Costs of Archival Storage

Anders Bo Nielsen, The Danish National Archives, Copenhagen, Denmark; Alex Thirifays, The Danish National Archives, Copenhagen, Denmark; Ulla Bøgvad Kejser, The Royal Library, Copenhagen, Denmark

Abstract

This paper presents an analysis of the cost of Archival Storage. The study is part of a project conducted by The Danish National Archives, The Royal Library and The State and University Library to develop a generic cost model for digital preservation (CMDP). The purposes of the study were to determine the costs of establishing and maintaining a preservation solution destined for long-term preservation of digital materials; and to develop a tool capable of doing this operation. In order to fulfill the purposes, the project employed a combination of own and external experience as well as the OAIS Reference Model as a framework to fully understand and identify the cost critical activities of bit-preservation as described in Archival Storage. We found that the costs of Archival Storage are obviously closely linked to the data volume, but also to the required preservation quality, especially with regard to the required number of copies and the type of storage solution.

Introduction

The exponential growth of the digital information, which memory institutions must preserve, adds pressure on the institutions budgets and emphasizes their need for efficient preservation solutions. A prerequisite for gaining efficiency is to understand which activities are involved in the preservation; to know how much the activities cost; and what the cost of these activities depends on. Such modeling of costs also allows designing "what if" scenarios to help institutions choose between different preservation alternatives.

In order to compare costs it is important that they are related to the resulting quality of the activity. Thus, the costs of preserving digital information are closely linked to the quality of the involved technology, processes and staff.

Cost only represents one side of an economic model; the other is the benefit that preservation of digital information brings to institutions, their users and the general public [3]. In order for institutions to justify their expenditure with regards to preservation they must also be able to demonstrate the benefits of the efforts.

There are different models available for assessing cost and benefits of digital preservation [1], [6], [7], [11], 19]. An overview of current models is provided in [9]. Lately, Rosenthal provided a cost model for storage [16] and Rusbridge et al. are developing the Economic Sustainability Reference Model [17].

This study focuses on how to model storage costs: What are the activities involved in storing digital information for the long term? How can the activities and their quality be described in an unequivocal way? What factors does the storage cost depend on? How can we estimate how storage costs will evolve over time?

The work is part of an ongoing Danish project to develop a cost model, capable of calculating all costs pertaining to trustworthy preservation of digital information and estimating future costs on a 20-year horizon. It is performed in collaboration between The Danish National Archives, The Royal Library and The State and University Library [8][9].

Methods

CMDP uses the OAIS standard [4] as framework and activity based costing, as described by Cooper et al. [5] and in the International Standard Cost Model [2].

Identification of cost activities and dependencies

We have analyzed the descriptions of the functions in the OAIS standard and identified the involved cost critical activities. We have then analyzed each cost activity and identified factors, which influence the costs. Based on this analysis we have implemented the results as formulas in a spreadsheet tool available at www.costmodelfordigitalpreservation.dk.

Accounting principles

We have applied activity based costing and have divided the activities into costs of establishing, operating and maintaining the preservation system. Moreover, these direct costs have been split into costs of labor and to equipment. Indirect costs (overhead) of general administration etc. have not been implemented yet.

Implementation of CMDP

The CMDP tool consists of a series of interrelated sheets. One sheet provides an overview of the whole CMDP tool and seven others specify each of the OAIS functional entities (modules). So far, the modules Ingest, Preservation Planning and Archival Storage have been completed. In addition, the sub-functions under Administration, which interact directly with the completed modules, have been operationalized in the tool.

The first sheet (Input) allows the user to key in basic information about the data. Required inputs from the user are marked with orange, blue is used for parameters, green for formulas, and black for text and fixed values. To enable cost estimation the user must as a minimum type in the expected yearly amount of ingested data in TB. The results of the cost estimates, for each module, are summarized on the sheet named Cost.

Predefined values in CMDP

The CMDP is based on a series of predefined values. These default values are intended to serve as a best practice guideline for preservation institutions. However, there are many differences between institutions, with respect to their mission, mandate, financial resources, preservation strategy, the type and value of the collections that they hold, etc. Therefore, it is possible to change the default values in the tool to better accommodate specific institutional requirements.

Cost critical activities in Archival Storage

There are different types of storage solutions available, and the solutions have different costs and characteristics with regard to the availability, integrity, and confidentiality of the stored data.

It is best practice that preservation organizations store multiple copies of their data to avoid loss of information, and that they seek to spread copies geographically, technologically and organizationally to mitigate this risk.

In this paper we use the term *storage solution* about the system storing all copies, and the term *storage node* about a system holding one set of the copies. Thus, a storage solution may consist of one or more nodes. In addition, the storage solution includes hardware and software for the *storage system management*, such as functions for managing storage hierarchy, and indexing the data objects and their location.

In the following we describe how we have analyzed the OAIS functional entity Archival Storage, and its six functions, to identify cost critical activities.

Receive Data

As described in the OAIS standard: "The Receive Data function receives a storage request and an AIP from Ingest and moves the AIP to permanent storage within the archive." We have identified the following cost critical activities: Receive storage request, Receive AIP, Select media and storage devices, Perform transfer of data, Send storage confirmation.

In CMDP all these activities are considered to be an integrated part of the storage solution, which we will describe in more detail in the next chapter "Cost of storage solution".

We assume that information about use frequency and other service level requirements are defined, and accounted for, in the submission agreement (service level agreement).

Manage Storage Hierarchy

"The Manage Storage Hierarchy function positions, via commands, the contents of the AIPs on the appropriate media…" Under this function we identified two cost critical activities: Position content on media and Provide operational statistics, which are both assumed to be part of the storage solution. For simplicity we have placed the cost of positioning content on media under Replace Media (see below).

Note that the model does not use storage hierarchy, but storage equality, i.e. the data are copied indiscriminately on all the nodes, whereas storage hierarchy differentiates between frequently used and rarely used data.

Replace Media

"The Replace Media function provides the capability to reproduce the AIPs over time […] this function may perform 'Refreshment', 'Replication', and 'Repackaging' that is straightforward." We identified one cost critical activity, namely: Replace media.

For the time being, all costs of media consumption, i.e. both cost of investment and maintenance, are calculated as an average and gathered under Replace Media. We assume that the costs of replacing media correspond to the cost of the initial storage media accounted for in the provision of the storage solution.

Error Checking

"The Error Checking function provides statistically acceptable assurance that no components of the AIP are corrupted during any internal Archival Storage data transfer." According to our analysis the function includes the cost critical activities Check for errors and Correct errors.

Errors typically occur during transfer and storing, and should be found by integrity checks and monitoring error logs. Based on our experience we estimate that it takes 5 min per TB for an operator to check error logs regardless of the type of storage node. There is a lack of studies of the frequency of bit errors during storage. Our best guess is that the occurrence of bit errors is 0.1 per TB for all types of storage systems. Likewise, empirical data on how long time it takes to correct errors once they are identified are scarce. In the CMDP we estimate that it takes 15 minutes to correct an error.

Errors also occur after transfer and storing e.g. due to media decay. In order to find these errors it is necessary to perform integrity checks. The required frequency of the checks and their extension, be it random checks or full scale, is hard to estimate due to lack of empirical data about the occurrence of the errors. The occurrence depends on the media type, but again this assumption is based on limited knowledge [18].

The reason why the frequency and extension of error checking are important for cost is due to the labor needed to transfer and load/unload off-line media. For on-line and near-line nodes integrity checks are generally not cost critical. So far the tool does not calculate the cost of ongoing integrity check after transfer and storing the data.

Disaster Recovery

The OAIS standard states that: "The Disaster Recovery function provides a mechanism for duplicating the digital contents of the archive collection and storing the duplicate in a physically separate facility." Currently, the tool does not include the cost of disaster recovery. We see this function, as part of the storage solution, in the sense that the so-called "duplicate of the digital content" is integrated in the storage solution as a node.

Provide Data

"The Provide Data function provides copies of stored AIPs to Access". We identified three cost critical activities within this function: Receive AIP request, Provide AIP and Send notice. All these activities are assumed to be part of the storage system. The tool does not consider cost of a temporary staging area.

The costs of providing data depend on the type of storage solution. Provision of data in on-line or near-line systems is handled automatically, while labor is required to handle provision of data from off-line systems. This cost is dependent on the frequency of AIP requests and on other nodes available. Normally, any AIP request will be sent to the node with the fastest transfer rate in order to lower cost and increase delivery speed. The uncertainty about future frequency of data provision will normally lead to the use of one on-line or at least one near-line node instead of using only off-line nodes. However, archives with decade long confidentiality clauses may find solely off-line nodes appropriate. These costs of manual provision of data are not included in the model yet.

 ©2012 Society for Imaging Science and Technology

Costs of storage solution

The first challenge in assessing the costs of a storage solution was to make a baseline description of a storage node in a way, which allows for comparison of the costs of different technologies.

Each storage node is based on a specific storage device and media; they can be on-line, near-line or off-line, and use a robot or a loader to automate processes, thereby lowering the costs and increasing the speed of access.

In CMDP it is currently possible to select between six different nodes: RAID-6 SATA, LTO-5 tape library, LTO-5 with autoloader, a DVD-R disc robot, a Blue Ray (BR-R) disc robot and a HDD SATA standalone with docking station. The costs of these systems are described in more detail below.

Storage devices

The capacity of a device has significant importance for the costs per data volume. In order to compare the capacity of different systems, we calculated the capacity in the following three ways and used the lowest figure to define the maximum capacity of the system:

1) The number of media, which can be fitted into the system, limits the capacity of on-line and near-line systems.
2) The capacity can be limited by the maximum time allowed to recover from a disaster: If your organization has opted to preserve three copies of the data and one of those is destroyed, for how long will you accept to have two copies only? This depends on the security requirements of your organization and on the I/O capacity of the surviving nodes.
3) There is an inverse proportionality between the number of times off-line media can be read and/or written, and the lifetime expectancy of the drive.

We have divided the cost of the storage devices in the cost of a base unit, one or more readers and/or recorders, and a number of media frames. The cost of the storage devices is depreciated linearly over their lifetime. Table 1 displays the estimated average lifetime of a selection of devices.

Table 1. Estimated lifetime (years) of device components.

	RAID 6	LTO-5 library	BD-R robot
Base unit	5	10	10
Frame	5	10	10
Rec./reader	n.a.	5	3

The cost of setting up the storage solution is set at 0.5-2 person weeks, and this cost is also depreciated over the expected lifetime of the system. The costs of maintaining the solution are set to 0.5-2 person weeks per year. The cost of planning, selecting, negotiating with system providers has not been included in the tool yet.

Storage media

Storage media mainly encompass magnetic and optical media, and they are available in different formats, such as disks and tapes.

The costs of the media are based on the raw media prices and the effective utilization of the media. For example, only 92% of a hard disk in a RAID 6 system is utilized, while 99% of the capacity of tapes and optical discs can be used. One should note that storing many small data packages results in a lot of wasted space, making these figures invalid. The figures are therefore based on storing large data packages, minimum 1 GB. We have estimated that the cost per TB is € 93 for HDD in RAID 6, € 27 for LTO-5 and € 64 for BD-R.

The costs of the media are depreciated linearly as a function of their life expectancy. HDD are expected to last 5 years on average, LTO-5 and BD-R 8 years.

Furthermore, we have estimated the annual change in cost for each media type. E.g. we estimate that the cost of HDD decrease with 2% per year, LTO-5 with 3% and BD-R with 1% per year. Nevertheless, due to the overall uncertainty of these figures, we have decided not to use this adjustment yet.

The cost of power for the media is also included, since especially on-line hard disks use considerable amounts of energy. The cost of power constitutes € 6 per TB per year, based on 7 W and a kWh price of € 0.1. This is to be compared with the costs of the media in RAID 6 equal to € 19 per TB per year.

The cost of labor to bulk load and unload 1 TB into a storage unit is estimated to 2 minutes for HDD in RAID 6, and likewise 2 minutes for tapes in a tape library, while loading tapes into an autoloader is estimated to take 3.3 minutes per TB. We assume that it takes 8 minutes per TB to load BD-R in a disk robot and 42.6 minutes per TB for DVD's due to their small storage capacity.

When it comes to bulk operations, e.g. integrity check, off-line tapes are not significantly more expensive to handle than near-line tapes. However, if specific tapes are requested, e.g. for access, off-line tapes become much more expensive.

When all these aspects are taken into consideration the average annual cost of media per TB is estimated to be € 25 for HDD in RAID 6, € 4 for LTO-5 in tape libraries, and € 8 for BD-R in disc robots.

The reliability of the media has significant importance for the selection of storage nodes. It is crucial to know the quality of products when comparing their prices. Therefore, when comparing the costs of different storage nodes, it is important to know the bit integrity they provide [20]. Unfortunately, we do not have reliable data for the probability of bit integrity for different media [15]. We can only inform the reader about our crude estimates based on a few studies and our own experience. In CMDP the reliability of the media is estimated as the annual failure rate (AFR) per raw media. The AFR is estimated to be 3.0% for HDD, 0.2% for LTO-5 and 0.1% for BD-R. For the same reason HDD are mostly used in RAID 6 solutions.

Results

The CMDP tool calculates the costs of the storage devices and media over 20 years as an average annually linearly depreciated cost, which includes replacement of system and media as well as labor costs of operating and maintaining the solution. The tool also incorporates the cost of the storage solutions' consumption of power, except power used for cooling. Cooling may constitute a considerable item of expenditure and the reason we have omitted it from the calculations is that it is highly dependent on the specific geographical location and the construction of the facility. This will be included as an option in future versions of the tool. Table 2 provides an overview of the average total annual costs per TB of media and devices, at 100% capacity, including power and labor.

Table 2. Annual cost (€) per TB at full system capacity.

	Device	Media	Total
RAID-6 SATA	27	25	52
LTO-5 library	14	4	18
LTO-5 autoloader	6	4	10
DVD-R disc robot	36	9	45
BR-R disc robot	17	8	25
HDD SATA stand-alone	2	18	20

Due to the fixed storage capacity of the systems the calculated cost per TB reflects how you exploit the maximum capacity of a storage node. For example, if you need to store 300 TB on an on-line RAID 6 system, you will need three systems each with a maximum capacity of 129 TB in total 387 TB, thereby wasting capacity corresponding to 87 TB.

So far the tool does not take into account the cost of investments made in year 20, which should be depreciated over the coming five to eight years, depending on life expectancy.

The storage solutions are not optimized for general access but for access for integrity checks and in case of disaster recovery.

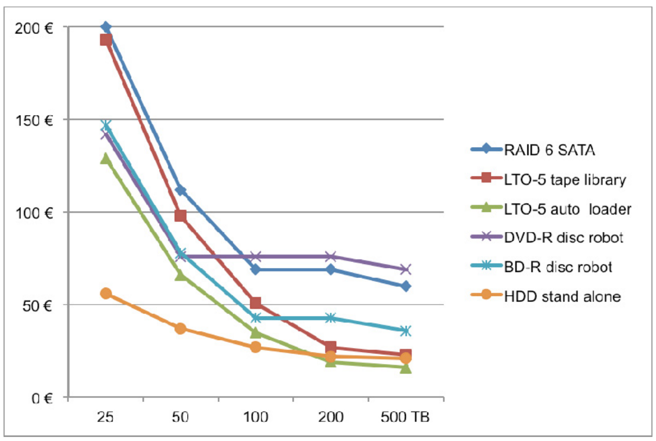

Figure 1. *Annual cost (€) per TB for different types of storage nodes.*

Figure 1shows that the on-line and near-line solutions are the most expensive at low data quantities. This is because the exploitation of their capacities is poor, 25 or 50 TB for maximum storage capacities of respectively 129 and 279 TB.

Once the data quantity passes from 50 to 100 TB, the on-line and near-line solutions drop below the cost of the off-line DVD solution. This is especially because of the low capacity of the DVD-R (4.7 GB), which entails considerable labor. Besides, it is due to the low capacity of the disc robot (58 TB) and the relatively high price of DVD-R per TB (€ 53).

At high data quantities (more than 100 TB), the HDD with a docking station and the LTO tape library solutions converge and join at 500 TB. In comparison the LTO autoloader drops fast and becomes cheaper due to the unit's relatively high capacity (322 TB), low cost (6 €/TB/year) and the media's price (4 €/TB/year).

It is important to note that the chart does not take into account any quality parameters or any form of organizational requirements (political or legal): For example, if an organization needs quick access to its data, the online solution (RAID 6 SATA) is the only viable; on the other hand, if the data are subject to long access dates, the best solution may very well be the LTO auto loader.

Figure 2 shows the time in person-weeks spend on installation and maintenance of the different systems. The time increases with the number of units needed in the storage system: Hence low capacity storage system types cost more at higher data quantity levels (200 and 500 TB), and the time spent handling the DVD Disc robot increases exponentially.

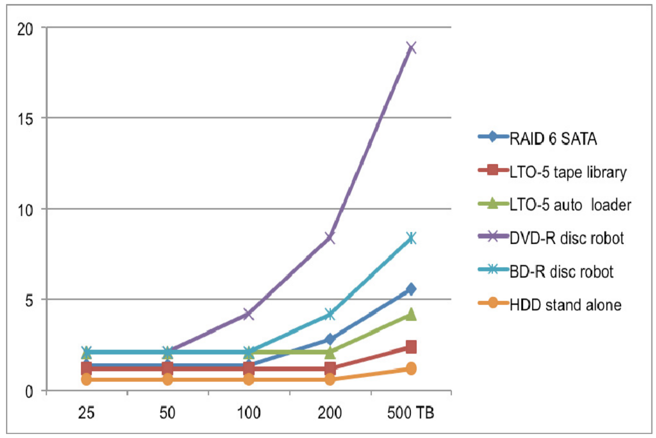

Figure 2. *Annual labor (person weeks) for installation and maintenance.*

Discussion

Generally, there is much uncertainty in the model. Many factors are unknown, among them the lifetime of media, the error rate of media and therefore the necessary number of copies of media and frequency of integrity checks.

The data volume and the number of copies are the most significant cost drivers. Regarding the data volume there are cost of scale for small volumes, i.e. less than 100 TB, due to the capacity of each storage node. Therefore, institutions with small data volumes can benefit from cooperation to overcome these thresholds. One can argue that it is possible to find equipment with much smaller capacity and fixed cost than used in the model. Nevertheless, even though the amount of labor needed to maintain the storage types is small, you still need that skilled labor, which rarely comes at less than full time. This is even truer if you decide to use storage nodes with different storage types in order to minimize the risk or lower cost, e.g. one storage node with on-line hard disc drive RAID-6, one off-line LTO-5 tape library, one off-line Blue Ray optical disc.

In general, independence between copies (technical, geographical or organizational) increases cost and the probability of integrity. Another reason that institutions with small data volumes should cooperate is to provide geographical independence (reducing risk due to theft, fire, natural disasters) and organizational independence (reducing risk due to financial cuts, mismanagement).

It is difficult to specify the necessary number of copies in order to ensure integrity, i.e. a sufficient high probability of

©2012 Society for Imaging Science and Technology

integrity. Nevertheless, most preservation institutions use at least three copies.

The frequency of integrity checks is debated. Regarding cost it seems only relevant to discuss for off-line media, due to the cost of labor handling media. Furthermore, the amount of errors found, especially on tape and optical disc, seems so small that it is justified to take small samples annually combined with full checks every 3 years or less. It should be noted that this justification is based on limited experience, especially regarding LTO-tape. We have experience with tapes since the late 1970'ies, but not with LTO. In general, the technological pace makes it hard to accumulate data before the next generation arrives, and this forces one to make estimates based on the previous generation. Estimating across technologies seems very unreliable. As for errors on optical media it is primarily based on our own experience using less than 10,000 discs for the last 15 years, and for hard disc drives it is mostly based on others' experience [18], [13],[14]. Even though the uncertainty is high, we dare estimate the cost to be low, and for this reason we have not included it in the tool yet.

An important factor for the capacity of the storage types is the time required for recovering from a disaster. It depends on the ratio of storage volume / transfer speed, which is very different depending on the storage types. We have limited the capacity according to the time it takes to read the data in case another node totally fails. One could argue that in case of disaster more reading capacity can be purchased.

Off-line storage is estimated to be less expensive than near-line or on-line storage, but with uncertainty due to the frequency of integrity checks and frequency of provision of data.

Near-line storage in the form of a tape library is less expensive than on-line storage in the form of HDD RAIDs. These well-known ratios are expected to last for the foreseeable future, but the validity of predictions of storage cost over 20 years and even 10 years is small.

We assume that prices on media decrease 1-2% per year, but the general uncertainty is so high that we have not used it in this version of the tool. The price of hard discs has decreased gradually and continuously, whereas the price of tape has decreased drastically with the introduction of new generations. Currently, the flooding in Thailand has caused a major increase in the cost of hard disc drives instead of a small decrease. Apart from natural disasters with long time repercussions the general development of storage technology is hard to predict. A few years ago there was serious speculation that LTO-5 tape would not be introduced, LTO-4 would be the last generation and tape would die, losing to the hard disc drive. Now, speculations are that users in the near future only want to use small capacity solid state drives and on-line cloud storage, and high capacity hard disc drives will be a rarity, used only in large data centers, increasing cost for small scale users. Whatever may happen regarding media, the need for qualified labor and the use of technical, geographical and organizational independence between copies point in the direction of cooperation between preservation institutions.

Conclusion

In this paper we describe how we have used the OAIS Archival Storage to model the cost of storage for a preservation institution and partly implement the model in a tool. Our analysis shows that obviously the cost of storage depends on the data volume, but also on the institution's requirements for the quality of preservation. More specifically, quality refers to the required degree of technological, organizational and geographical independence between the copies. Especially the required number of copies of data and the storage types used are important.

In conclusion we find the model and tool suitable for estimating the cost of Archival Storage, at least showing a minimum cost level, and constitutes a basis for further development. However, several challenges remain both related to the model and the functionality of the tool and its user interface. We especially need more data for the occurrence of errors on different media as well as the cost of maintenance for different storage types.

References

[1] Beagrie, N., Lavoie, B., Woollard, M., 2010. Keeping Research Data Safe 2, Final Report, Charles Beagrie Limited, www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf

[2] Bigelow, Scott, Don't hate the tape! LTO vs. hard drive backup, http://www.scottbigelow.com/article/dont-hate-tape-lto-vs-hard-drive-backup

[3] Blue Ribbon Task Force on Sustainable Digital Preservation and Access. 2010. Sustainable Economics for at Digital Plant: Ensuring Long-Term Access to Digital Information, Final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

[4] Consultative Committee for Space Data Systems (CCSDS). 2002. Reference Model for an Open Archival Information System (OAIS), 650.0-B-1, Blue Book (ISO14721:2003). http://public.ccsds.org/publications/archive/650x0b1.pdf

[5] Cooper, R., Kaplan, R.S., Maisel, L.S., Morrissey, E. & Oehm, R.M. 1992. Implementing Activity-Based Cost Management: Moving from Analysis to Action. New Jersey. Montvale, Institute of Management Accountants

[6] Data Archiving and Networked Services (DANS). Costs of Digital Archiving vol. 2. www.dans.knaw.nl/en/content/categorieen/projecten/costs-digital-archiving-vol-2

[7] Hole, B., Lin, L., McCann, P. & Wheatley, P. 2010. LIFE3: A Predictive Costing Tool for Digital Collections, In: Proceedings of iPRES 2010, 7th International Conference on Preservation of Digital Objects, Austria www.ifs.tuwien.ac.at/dp/ipres2010/papers/hole-64.pdf

[8] Kejser, U.B, Nielsen, A.B., Thirifays, A. 2011. Cost Model for Digital Preservation:Cost of Digital Migration. In: The International Journal of Digital Curation, Issue 1, Vol. 6, pp. 255-267. www.ijdc.net/index.php/ijdc/article/view/177

[9] Kejser, U.B, Nielsen, A.B., Thirifays, A. 2011. Cost Aspects of Ingest and Normalization, Proceedings of the iPRES2011 Conference, November 1-4, 2011, Singapore, pp. 1-10.

[10] Kennedy, Patrick, The RAID Reliability Anthology – Part 1 – The Primer http://www.servethehome.com/raid-reliability-failure-anthology-part-1-primer/

[11] NASA Cost Estimation Toolkit (CET). http://opensource.gsfc.nasa.gov/projects/CET/CET.php

[12] OECD. 2004. International Standard Cost Model Manual to reduce administrative burdens. www.oecd.org/dataoecd/32/54/34227698.pdf

[13] Panzer-Steindel, Bernd, Data integrity, Draft 1.3 8. April 2007, http://indico.cern.ch/getFile.py/access?contribId=3&sessionId=0&resId=1&materialId=paper&confId=13797

[14] Pinheiro, Eduardo, Weber, Wolf-Dietrich and Barroso, Luiz André, Failure Trends in a Large Disk Drive Population, Proceedings of the

5th USENIX Conference on File and Storage Technologies (FAST'07), February 2007, http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/da//archive/disk_failures.pdf
[15] Rosenthal, D.S.H. Bit Preservation: A Solved Problem?, In iPRES 2008: The Fifth International Conference on Preservation of Digital Objects. 2008. http://www.bl.uk/ipres2008/presentations_day2/43_Rosenthal.pdf
[16] Rosenthal, D. Economic model of Storage. November 2011. http://blog.dshr.org/2011/11/progress-on-economic-model-of-storage.html
[17] Rusbridge, C. Update on the state of the Economic Sustainability Reference Model. https://unsustainableideas.wordpress.com/2011/10/17/update-state-ref-model/
[18] Schroeder, Bianca and Gibson, Garth A, Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you? Proceedings of the 5th USENIX Conference on File and Storage Technologies - Paper, Pp. 1–16, http://static.usenix.org/events/fast07/tech/schroeder/schroeder_html/index.html
[19] Slats, J. and Verdegem, R. 2005. Cost Model for Digital Preservation. Proceedings of the IVth triennieal conference, DLM Forum, Archive, Records and Information Management in Europe. http://dlmforum.typepad.com/Paper_RemcoVerdegem_and_JS_CostModelfordigitalpreservation.pdf
[20] Zierau, E., Kejser, U. B., Kulovits, H.: Evaluation of Bit Preservation Strategies, In: Proceedings of the 7th International Conference on Preservation of Digital Objects, Vienna, Austria, pp. 161-169 (2010).

Author Biography

Anders Bo Nielsen holds a master in economics from the University of Copenhagen (economic history and IT as specialty) from 1997 and a master in IT (software development as specialty) from the IT University of Copenhagen from 2007. He is a principal consultant at the Department of Appraisal and Transfer at the Danish National Archives, where he has been employed since 1997.

Alex Thirifays is IT-archivist at The Danish National Archives. He obtained his MA in history in 1999. He has been working as an IT consultant for 6 years, and is now working with digital curation.

Ulla Bøgvad Kejser is preservation specialist, PhD, at the Royal Library in Denmark. She works with preservation imaging, preservation planning and curation of both physical and digital collections, especially images.

©2012 Society for Imaging Science and Technology