data, In *Journal of Computers in Biology and Medicine*, vol. 33, issue 3, 277-292.

[14] ISO 16363:2012. 2012. Space data and information transfer systems - Audit and certification of trustworthy digital repositories.

[15] ISO 14721:2012. 2012. Space data and information transfer systems - Open archival information system (OAIS) – Reference model.

[16] Zierau, E. 2017. OAIS and Distributed Digital Preservation in Practice. In *Proceedings of the 14th International Conference on Preservation of Digital Objects*, Bern, Switzerland.

[17] Zierau, E., Schultz, M. 2013. Creating a Framework for Applying OAIS to Distributed Digital Preservation. In *Proceedings of the 10th International Conference on Preservation of Digital Objects*, 78-83.

[18] Jurik, B. A., Nielsen A. B., Zierau, E. 2012. Flexible Bit Preservation on a National Basis. In *Proceedings of the IS&T Archiving 2012*, 2-7.

[19] Zierau, E. 2011. A Holistic Approach to Bit Preservation. Available via overview of computer science PhDs at Copenhagen University.

[20] Strategy for long term preservation of digital collection materials at The Royal Library, Web archive: archive.org archive date: 2018-04-08 10:51:28 UTC, part, archived URI: http://www.kb.dk/export/sites/kb_dk/da/kb/downloadfiler/PreservationStrategyDigitalMaterials-KB-DK-2014.pdf.

[21] Reich, V., Rosenthal, D. 2009. Distributed Digital Preservation: Lots of Copies Keep Stuff Safe. Web archive: archive.org, archive date: 2016-04-05 01:05:22, archived URI: https://lockss.org/lockswiki/files/NIST2010.pdf.

[22] Zierau, E. 2010. Representation of Digital Material preserved in a Library Context. In *Proceedings of the 7th International Conference on Preservation of Digital Objects*, 329-337.

[23] Zierau, E, Kejser U. B., Kulovits, H. 2010. Evaluation of Bit Preservation Strategies. In *Proceedings of the 7th International Conference on Preservation of Digital Objects*, Vienna, Austria, 161-169.

[24] Zierau, E, Jensen, C. 2010. Preservation of Digitised Books in a Library Context. In *Proceedings of the 7th International Conference on Preservation of Digital Objects*, Vienna, Austria, 61-69.

[25] Rosenthal, D., Vargas, D. L. 2013. Distributed Digital Preservation in the Cloud. In *Proceedings of 8th International Digital Curation Conference*.