Deliverable D2.1.4

Tools for quantitative comparison of preservation strategies

Matthew Addis, Mariusz Jacyno, Martin Hall-May, Stephen Phillips
(IT Innovation Centre)

Document administrative table

Document Identifier	PP_WP2_D2.1.4_PreservationStrategyTools_R1	Release	1
Filename	PP_WP2_D2.1.4_PreservationStrategyTools_R1_1.00.pdf		
Workpackage and Task(s)	WP2 Models and environments for long-term audiovisual content preservation WP2T1– Models for Audiovisual Preservation		
Authors (company)	Matthew Addis, Mariusz Jacyno, Martin Hall-May, Stephen Phillips (IT Innovation Centre)		
Contributors (company)			
Internal Reviewers (company)	Adil Hasan, University of Liverpool. Tom Heritage, BBC (reviewed workflow section).		
Date	26/10/2012		
Status	Draft		
Type	Deliverable		
Deliverable Nature	Prototype		
Dissemination Level	Public		
Planned Deliv. Date	M45 - 31/09/2012		
Actual Deliv. Date	M46 - 26/10/2012		

Abstract

This report describes the tools developed by IT Innovation for quantitative comparison of preservation strategies. The tools have been open sourced and are publicly available on a website which includes documentation and features for bug reporting and new functionality requests.

DOCUMENT HISTORY

Version	Date	Reason of change	Status	Distribution
0.5	19/12/2011	First release v0.50	Release 0	Confidential
0.1	01/10/2012	First Draft of new release	First draft	Confidential
0.2	05/10/2012	Addition of material	Second draft	Confidential
0.3	25/10/2012	Revision in response to internal reviewer	Third draft	Confidential
1.00	26/10/2012	Finalised	Release 1	Public

Table of contents

Scope

The European Commission supported PrestoPRIME project (http://www.prestoprime.eu) has researched and developed practical solutions for the long-term preservation of digital media objects, programmes and collections, and found ways to increase access by integrating the media archives with European on-line digital libraries in a digital preservation framework. This result will be a range of tools and services, delivered through the PrestoCentre.

This report describes tools developed by the IT Innovation Centre for modelling and simulating a range of audiovisual preservation strategies, in particular the use of IT systems for the ingest, storage, access and processing of file-based audiovisual assets.

The tools are available online at http://prestoprime.it-innovation.soton.ac.uk and have been widely disseminated to the audiovisual community through international conferences and journals, blog posts on the PrestoCentre website, PrestoCentre training events in Paris and Los Angeles, and demonstrations and user sessions at the PrestoPrime Testbeds.

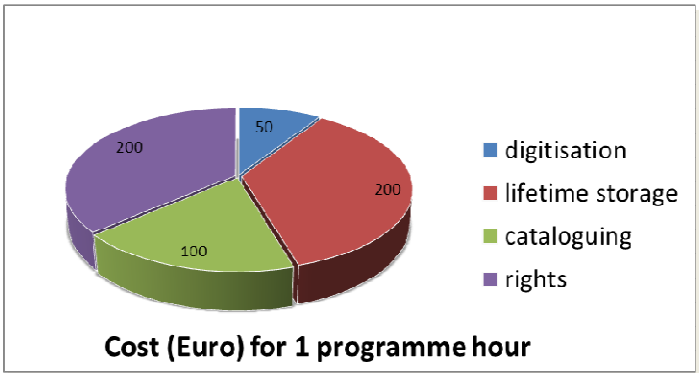
1 Key findings, recommendations and conclusions

"Digital information lasts forever - or five years, whichever comes first."
Jeff Rothenberg

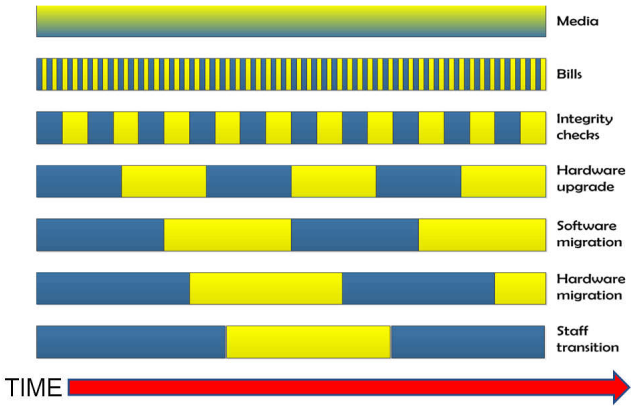
"Eternity is a long time - especially towards the end."
Woody Allen

This report describes an approach to predicting the costs of preservation activities needed to create and keep digital audio-visual (AV) content accessible for the long-term, which can be 20, 50 or 100 years, or more. The focus of this report is the costs of ingesting, storing and accessing file-based content using IT systems and the associated people and processes needed to operate these systems.

For AV material where data volumes can be huge (Petabytes of data for a large archive), the long-term Total Cost of Ownership of storage, i.e. the 'lifetime storage cost', is a significant percentage of the overall costs. For AV material, despite the falling costs of storage, it is rare to be able to say that 'storage is effectively free' - at least not yet! Other costs, e.g. digitisation of analogue material, cataloguing or rights clearance might well have higher costs, but they are typically one off costs incurred during ingest or access. Storage on the other hand is an on-going cost. It is also one that can't be ignored – in the absence of everything else, files have to live somewhere and that somewhere is on storage, and that storage costs money.



Costs of different activities for preserving AV content

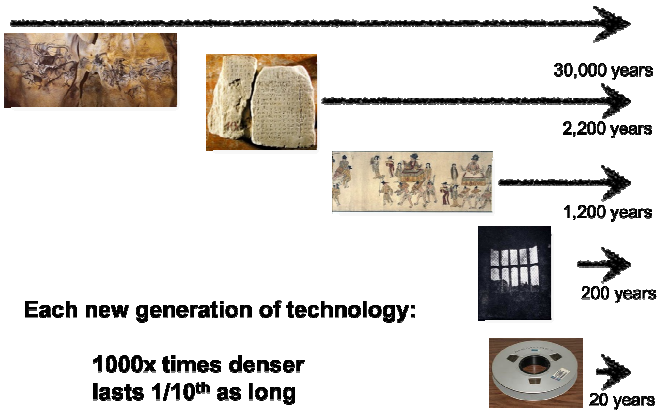


Long-term storage is a series of changes

Over the long-term, storage using IT systems includes the need for various preservation activities, for example media migrations and fixity checks. These activities keep files 'alive'. Preservation of files using IT systems is very much something that requires an active approach. A failure to be pro-active, a failure to invest in storage, and a failure to maintain the investment needed over the lifetime of the content being stored, all puts content at risk of loss. This makes long-term planning and cost estimation of storage an important part of preservation.

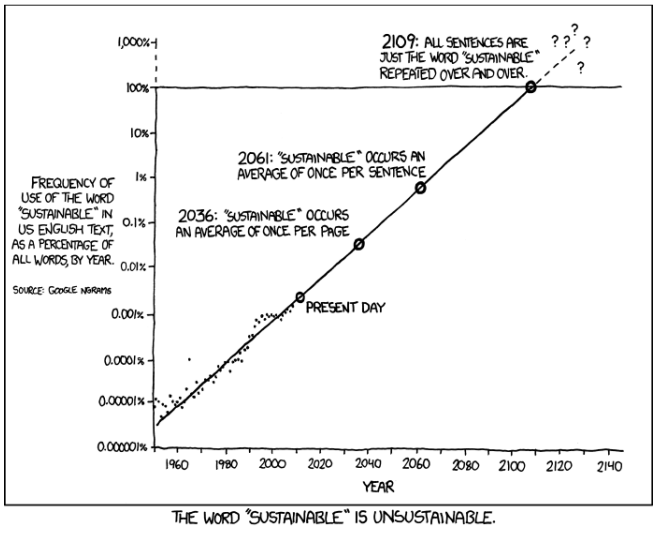
Many AV archives are driven by a mission to make their content accessible, including online, and to do so for a very large volume of content - but with a limited budget. This is over and above the core need to preserve the content so it remains accessible in the future.

The use of IT technology offers the potential to drive costs down, in particular by offering archives the ability to take advantage of the advances to technology that are driven by the much wider and global explosion in data volumes that humanity is creating and processing for scientific, medical, environmental and social networking to name but a few. However, IT systems are not 100% reliable and there are inevitable trade-offs to be made. New technologies offer the ability to store ever larger volumes of data for the same cost, but their lifetime is limited, their reliability a concern, and new and ever more sophisticated techniques and processes are needed to deal with these limitations.



Long-term storage: a series of new technologies

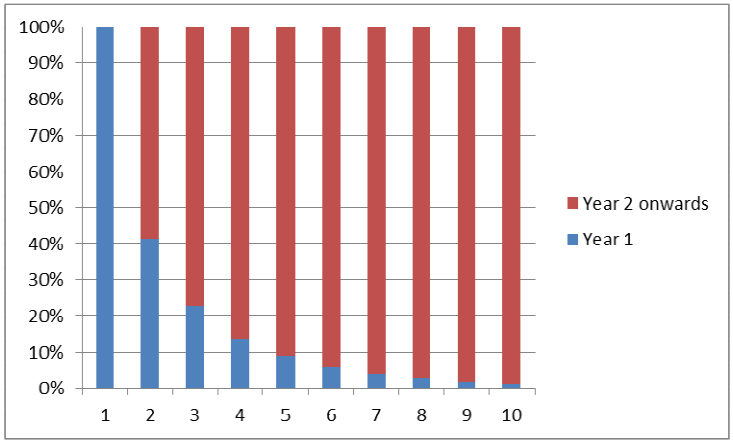
There is no ‘silver bullet’ option. No single technology offers high levels of data safety, long lifetime, fast access, minimal intervention, and low cost. Making too many compromises in order to drive costs down has the counter-effect of increasing the ‘risk of loss’ of content in the long-term. There are many technology options to choose from, each with its own balance of cost, safety, longevity, access, and expertise to install and use it. There is also preservation best practice, which essentially comes down to using ‘diversity’ as a way to mitigate the long-term risks of content loss. Make multiple copies of content, in multiple locations, using different technologies, and ideally make them independent and managed by different teams of people or organisations. The approach becomes one of designing a ‘preservation system’ that follows these principles, and one that meets an organisation’s specific set of requirements for content safety and accessibility - and of course cost.



Past trends are not always an indicator of the future
(reproduced from XKCD HTTP://XKCD.COM/1007/)

Such preservation systems can be designed today, and several organisations have successfully implemented them, but these systems are also short lived due to rapid obsolescence. They need refreshing and replacing. The question then becomes one of what systems will exist in the future and what will their characteristics be. It is tempting to use historical trends in technology improvement and extrapolate them into the future, for example following Kryder’s law of increasing storage capacity for a give cost. However, not only are there fundamental limits that prevent this trend continuing long-term for many technologies, e.g. hard drive based storage, but as the recent tsunami that wiped out key global manufacturing plants in Thailand illustrates, there can be disruptions to the overall trend that take years to recover from and significantly impact long-term cost projections.

Alongside this comes the need to consider all elements of Total Cost of Ownership (TCO), including power, cooling, space, maintenance, and people costs. These factors can have different long-term trends which means they can become more dominant percentages of the TCO in the long-term. The approach to dealing with all this is called Discounted Cash Flow with different discount rates applied to the different elements of the TCO.



Today's content is set to become a marginal fraction of the total content held by archives given high CAGR

It's not just the trends in TCO for storage that needs consideration, but ingest and access too. Growth rates in archive content and how often the content is accessed have a big impact. There are significant economies of scale when using IT based approaches. These come through automation and using large-scale (enterprise) solutions. But what's considered 'large scale' today rapidly becomes 'small scale' tomorrow unless the rate of growth of archive content and access at least matches the rate at which technology advances. In other words, archive growth rates have significant bearing on long term costs and what might seem like a huge volume of content that needs preservation today can actually become marginal (and hence a marginal cost) compared to the volume of content that needs to be preserved in the future. Long term costs become dominated by long-term growth, or lack of it. There are likely to be limits to this, however. Taking today's Compound Annual Growth Rate (CAGR) and extrapolating too far into the future is potentially just as flawed as doing the same with Kryder's law for technology– there are fundamental limits that will eventually come into play – not least the limited capacity of mankind to 'access' AV material. A fixed number of eyeballs on the planet available to view AV content may ultimately limit the rate at which it is produced and the quality it needs to be produced to. What is true is that there are many uncertainties in long-term trends, including capabilities of technology, the elements of TCO that will dominate in the future, how volume of content will change over time, and how much of that content will be accessed. Uncertainties in these trends generate the largest uncertainties in long-term costs or risks of loss.

Given all the issues that have to be considered when estimating long-term costs of ingest, storage and access, and having done a detailed analysis of costs and risks, we make the following recommendations for those developing cost models in this area.

When building a model, the following overall approach should be taken:

- The model should be for the 'system as a whole' including the interactions between the component parts (storage, processing, networking) and all the functions that need to be supported (transcoding, replication, fixity checking etc.)

- Best preservation practice should be embedded in the model, e.g. multiple copies of content stored in different locations using different technologies.

- The model should include the ability to simulate events that can occur during the lifetime of the system, e.g. migrations, failures, access requests, changes in ingest rate. Events can be both scheduled, e.g. a migration, or stochastic, e.g. failure of a storage server.

- The model should allow trade-offs to be analysed, e.g. the trade-off between cost and content safety (cost of risk of loss) or the trade-off between having limited resources (people or equipment) and the ability to meet ingest and access needs.

When modelling the costs of storage systems, the following need to be considered:

- TCO of storage needs to include the trends over time for each of the cost components (e.g. media, servers, power, space, cooling, maintenance, people).

- TCO needs to include periodic activities maintain accessibility and keep costs down (e.g. media refreshes and server migrations).

- Projections need to be made of data volumes and rates of data ingest and access over time, i.e. the usage levels and the storage capacity needed.

- Net present value and discounted cash-flow techniques should be used to allow future costs to be converted into ‘today’s money’ in order to allow proper comparison of different options including Pay As You Go and Paid-Up models.

When modelling the failure types and modes that can put content at risk, the following need to be considered:

- Latent and extant failures can happen from a wide range of causes (e.g. media, systems, people).

- Failures can occur when writing data to storage, reading data from storage, or retaining data in storage

- Failures can be either data corruption or data loss and can occur at a range of scales, e.g. bit, byte, sector, block, drive or system levels.

- Failures where one data object is damaged can happen when other data objects are being read or written, e.g. misdirected writes for a HDD array or damage to a data tape in a drive.
- Failures can be correlated or random, and the trends for the rates of these failures will evolve over time.

- Rare events can happen with major impact, e.g. fire, flood, storage server crash, malicious attack. Resulting data loss can propagate between systems, e.g. through replication or human errors.

In order to model the interaction between cost and failure modes or other factors that can cause content loss, a risk management approach should to be included:

- The ability to model a broad range of risks that arise when using storage systems for data retention and access.

- Inclusion of countermeasures that address these risks, e.g. fixity checks and repair of corrupted data from replicas, including the costs of these countermeasures and the residual risks that will remain after their application.

- The scheduling of countermeasures (e.g. scrubbing) on a regular basis and the triggering of countermeasures based on events occurring (e.g. detection of data loss triggering a repair).

Having implemented a model based on these principles, we make the following conclusions:

- Considerable effort is needed to implement models that are both detailed enough and flexible enough to describe a wide range of ingest/storage/access scenarios. Our iModel simulation tool took approximately 18 person months of development effort to implement.

- Stochastic and 'Monte Carlo' approaches are invaluable when exploring a range of different options or understanding the range of outcomes that can occur because of random events (e.g. failures in storage).

- Systems with adequate resources at the outset can quickly become overloaded in scenarios of rapid content growth. At this point, many problems start to manifest, including increased rates of content loss and delayed ingest or access.

- Replication and proactive integrity management strategies, e.g. periodic fixity checks and repairs, can have a big impact on reducing long-term content loss and are essential features of a preservation storage system.

- Activities associated with ingest, access, storage and preservation all use resources and all take time. Modelling of limited resources and finite execution times is essential to reveal bottlenecks, under or over provisioning, and the impact of unexpected events, e.g. timescales for disaster recovery.

The rest of this report contains a more extensive review of how to estimate the costs and risks of loss when using IT systems for ingest, storage and access of file-based digital content. This includes how our approach of building a simulation tool fits with other strategies for cost modelling, e.g. empirical models based on content lifecycles and cost estimation based on historical data.

Finally, we make one last observation:

- Costs and the risk of loss need to be balanced by the value of preserving content and making it accessible. This requires a wider cost/benefit analysis, which is unfortunately beyond the scope of the work described in this report.

2 Overview of tools

Three tools have been developed during the PrestoPRIME project.

1. **Interactive simulation of ingest, storage and access** (iModel), which is a simulation tool for investigating in detail different strategies for storage, transcoding, ingest, access and file-format migration of digital audio visual assets.

2. **Web-based storage planning tool**, which designed to support decision making on what storage strategy to use, for example how many copies to make of files in archive, what storage technologies to use to hold them, and what measures to take to maximise the long-term integrity of these files.

3. **Simulation of digitisation and transfer workflows** (iWorkflow), which is a simulation tool for digitisation/migration workflows of discrete assets (e.g. digital video tapes) and has been developed for a specific scenario at the BBC for their D3 project. More details are in the appendix to this report.

2.1 Interactive simulation of ingest, storage and access (iModel)

The interactive simulation tool (iModel) provides a comprehensive facility for estimating the costs and risks of using IT systems for storing, accessing and processing audiovisual assets.

The tool is intended to allow a wide range of questions to be considered when planning, selecting or operating such storage and access systems. Just some of the questions that the tool can help answer include:

- When storing content, how many copies should be made, what technologies should be used, how much will it cost, what are the long-term risks of losing files?

- What impact does the choice of codec (e.g. compressed or uncompressed video) have on costs and risks?

- What are the pros and cons of just in time generation of access copies compared to creating and storing a full set of proxies in advance?

- When storing data, how often should it be checked to make sure integrity is intact, and when does this become counter-productive (e.g. act of checking causes more damage than it might repair)?

- When should media migration take place (e.g. between LTO generations): regularly or at the point of obsolescence?

- What is the impact of ingest and access on shared resources for storage and data safety: what level of resources is needed to support both?

2.2 Web-based storage planning tool

The simple web-based tool for storage planning is available to use on an IT Innovation hosted website (http://prestoprime.it-innovation.soton.ac.uk) and is described in detail in PrestoPRIME deliverable D2.1.2 Preservation Modelling Tools, which is publically available. The tool has not been extended or changed since the release of D2.1.2 and isn't described further in this report.

2.3 Simulation of digitisation and transfer workflows

The simulation of digitisation and transfer workflows was an investigation into what level of detail is needed to model the operational characteristics of 'factory' scale transfer set-ups.

For the simulation of migration workflows, e.g. the BBC D3 project[1] used as the basis of this particular tool, it is very difficult to build a general purpose tool that can be applied to the diverse range of workflows that archives use. Instead, each migration scenario needs to be modelled on a case-by-case basis. Only in this way can the model capture the necessary detail to allow use forecasts to be made

The question becomes one of how much effort does it take to create a bespoke model of a digitisation or transfer workflow that adds real value to the organisation spending this effort? The organisation might be an archive that is planning to do a preservation project, or it might be a service provider that wants to cost-up and plan a project for a customer. The time and effort to do effective modelling is a critical factor on estimating the return on making this investment, especially when time and funds may be particularly tight in the planning stages of a project or when putting a business case together for investment.

For example, consider a large national archive that is about to make a business case or a detailed plan to migrate a 100,000 hrs of material from digital video tapes onto IT mass storage. This is likely to be a multi-million Euro operation that will take several years to complete. The archive could benefit significantly from a month or two spent developing a tool that allows them to explore and cost different options since this could easily save far more time and money in the implementation and operation phases. On the other hand, the return on investment for doing this for smaller scale preservation projects, e.g. a few hundred hours of material, would be a lot lower.

Without knowing what level of modelling detail can be achieved with what level of effort and expertise, it is very difficult for archives to make an informed judgement on whether to invest in this type of activity. Investigating in PrestoPRIME what it takes to model digitisation/migration workflows helps answer this question. It also allows some examples to be provided to the community on approaches to take and results that can be achieved.

The specialisation of the tool to the BBC workflow being modelled means that the tool is not available for general download and use outside of the PrestoPRIME project. Instead, the findings of the work are presented in this report as information for others who are also considering detailed modelling of current or future transfer projects.

[1] http://www.bbc.co.uk/rd/publications/whitepaper155.shtml

3 iModel

3.1 Why the need for a cost and risk simulation tool?

Audiovisual (AV) archives are making the transition to IT storage systems as the basis of long-term retention and access to their assets. As a result of digitizing analogue holdings and new 'born digital' production processes, audiovisual archives face the challenge of how to preserve and manage huge file-based repositories of digital audiovisual assets.

For example, 1 hour of high definition (HD) programme material stored in an uncompressed format will typically require 1TB of storage. A large national broadcaster will typically generate several hundred hours of new content each week. Preservation best practice recommends that audiovisual material is stored either in its native digital file-format if there are no immediate risks of format obsolescence, or otherwise is converted into uncompressed or loss-less compressed preservation formats (e.g. WAV for audio, uncompressed or JPEG2000 lossless for video, TIFF for scanned film frames). Irrespective of whether content is in high-bit rate professional production formats or is converted into long-term preservation formats, the storage requirements are onerous and the costs are a major part of an archive budget.

The requirement for long-term safe keeping and access to this material mean that archives are looking at 50 year horizons (or more) over which they need to be sure that their digital assets will be safe. This horizon and requirement for long-term data safety is a non-trivial problem given the rapid obsolescence of storage technologies (3-5 year cycles) and less than ideal reliability and data integrity characteristics (e.g. component or system failure rates and so called 'bit rot').

Preservation best practice suggests the solution is to make multiple copies of content, in multiple locations, using a diverse range of technologies for storage, and to monitor/migrate frequently in order to address inevitable failures and obsolescence. This guards against a wide range of risks, but archives typically do not have the budget to achieve the degree of data safety and accessibility that they desire. Long term retention and access to AV assets inevitably becomes a compromise in order to achieve a balance between acceptable costs and acceptable risk of loss.

What archives lack is a way for these compromises to be objectively and quantitatively assessed. Whilst designed with AV archive scenarios in mind, we believe the tools and techniques we have developed are applicable to a wider class of storage simulation and planning scenarios and are of interest to a broad audience.

3.2 Review of state of the art in cost modelling

Much work has already been done on the cost and reliability of storage systems, including for preservation of audiovisual content [1]. From a cost perspective, Google [2], the San Diego Super Computing Centre [3] and others have reported the Total Cost of Ownership (TCO) of storage. These studies typically present a breakdown of costs into media, servers, power, space, cooling and maintenance (which includes labour costs). For long-term storage, a further set of factors need consideration. These include the falling cost of storage (effectively Kryder's law noting that whilst the capacity per unit of storage such as a data tape or HDD may double every 18 months or so, the cost doesn't and therefore the

effect is that the cost per TB will halve at the same rate). The rate of fall of the TCO of storage is slower than the trend for media because of the other cost elements, but still has a downwards trend. This is evidenced by the long-term price per TB charged by Amazon S3 as a storage service provider which has halved every 2-3 years. Short-term disruptions to the cost of storage, such as the flooding in Thailand in 2011 which halted manufacturing of HDD by Western Digital, can have long-term consequences on the lifetime cost of storage. David Rosenthal has modelled this effect [8] in terms of the endowment needed today to fund 100 years of storage with a high probability (98% in his example) that the money will be sufficient.

Other factors that come into play are the need for migration to counter technical obsolescence or storage failure rates. There has been work in this area for some time, especially the comparison of the long-term TCO of data tape vs. HDD, for example the recent Clipper Group [10] and Enterprise Strategy Group [9] reports, which are valuable in terms of their enumeration of all the factors rather than specific findings. The extreme of long-term cost modelling is the emerging body of work on ‘forever costs’, for example the work by Princeton University [11], which uses simple model of periodic storage system migration where each successive system falls in price according to Kryder’s law. The result is a prediction of the lifetime cost of the storage of data as a multiplier of the year 1 cost. Similar models have been developed for storage of digital media, e.g. by Sun [14] and AMPAS [13], although the conclusions on the cost of digital storage are radically different ranging from ‘half the price of analogue’ to nearly ‘twelve times higher'.

Further factors affect long-term costs of storage, or more importantly the amount of money that is needed to cover these costs over the lifetime of the data concerned – especially if this money is allocated today using an ‘endowment’ approach where future interest rates can have a big influence [15]. A particularly important factor is the data volume that needs to be stored as a function of time. The cost of storing data is a function of the amount of data to be stored. For example, consider the marginal cost of adding 1TB to an already large tape library compared with the cost of only having 1TB to store using a dedicated HDD server. In general, the larger the volume of data to be stored then the lower the cost per TB of storing that data. This can be seen both in the capital expenditure and operating costs of an ‘in house’ storage solution or in the pricing tiers of online service providers (e.g. at the time of writing, Amazon’s S3 storage service is $0.125 per GB per month for the first TB, $0.08 per GB per month when over a PB is being stored, and then $0.055 per GB per month for 5PB or more of data).

What is considered a large volume of data today, and hence attracts a lower cost per TB because of economies of scale, will rapidly become a relatively small volume of data in the future (as seen in Kryder’s law). Therefore, the cost of storage of data only remains relatively low if the volume of data to be stored grows continuously so that the data volume remains large in relative terms.

What has become clear is the requirement for a long-term cost model to include:

- TCO of storage and trends in each of the components (e.g. media, servers, power, space, cooling, maintenance).
- How TCO varies with time (e.g. media refreshes, server migrations, proactive measures to ensure data safety e.g. scrubbing).
- Projections of data volumes and rates of data ingest and access over time, i.e. the usage levels and storage capacity needed.

- Financial parameters to allow long-term budgeting, e.g. interest rate trends that allow the use of net present value and discounted cash-flow techniques.

Cost modelling for long-term preservation and archiving is a rapidly moving area. A more detailed look at some recent work in this area is provided in Section 0(Appendix on cost modelling).

3.3 IT storage: not a perfect solution for long-term data storage

There are many reports on the reliability (or lack of it) for storage technology and storage systems, including the types and origins of failures [4], mostly for Hard Disk Drive (HDD) based systems, supported by field studies and evidence of failure rates seen in practice [5] [54] including for AV archives [6].

It is natural to think of failure modes in terms of the media or underlying physical devices used for storage, for example HDD failures as examined in the analysis by Google of their storage infrastructure [16], but data corruption can take place in all types of IT storage and at all levels, including in systems explicitly designed to prevent it, for example RAID arrays of HDD [5]. The study done by CERN [17] highlights many issues seen in practice with large scale storage, including firmware bugs, and failures in other links in the chain of getting data to and from storage, e.g. imperfect computer memory and incorrect error propagation at the file system level [55]. This corruption can be both silent (otherwise known as latent corruption or 'bit rot') and permanent. Most worryingly, corruption events (in terms of checksum mismatches) have been found to be non-independent, both on the same disk and between disks within the same storage system [53].

A good overview of latent and extant faults and their causes was provided in [18] in 2006. This and subsequent work has looked at how failures translate into various metrics for characterizing the 'safety of storage', for example Mean Time to Data Loss (MTTDL) [18], 'bit half-life' [19], Mean Latent Error Time (MLET) [20], and Normalized Magnitude of Data Loss (NoMDL) as described in the paper "Mean time to meaningless: MTTDL, Markov models, and storage system reliability" [21]. The latter serving to highlight the challenges involved in defining a meaningful metric and a realistic underpinning storage model.

It should be said that corruption levels both published by manufacturers and seen in practice are actually remarkably low, which is testament to the levels of engineering in these technologies. For example, a modern hard drive has a Bit Error Rate of 1 in 10^{14} with LTO data tape being lower at 1 in 10^{17}. However, with 1TB requiring approx. 10^{13} bits of storage it becomes clear that errors are inevitable at the PB storage scale seen in many large archives. The problem is that whilst the capacity of a storage system of a given cost will typically double every 18 months or so, the rate at which this system can be loaded with data and the rate at which errors will occur do not keep pace with this trend. This results in increased data integrity recovery operations (e.g. RAID array rebuild times) and increases in complexity of error protection schemes (e.g. the transition from single parity RAID-5 [24] to double parity RAID-6 [25] with triple parity already on the horizon [26]).

The responses developed to cope with, or counter, failure modes and data corruption in storage are manifold. Techniques include RAID for storage arrays, ever more advanced techniques for dealing with latent errors in HDD storage as reviewed in [23], and a wide range of approaches to distributed storage, including erasure codes, e.g. as described as

the basis of Redundant Array of Independent Nodes (RAIN) [22] and more recently in archive applications such as Pergamum [27] and DAWN [29]. Simple replication techniques also have a part to play, e.g. as used by Lots of Copies Keep Stuff Safe (LOCKSS) [28]. These techniques can be combined, e.g. a layered combination of replication, integrity checking, erasure coding and other techniques are now seen in advanced file systems designed to manage data integrity from the outset such as ZFS [30].

Recognizing that it is possible to reduce the level of data loss at increased cost, but not eradicate it completely, a flexible and pragmatic approach needs to be taken to storage failure modelling that incorporates:

- Latent and extant failure modes with a wide range of causes (e.g. media, systems, people).
- Failures that occur when writing data to storage, reading data from storage, or retaining data in storage
- Failures that can be either data corruption or data loss at a range of scales, e.g. bit, byte, sector, block, drive or system levels.
- Failures where the data damaged is not necessarily the data being read or written, e.g. misdirected writes for a HDD array or damage to a data tape in a drive.
- Correlated or random failures with trends for how the rates of these failures will evolve over time.
- Rare events and analysis of their impact, e.g. loss of a whole array, propagated corruption between systems, human errors.

3.4 Risk and risk management

Risk assessment and risk management techniques are well suited to the use of storage technologies for data retention and access. Risks to content from IT storage technology come from many sources in addition to the failure modes discussed already, for example technical obsolescence and human factors (e.g. deliberate or accidental deletion of data, failure to budget properly, failure to follow data management processes). Short lifetimes of storage technology require frequent migration to avoid loss from technical obsolescence.
Risk management is a cyclic activity [31] of assessing and dealing with risk, including the selection and application of one or more treatments. Risk management as a methodology is ideally suited to assessing ‘whether IT systems are safe’ in the context long-term storage and access of data assets. Not surprisingly, application of risk management techniques is widespread in critical applications, e.g. information security [32]. In the digital preservation domain, the CCSDS (producers of OAIS) are currently combining the efforts of TRAC [33], DRAMBORA [34], Nestor [35] and ISO 27001 [36] to ISO standardize the results (Trusted Digital Repository Checklist ISO16363 [37]).

We have applied a risk assessment approach in previous work [38][39] that looks at the risks and counter-measures for a wide range of storage technologies (optical, HDD, data tape, printing bits to film and others). This work also considers the impact should a particular risk materialize. For example, even infrequent bit-level data corruption can have significant consequences on the usability of audiovisual assets. In effect, corruption is amplified, particularly if the file is compressed, e.g. studies show [40] that a single byte corrupted in a JPEG2000 image (lossless or lossy) can result in 30% or more of the

decoded pixels being affected and in many cases with major visual artifacts being visible across the whole image.

Therefore, when looking at the cost and safety of long-term storage, a wide range of risks need to be considered along with the impact on the utility of content should those risks materialize. Each countermeasure employed will have a cost associated with it, and in some cases the countermeasure itself, if applied too frequently, can be counterproductive and increase the risk. Consider data scrubbing as an example. The act of accessing data to generate a checksum can cause failures (e.g. increased wear, possibility of head crashes etc.). Therefore, checking data too often can be counterproductive, for example as presented by Mary Baker at MSST 2011 [42]. Checking data can also increase costs, especially in archive storage systems that keep data on media that is at rest, e.g. data tapes or spun-down disks in MAID type arrays. These costs need inclusion in the overall cost model of storage TCO over time. However, there is relatively little work that investigates the trade-offs that exist between cost and loss [43] [7] by looking at the relationship between the two over time.

In order to model the interaction between cost and risk, the following factors need to be included:

- The ability to model a broad range of risks that arise when using storage systems for data retention and access.
- Inclusion of countermeasures that address these risks, including their costs and the residual risks that will remain after their application.
- The scheduling of countermeasures (e.g. scrubbing) or the triggering of countermeasures based on events occurring (e.g. detection of data loss triggering a repair).

3.5 How the simulation tool works

3.5.1 Cost and risk combined

There are many approaches to long term preservation of digital audiovisual content. Each one has associated costs and risks as well as delivering differing degrees of content accessibility. Examples include HDD in servers, data tapes in libraries, archival grade optical media (e.g. Millenniata discs [51]), printing digital bits to analogue media (e.g. polyester film [52]) and many more. No single technique provides a complete solution. Many archives face the challenge of how to compare, assess and combine the options in a consistent way.

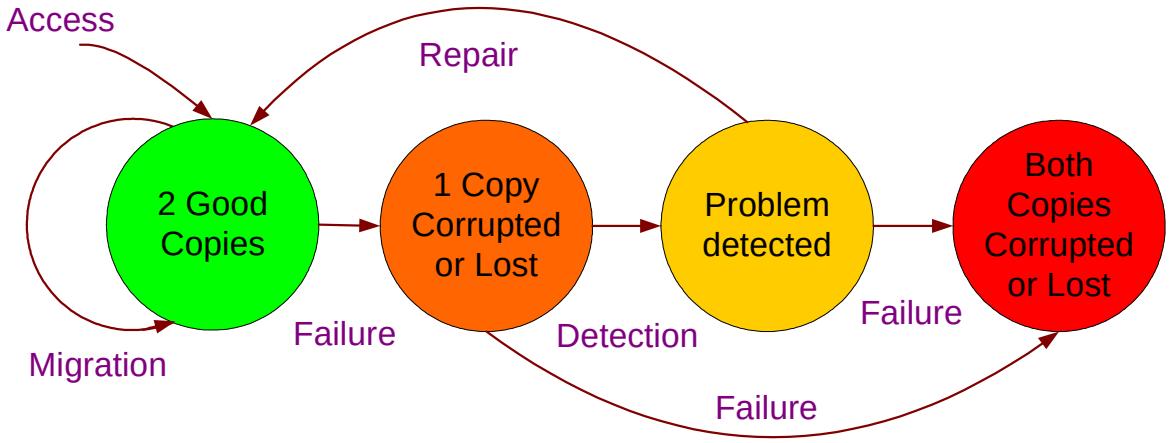


Figure 1 Conceptual model of storage, access, migration and integrity management

Figure 1 presents a simple conceptual model for analyzing cost and risk for data storage and access. With reference to Figure 1, the bedrock of data safety is to keep multiple copies of each data object (green circle), e.g. by using different technologies in different locations, and ideally operated by different people. This guards against major risks, e.g. by enabling disaster recovery, but also guards against unanticipated problems with individual technologies and processes, i.e. it ensures eggs are not 'all in one basket' at any level. The diagram shows the simplest version of this approach: keep two independent copies of each data object. Each copy is stored in a storage system of some description. One or more of these storage systems is used to serve requests (access) for data objects already in the systems, or to receive new data objects (ingest).

For each storage system used to hold a copy of each data object there is the need to regularly migrate each component of the technology stack (hardware, operating system, management software, formats etc.) to address technical obsolescence, media degradation, and to provide improved capacity or performance. At the same time as data is being stored, accessed or migrated there is always the chance that one of the copies is damaged or lost resulting from a failure in the corresponding system used to store it (orange circle). This can be modelled as a probabilistic process where risks (e.g. data corruption) are represented as probabilities of transitioning between the states. But only after the corruption is detected (yellow circle) can any action can be taken, e.g. to repair or replace the damaged or lost copy by using the remaining good copy. If at any time something happens to the second copy (the only remaining good copy), then there is a risk that both copies are permanently lost or damaged (red) – i.e. the data object is lost. This is of course a simplistic model, in reality there are many cases where both copies only become partially corrupted and the remaining 'good' sections of each can be combined to recreate a new copy that has no corruption.

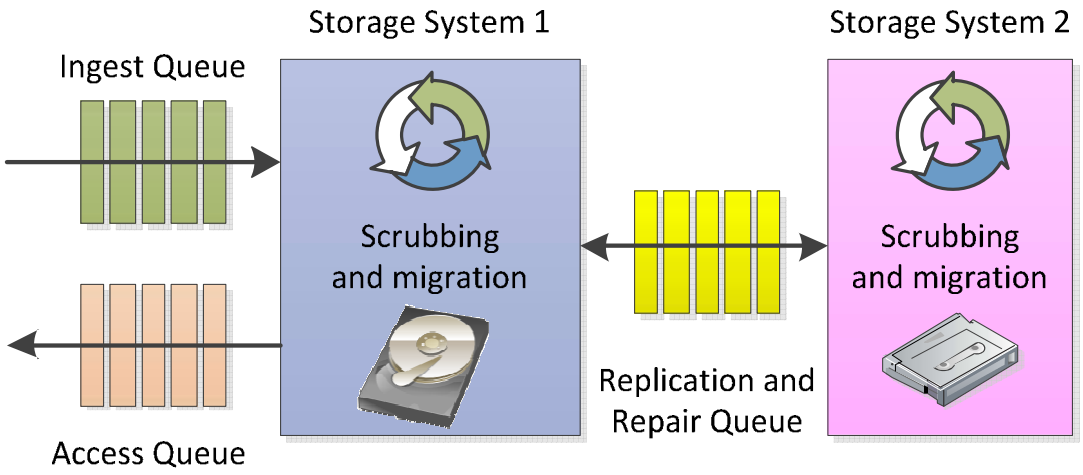


Figure 2 Example storage system configuration

Figure 2 shows an example storage configuration that might underpin the two copy model shown in Figure 1. The configuration consists of two storage systems (on each of which one copy of the data objects are held). Storage System 1 (SS1) is used to handle ingest and access requests and Storage System 2 (SS2) is used to provide a second safety copy of the data objects. Data objects are replicated from SS1 to SS2. In this configuration, SS2 provides a disaster recovery capability as well as a source of 'good' data to repair any corruptions or loss in Storage System 1 (and vice versa). The ingest, access, replication and repair activities are modelled as queues. Within each Storage System, internal processes check data integrity or perform local migrations, e.g. media or file formats. Each storage system has a cost associated with its operation, as do all state transitions in Figure 1. In this way, the model allows both the risks and costs to be combined into a single model. The model can easily be extended to include more than two copies of each data object. In the simulation tool we have developed, we can model up to 3 separate storage systems each of which can hold 1 or more copies of each data object.

The model shown in Figure 1 is similar to the Markov Chains often used to model failures in storage systems [49][50][18]. Although we use a discrete event simulation implementation, the behaviour of our system is degenerate with a Markov approach when simplifications are applied so that the Markovian 'no memory' property holds (e.g. no queues, failures are independent of each other). We use these simple cases to validate that the model is behaving correctly.

3.5.2 Data corruption model

Our approach to simulating data corruption or loss is to consider a storage system as having functions of: (a) accepting files for storage, i.e. writes, (b) returning files from storage, i.e. reads, and (c) storing the data inside using some form of physical media (hard drive, data tape, optical disc etc.). With the storage system is some form of 'controller' (manual or automatic) that mediates these processes.

The model can be applied to storage that is fully automated through hardware/software or to more manual process, e.g. 'data tapes on shelves'. When writing or reading files, various operations may be applied by a storage system, for example encoding or applying error correction. Depending on the system being modelled, this could be by firmware on a

HDD, the RAID controller in a HDD array, integrity management in a ZFS filesystem, manual integrity verification by an operator - or a combination of these. Likewise, various failures or errors can occur, both latent or extant, which range from 'bit rot' in a HDD system through to accidental damage from manual handling of discrete media, e.g. a data tape. These can happen (a) when data is written, (b) when data is read, and (c) when the data on the physical media is in effect 'doing nothing'. These are represented as error rates for read, write and store actions. This simple but flexible approach allows a wide range of storage approaches to be included in the simulation provided that they can be characterized in terms of failure rates. Rather than attempting to model the detailed mechanics of how a storage system physically holds data objects (files in our case), we consider the type of corruptions that occur and their impact at the file level. Conceptually a set of files to be stored can be represented as a set of files F={F1, F2…Fn} as shown in Figure 3. This assumes that the files are laid out contiguously on the storage medium.

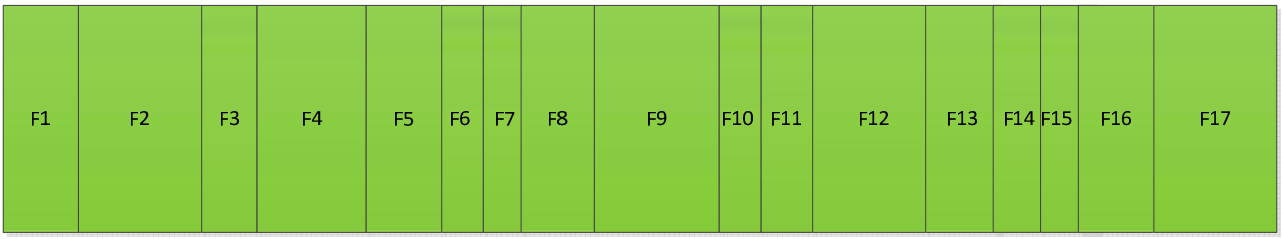


Figure 3 Representation of a set of files within a storage system. The files are labelled F1, F2 etc. The width of each file represents its file size.

File corruptions are modelled by overlaying corruption patterns on top of the set of files. For example, random bit level corruption can be visualized as a series of bit flips aligned against the line of files as shown in Figure 4.

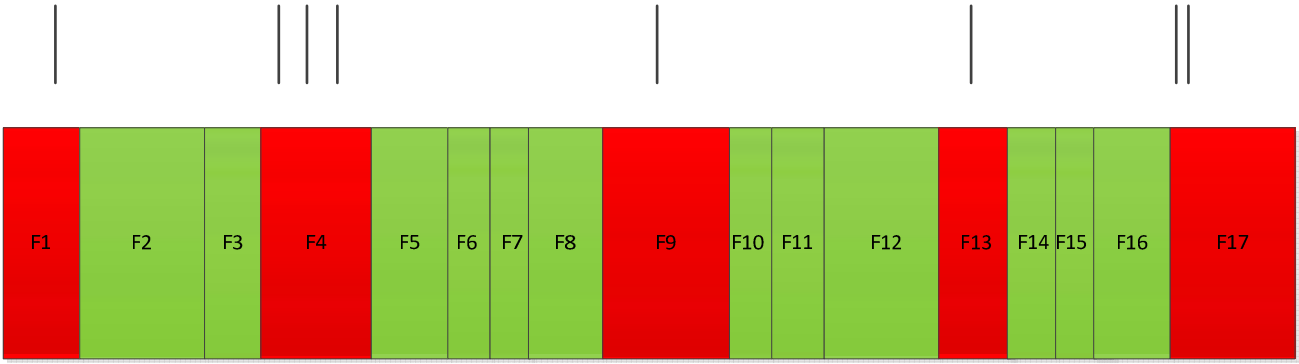


Figure 4 Example of bit level corruption (thin lines at top of the figure). Where one or more corruptions align with a file then the whole file is considered as corrupted (shown in red, e.g. F1, F4 etc).

Where corruptions are more than a single bit in extent, e.g. RAID blocks, then the corruption might hit a single file, or it might span multiple files. The approach of aligning corruptions of varying extent with the set of files can be extended to much larger scale failures, for example modelling a whole HDD failure or whole data tape failure as shown in Figure 5.

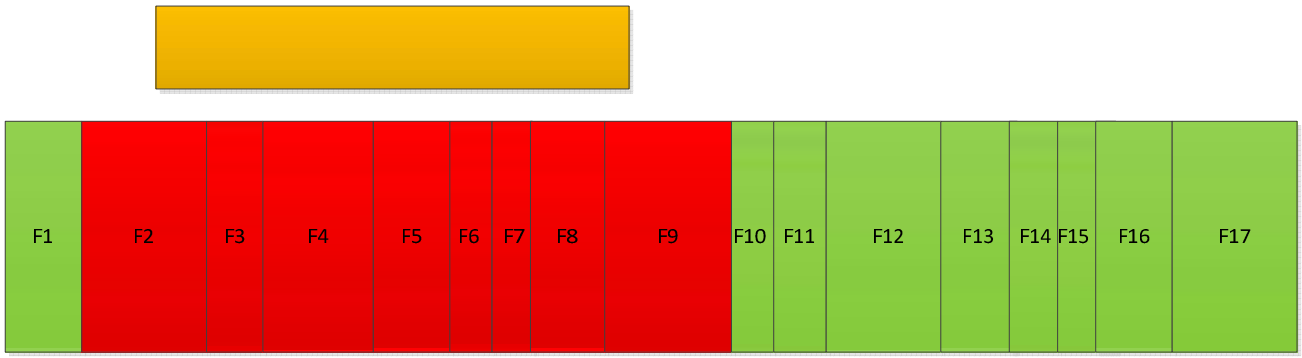


Figure 5 Example of a corruption that has a large extent and impacts a set of files, for example a data tape failure.

A set of corruption patterns can be combined to model the corruption of a set of files from a range of causes, for example as shown in Figure 6.

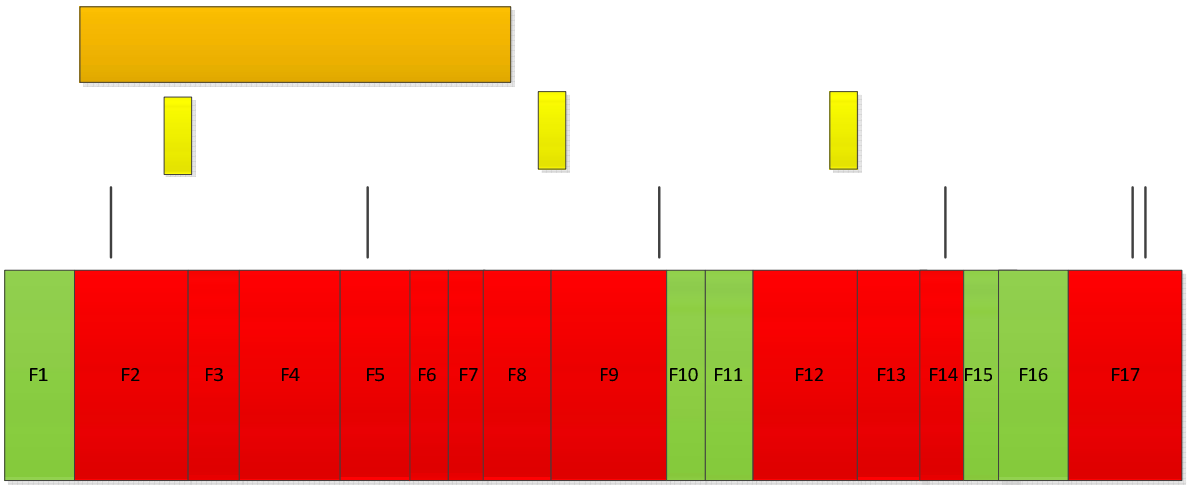


Figure 6 Multiple corruption types applied to a set of files, for example failures at bit, block or device levels.

The final part of our corruption model is to address the issue that the usability of a file or the ability to repair it in some way will depend on where in a file a corruption takes place and to what extent. For example, in the case of many audiovisual files, the corruption of the header of a file or key metadata in the file makes the whole file unusable. However, small scale corruption of other parts of the file, such as bit flips within a video stream can sometimes be repaired (e.g. using interpolation from adjacent video frames) and hence these types of corruption are less critical. We model this in terms of a file having a size f where a fraction p_c is 'critical' in the sense that any corruption of this will render the whole file as irrevocably damaged. The rest of the file, $f(1-p_c)$, is 'non critical' in the sense that provided that less than x% of this part of the file is contiguously corrupted then the file is considered as 'repairable', although at additional cost. Of course, if another copy of the file exists then this would be the default route to repair, i.e. it would be merged with or replace the corrupted file so full integrity is restored.

3.5.3 Storage migration model

Long-term retention of data on IT based mass storage system inevitably involves multiple migrations. Typically this takes place by having two systems (old and new) that operate

side-by-side for a period during which files are migrated from one to the other. Migration takes time and resources. During this time one or both of the storage systems still need to service access requests to the data as well as ingest any new data that needs to be stored.

In our model, a series of storage system generations can be represented on a timeline as shown in Figure 7. In this example, files are migrated between storage system SS1 and its replacement storage system SS1'. The two storage systems overlap in time (during which migration takes place). The green lozenges represent the number of files in the two systems and shows how files are moved from SS1 to SS1' during the window of migration.

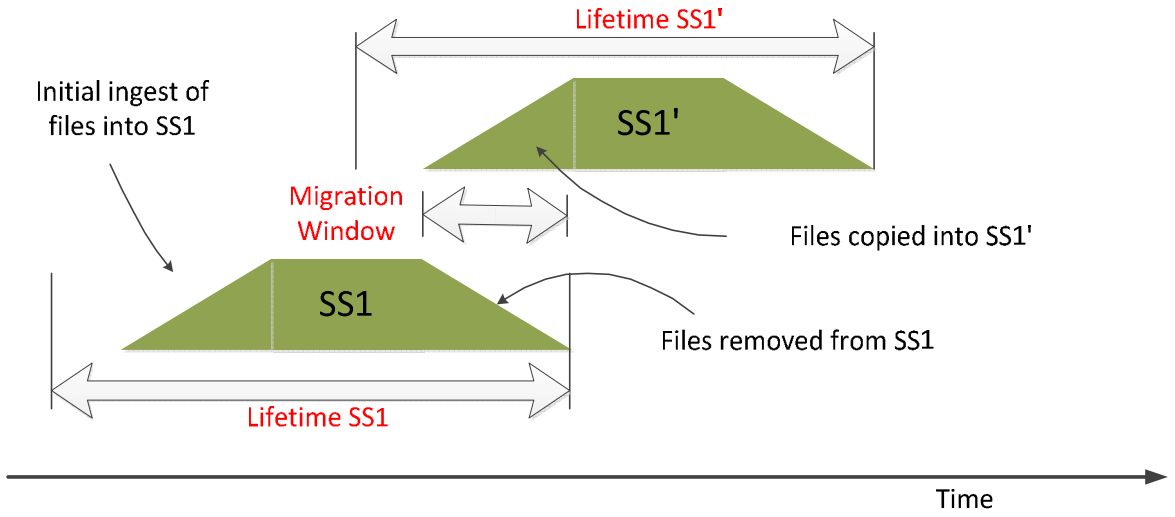


Figure 7 Model of storage system migration from storage system SS1 to storage system SS1'. Height of the green lozenges represents the data volume in a storage system.

Each storage system has a set of parameters. These include: start date when the storage system is first commissioned; lifetime of the storage system; system cost (capital equipment cost and amortization period plus operational cost per unit time); cost of adding data; cost of storing data per unit time; and cost of accessing data. The start date can be before data is migrated into the storage system, e.g. to model set-up time. The amortization period can be less than the 'lifetime' of the system, e.g. amortization over 3 years for a system that has an operational life of 5 years. The capacity of the storage is 'elastic' and is assumed to have no limit, i.e. as much data can be stored as desired. For a migration is between systems SS1 and SS1' the migration starts at T1 and it ends at T2, where the user of the model sets the date for T1 and then T2 occurs when the simulation has finished the migration. If T2 extends beyond the lifetime of S1 then the lifetime is extended until migration is complete (costs are still incurred, e.g. opex and data storage). If T2 occurs before the end date of SS1 then the end date of SS1 is brought forward to T2. Files are copied between SS1 and SS1' in-between times T1 and T2. Ingest of new files that come into the archive after T1 go to SS1'. Files are deleted from SS1 straight after a copy is made to SS1'. Access to files is serviced by SS1 or SS1' depending on whether the file has been migrated yet or not. The model described is simple, but able to describe the main characteristics of many types of storage. For example, a migrating between LTO generations in a tape library, or migrating between storage servers for a HDD solution. There is an initial capex (capital expenditure) cost and then growing the storage means extending the system with disk cabinets which can be modelled as having a cost that is proportional to the amount data being stored. Tape libraries and media migrations can also be modelled. Tape libraries (frames, robotics, etc.) have high capex but can last for say 15

years. In this time the drives and media can be upgraded, e.g. following the LTO roadmap. This means being able to model a migration between LTO generations within the same library. We do this by having the capex in SS1 and then migrating to SS1' where the data storage costs are halved and there is no further capex.

Modelling cloud storage is an interesting case as there is no explicit migration from a user perspective. One approach is to model a cloud service as a series of discrete storage systems. In some ways this works well: cloud prices fall, but in big steps and not that often. However, costs should not be artificially incurred when migrating. We therefore set the 'add data' costs to zero for existing data and only incur costs for new data that is added to the cloud storage.

3.5.4 Simulating events

The interactive simulation tool uses a discrete event simulation approach. The simulation contains one or more storage systems, each of which is modelled as providing a set of services (e.g. ingest, access, checksum validation). Several storage systems can be composed to simulate a hybrid storage system, such as HSM. Each service uses one or more resources (e.g. copying data, checking integrity). Requests to use a service are added to a queue for that service (e.g. queue of files to be ingested) where each request is then taken from the queue for processing if sufficient resources are available. The queues are 'first-in-first-out', with the ability to prioritize user access to archive content above background processes of migration and fixity checking if needed.

During the simulation, time ticks away and events are generated (e.g. random corruption of files in a storage system, requests to access a file, new files to be added to the archive). These events can trigger actions, e.g. a copy/repair process might be triggered if a file access event identifies that a copy of a file is corrupted. These actions then are added to the relevant service queues (e.g. file access queue for access events, file copy queue used as part of a repair process or scheduled file migration).

A storage system will process items in the queues for its services according to how much resource it has available (e.g. serving access requests sequentially or in parallel). The available capacity of the resources used by each service determines how many items in the queue for that service will be processed for each tick of the clock. If there is insufficient resource then not all items in a queue will be dealt with and the unprocessed items remain in the queue and are carried over to the next tick of the clock.

For a simulation of more than one storage system, a series of interactions are defined between storage systems, for example replicating files. In this way, the services for the storage systems become coupled. For example, if storage system 1 is used for ingest of files and the policy is to replicate those files to storage system 2 and storage system 3 before ingest is considered successful, then the rate at which items will be processed on the ingest queue is dependent on the copy resources available to create replicas of the file on the other storage systems. A set of template configurations are provided that correspond to common patterns for real world storage configurations, e.g. mirrored servers, HSM, online primary storage server plus deep archive for disaster recovery.

The core of the simulation is a relatively simple one – a set of services with queues and resources, a set of event generators and a set of template configurations for how storage systems are connected together.

On top of the core simulation is the user interface that allows the user to set parameters, interact with the simulation, and view results. This is where specific UI features are used, e.g. sliders, radio buttons, auto scaling graphs, easy tabbing between storage systems – all of which are designed to make the tool easy to use and tailored to the problem of cost and loss simulation.

The cost model used by the simulation is based on the premise that use of resources by each service will incur a cost (e.g. ingest, access and storage all have a cost). The cost is accumulated as the clock ticks. By attaching costs to resources, the different costs for each storage system can be accounted, e.g. resources used for copying files, checking their integrity, performing local repair or providing access. This allows the simulation to be easily extended if needed by simply adding further resources and costs. For example, should the model need to include the costs of archive activities such as cataloguing or rights clearance then these can be added to the ingest service. The tool is implemented in Java and is available online as open source (LGPL license) [44].

Existing simulation frameworks were considered (e.g. Simul8 [46], iGrafx [45], SimEvents [48], PRISM [47]). Whilst some are able to cover the core of the simulation, they all have difficulties when it comes to building custom user interfaces, using non-standard probability distributions or queue disciplines, and allowing user interaction and changes to the settings during simulation. These factors would make the tool hard to develop on one of these platforms and in particular hard to extend to include more complex functionality. There is also the major problem that these frameworks are mostly commercial and expensive to license which would significantly limit the ability to provide the tool to the community to use for free.

3.5.5 Data storage, management and access processes

Whilst the architecture of the simulation is relatively simple, the tool is provided with functionality that aims to model realistic corruption and storage system management processes. For this purpose, a detailed data storage model has been developed where archived assets are represented as file objects that include such properties as name, size, and corruption details. Each asset can have more than one file (replica) representing it within the system. Consequently, each storage system contains a list of such items and is responsible for their storage and integrity management. In the current model, file assets are homogeneous. In subsequent versions of the tool, assets will be distinguished according to different types and different preservation actions and behaviours can be associated with each type.

During the simulation files become corrupted as a result of latent (silent) corruption or access corruption (e.g. during file read/write or access). For each corruption type it is possible to define a number of corruption events that are probabilistically triggered by the simulation on a per-tick basis (with a tick typically representing a real-world increment of 1 day). For example, a possible corruption event might be specified as: corruption of a 1Kbyte block with a probability of occurrence of 1 in 10^{10} blocks over 12 months. Assuming a Poisson distribution, this rate is then converted into the probability of corruption on a per-

tick basis. The current model assumes that corruptions are randomly distributed across the data being stored. The probability of corruption can be varied over time, either interactively during a simulation or following parameters set for each generation of a storage system.

When corruption takes place, it starts at a randomly selected location within the storage system and damages the data size that was specified by the corruption event blocks described earlier, files are modelled as containing critical section (the size of which can be configured) and a non-critical section. When corruption hits the critical section, then file is considered as not repairable and this triggers a repair process by using one of the other replicas of the file.

On the top of corruption processes, the operation of each storage system includes ingesting new files, providing access to existing files, migrating files and storage, and managing integrity. Since all of these system-level activities consist of more than a single atomic action (e.g. access to an asset includes file integrity check, its potential repair and transcoding before it becomes accessed), they are defined as workflows consisting of series of actions that are, in the end, realized through the execution of storage system services. Some workflows are relatively simple, for example checking the integrity of a file by reading it, generating a checksum and then comparing that checksum with a reference checksum (Figure 8).

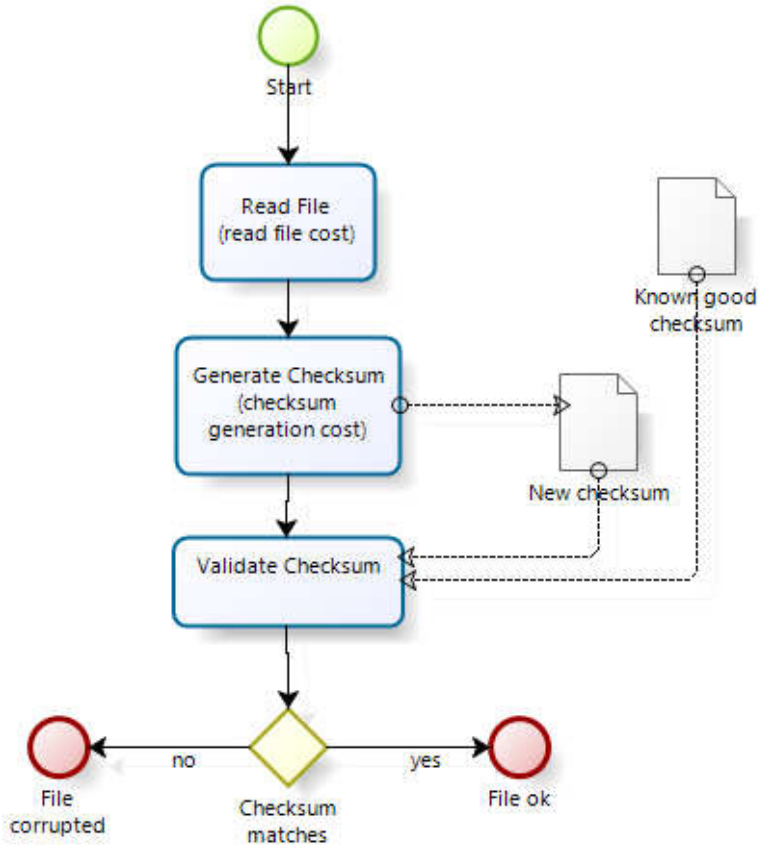


Figure 8 Validate checksum workflow (BPMN notation)

These workflows are then incorporated into other processes, for example copying a file which involves a read operation, a write operation, and a check that no integrity has been lost in the process (Figure 9). Further workflows can then be built up that have increasing

complexity, for example the integrity check and repair workflow (Figure 10) that verifies the integrity of the file, and if integrity is lost then performs necessary repair through a copy operation of a known good replica or if unsuccessful then an attempt to do a local repair.

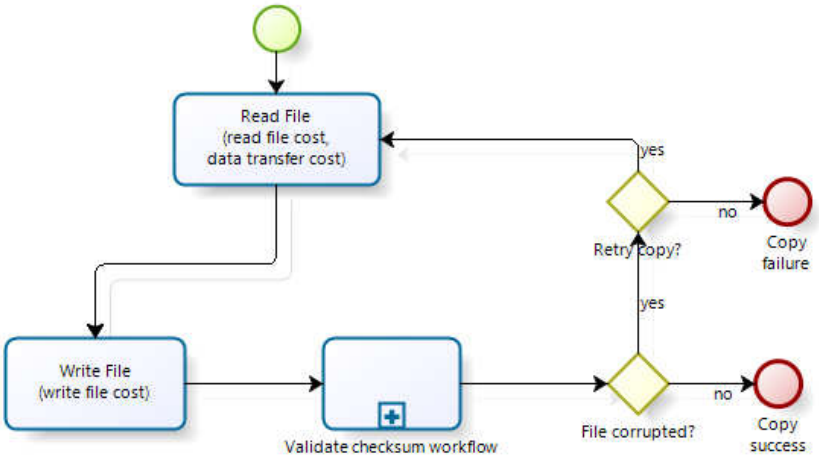


Figure 9 Copy file workflow

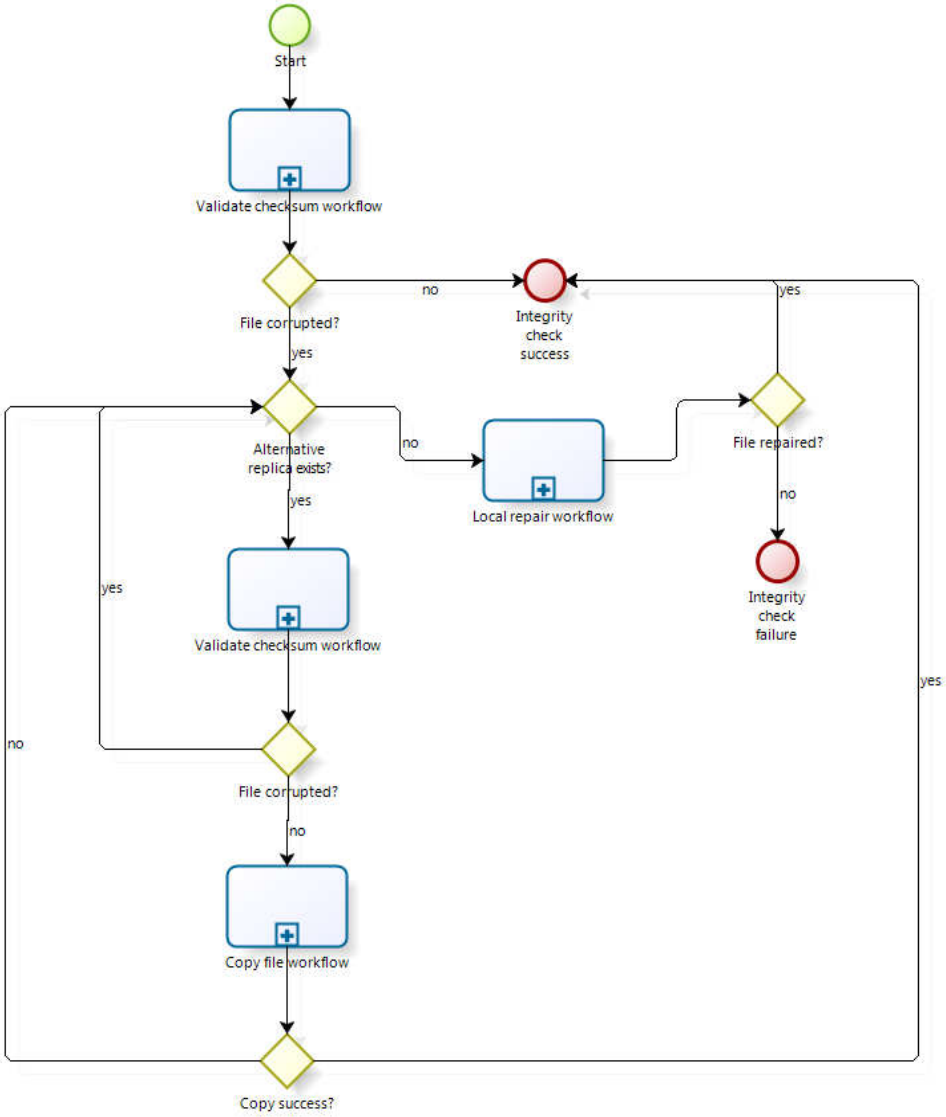


Figure 10 Integrity management workflow

The ability to break down and express each management activity in the form of a workflow (and its actions) provides several advantages. Firstly, the system-level functionality is broken down into atomic parts that are realizable through the execution of different services. Since services are resource constrained, this allows the simulation to be used to understand of the impact of under-provisioning of resources. Secondly, since the tool is customized to process workflows, the same system-level functionality can be realized using different set of actions (hence workflows) allowing for a broad customization to real-world archive system examples and their simulation.

Since most of the storage system operations are dependent on the execution of services, another important feature of the simulation tool is the ability to model the impact of resource constraints on the efficiency of the system. This can be achieved by setting up a specific allocation of resources to individual services that remains unchanged during the simulation or to allow different services to consume shared pool of resources. In the latter case, the model provides two simple resource allocation algorithms (round robin and greedy) that control the access to limited resources. These algorithms can be further extended to address more realistic allocation strategies aiming to optimize usage of resources without system performance loss.

3.5.6 Interactive and batch execution

There are two complementary ways in which the tool can be executed: interactive mode and batch execution mode.

In interactive mode the user is presented with GUI that provides the user with a number of performance charts reflecting the storage system size, operational cost, lost assets, assets at risk, actual state of the service, ingestion and access queues. These charts are updated continually to show the most up-to-date information. The tool exposes a set of configuration options through which the user may alter the course of the simulation (e.g. change the corruption probability, number of replicas stored in the system or frequency of integrity checks).

Using this interactive interface, the user is able to evaluate a number of ‘what-if’ scenarios and observe their real-time outcomes during the simulation run. Such interactive features not only serves as an educational tool for people that wish to understand how storage systems operate, but may also prove useful to experienced archive system administrators to understand the consequences of their administrative actions within the simulated environment.

A complementary approach is to use batch execution. This mode does not expose any GUI to the end user but instead offers the ability to run multiple simulations in an automated way. In this mode the user specifies a batch of file configurations (in a human readable format) that the model repetitively executes. The results of each simulation are stored in log files for further processing and statistical analysis.

3.5.7 Limitations and assumptions

The model and simulation tool focus on ‘bit preservation’ rather than ‘content preservation’. For example, the tool does not model operation or obsolescence of the software stack needed to make full use of digital AV content and metadata.

It is not sufficient to preserve the data if it cannot be interpreted owing to loss of metadata or loss of the software/expertise to interpret the meaning of the data. The tool described above represents metadata as a 'critical' section of the file that, if corrupted, causes the whole file to be irretrievably lost. The tool also simulates the process of migration to maintain data in long-term preservation formats that ensure successful access to and interpretation of the data. It does not explicitly simulate obsolescence of the software stack. It is future work to simulate the situation in which access to data that has not been migrated is attempted once the software required to interpret it (e.g. a video codec) is no longer available, thereby causing access to fail.

3.6 Input parameters and model validation

The quality of the outputs of a model depends on having a correct and complete set of input parameters that are then used in a model that is itself correct and representative of the real-world situation being modelled. The issues of input parameters and model correctness/validation are discussed in more detail below.

3.6.1 Input parameters

iModel has a large set of input parameters including costs, failure rates, time and resources needed for executing different activities, and how all of these change over time.

The question is where to get value for each of these parameters.

Some of the sources include:

- The initial PrestoPRIME Deliverable D2.1.2 on modelling provided some examples of costs and failure rates for storage and the trends for how these change over time.
- Failure rates for storage seen in large scale field trials as reported in the literature. See section 3.3
- Costs reported by other projects that have done cost modelling. See Section 3.2
- Failure rates seen in audiovisual archives using IT storage technologies, e.g. data tape libraries. Some examples are provided in PrestoPRIME Deliverable D3.2.1
- Costs of online storage services and how these have changed with time, which provides an estimate of TCO of storage (after taking out an estimate of margin made by the service provider).
- Costs of storage technology available from vendors or from websites such as StorageMojo[2]

[2] http://storagemojo.com/

- Costs of preservation activities as estimated by other cost modelling projects. See Appendix A: cost modelling and PrestoPRIME deliverable D6.3.1 “Financial models and calculation mechanisms”.

- David Rosenthal[3] provides regular reports on the costs, reliability and trends for storage that together provides an excellent gateway to sources of parameter values for iModel. http://blog.dshr.org

When using figures from these sources, the following should to be kept in mind:

- Failure rates for storage depend on many factors including the specific manufacturer of a given technology, the environmental conditions in which it is operated, the workload on the storage, maintenance and upgrade, etc. Therefore, whilst industry averages can be used, mileage can vary.

- Costs for storage can vary rapidly over time and long-term projections are very sensitive to current costs.

- Failures in storage are often quoted as averages, e.g. Bit Error Rate (BER). However failures in practice are rarely single bits. The BER needs to be converted into the size of the failure (bit, byte, sector, block, drive, tape, array etc.) and the probability of the failure.

- List prices for enterprise scale storage (e.g. tape libraries) are often a lot higher than prices that can be secured by negotiation.

- People costs are often underestimated or ignored.

- Costs can depend strongly on the amount of data being stored and hence the current and projected size of data in an archive is an important factor. A simple example of this is provided in Appendix C: storage cost-curves.

3.6.2 Validation

Knowledge that iModel is ‘correct’ is something frequently requested by users. Correct in this sense means that:

- The model correctly approximates storage systems for the purposes of making projections of cost and risk; and

- The model has been correctly implemented in software (i.e. it does not have bugs).

This is extremely difficult – only limited information is available on current storage technologies, especially their failure modes and frequencies, let alone future storage technologies that might become available in say the next 20 years.

The technologies that are starting to be used by archives today, e.g. data tape, have not been employed for long enough or widely enough for a model to be validated against

[3] http://blog.dshr.org

historical data collected from archive systems used in the field. Indeed, many archives who have operated IT storage systems for long-term storage of data have not recorded information on either costs or failure types and rates. Manufacturer estimates of reliability have been shown to be unreliable, so that doesn't provide a way to validate iModel either.

So what to do? The issue becomes one of building confidence in the design of the tool and that it has been implemented correctly. The approach taken for iModel includes:

1. Testing (to make sure it is working correctly and there are no bugs). For example, developing simple test cases for units of functionality (e.g. corruption, repair, ingest, replication, resource consumption, cost calculation etc.) based on calculations of correct behaviour. These tests are deterministic and check that the 'pieces' of the model are implemented correctly. These test cases can be found in the source code of iModel as part of the distribution.

2. Release of the model including documentation in an open way that allows independent assessment of its implementation and correctness. This was one of the main drivers for open-sourcing the tools and providing bug reporting and feature request services.

3. Compare the results of iModel with models developed by others or with the results of simple models where there is an analytical solution. This includes comparing the results of iModel with the results of our web-based tool (since they use different underlying modelling techniques), but more importantly comparing iModel with other published models. Most of the existing models focus on simple Markov techniques and therefore only cover subsets of iModel or degenerate cases. iModel does give the same results as Markov techniques. Some simple examples are given in the Appendix D: example iModel test case.

4. Comparing results of iModel with archive experiences with real systems. In the PrestoPRIME testbeds, iModel has been used to simulate storage scenarios provided by testbed users and has showed the behaviour expected by those users. It should be noted that detailed parameter values were not available during these tests and hence it was only the overall behaviour of the model that was validated.

3.7 Example: modelling the cost of risk of loss

An example of the use of iModel in batch mode followed by visualisation of the results set is presented below. The full version is available as an article on the website:

http://prestoprime.it-innovation.soton.ac.uk/imodel/visualisation/

Estimating the long-term costs and risks of archiving digital content is not an easy task! Nearly five years ago Richard Wright from the BBC expressed this challenge in terms of what an archive might want[4]. He asked for a cost curve for the percentage of loss of archive content over 20 years, including the probability of loss, the size of the uncertainties, and a cost function to show how the probability of loss varies against increase - or decrease - in investment. His request was for no more than basic actuarial

[4] Wright, R. 2007. Structural requirements for digital audiovisual preservation: Tools and Trends. International Conference on Digital Preservation, Koninklijke Bibliotheek, The Hague, The Netherlands

information needed for investment decisions. This article shows how we have approached the problem of providing quantified costs, risks and uncertainties for long-term storage of file-based assets using IT storage technology.

The interactive storage simulation tool allows a user to manipulate a storage model in order to observe the effects of changing the storage strategy on cost and on the risk of loss of assets. The tool can also be used to batch process a number of parameterised configurations in order to explore the space of possible storage strategies. Given the results of this, it is possible to compare directly the effect of, for example, keeping a greater number of replicas of each asset while scrubbing the files less frequently.

The storage simulation tool uses a stochastic simulation, which means that each time it is run using the same initial configuration the results may vary in terms of the number of files lost and the total running cost. By repeatedly running the tool using the same configuration, we can generate a probability distribution of asset loss (below). The figure was generated by sampling the model 1000 times, each time simulating 10 years of preservation activity, and indicates that we would expect to lose around 0.3% of the archive over 10 years (or around 75 assets for an archive of 25,000 assets).

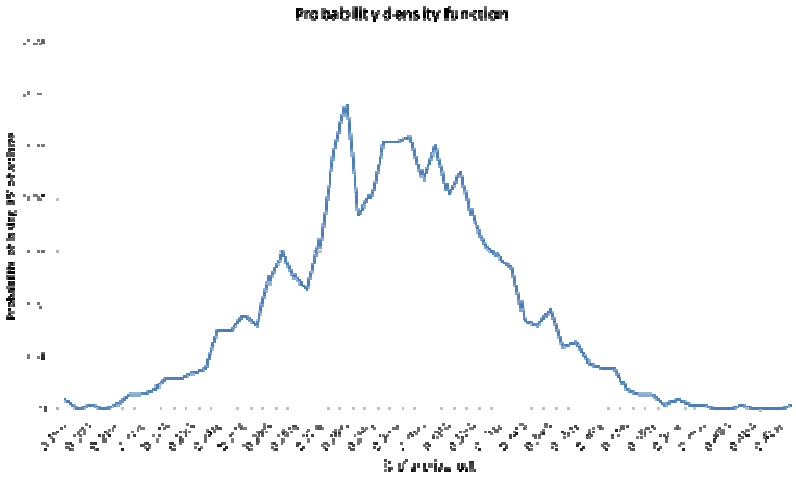


Figure 11 Probability Density Function

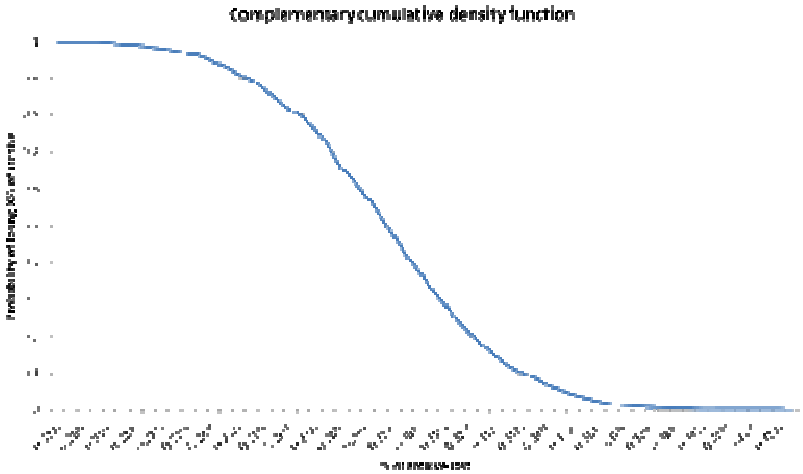


Figure 12 Cumulative Density Function

The probability distribution function (PDF) tells us the probability of losing exactly a given percentage of the archive as shown in Figure 11. By cumulatively summing the probabilities up to a given percentage of archive loss, we can generate a cumulative distribution function (CDF) of the PDF. The CDF (Figure 12) gives the probability of losing this amount of the archive or less. However, we are interested in the probability of losing a given amount of the archive or more. This can be found by taking the complement of the CDF, shown in Fig. 2. The probability of losing more than a very small amount of the archive is high, while losing more than large amounts of the archive is low.

Figure 13 shows a two-dimensional representation of the multi-dimensional data produced by the process outlined above. This shows a single storage system where the number of additional copies of a file stored in that system and frequency of integrity checking (scrubbing) have an impact on both the cost and the risk of file loss. The figure illustrates the risk and cost landscape for the loss of more than a specified percentage of the archive's assets (the acceptable maximum level of loss). The boundary between adjacent coloured bands represents configurations of equal cost. The white contour lines are lines of equal risk of loss. Each intersection of values from the X and Y axes represents a storage simulation that was actually executed (multiple times). The intervening values are interpolated.

This type of visualisation helps the decision maker to identify the optimal storage strategy given their constraints. Firstly, given a fixed budget, it enables them to select the storage strategy with the lowest probability of asset loss. For example, for a given budget of 50 million (Euros), the strategy with the lowest probability of loss of more than 0.1% of the archive over 10 years is to keep 3 additional copies of each asset and to check the integrity of the files every 10 months. In this case, it is not cost efficient to increase the frequency of scrubbing, as it will cost more but is unlikely to deliver any benefit in terms of data safety. Similarly, given that we are willing to accept the risk of losing 0.1% of the archive over 10 years with a probability of 1 in 5, then the strategy with the lowest cost is to keep 3 additional copies of each asset and to check the integrity of the files every 12 months.

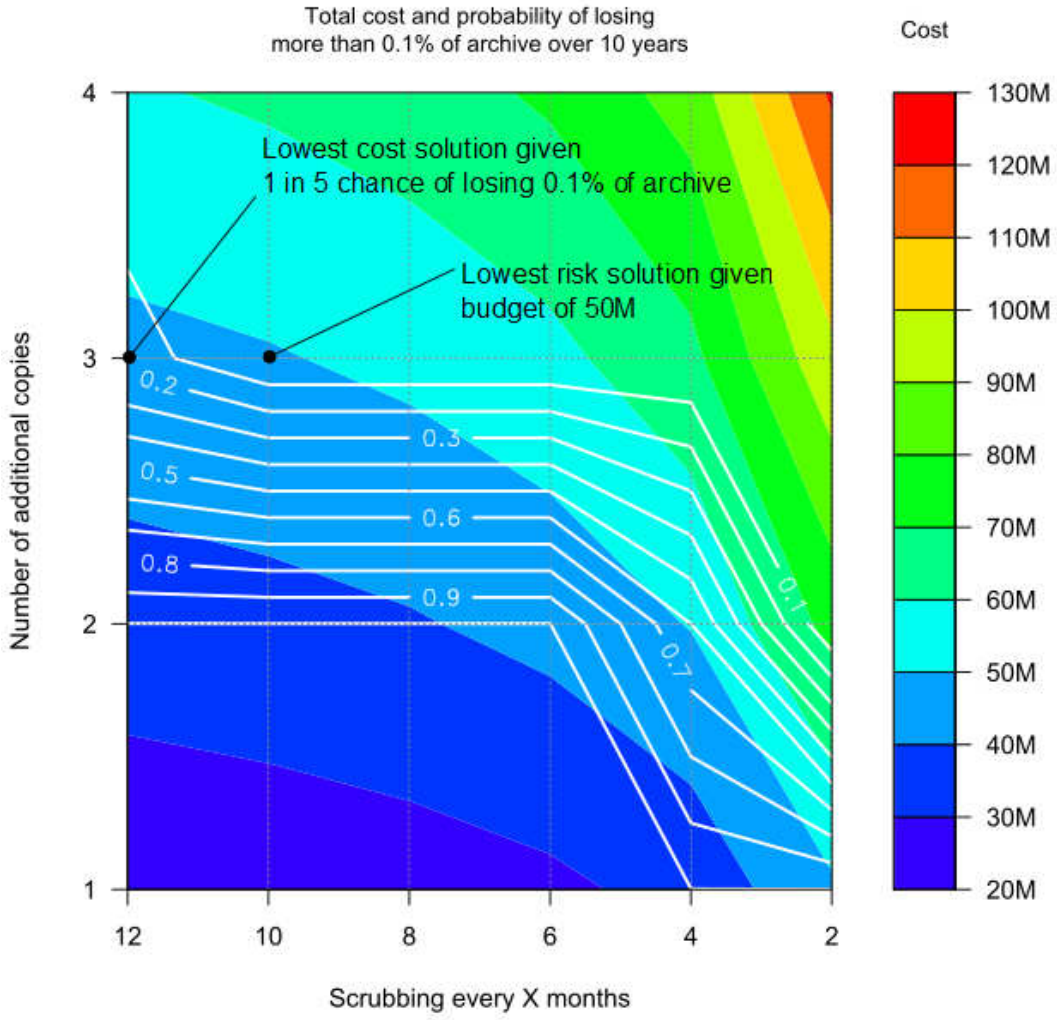


Figure 13 Cost of risk map

3.8 Example: resource contention

The previous case study demonstrated how to optimize for data safety requirements by balancing replication and integrity checking. This is not sufficient in an archive with finite available resources, for example a fixed number of tape drives in a library. As an archive grows with ingested content, the resource constraints affect the ability of the system to work effectively.

Consider a scenario where two copies are made of each asset being stored in an archive. Both copies are stored on HDD servers (JBOD storage with 1TB drives with an annual failure rate of 1% plus latent corruption rate of 1 in 10^{14} bits each year). Assets are ingest onto one storage server and then replicated to the other. In addition consider that there are two dedicated servers for checksum generation (each server capable of checksum generation at 100MB/sec). The ingest profile is one 400GB asset per day for the first 12 months, which then doubles every year, i.e. rapid storage growth. A checksum is generated for each file on ingest and then again on every file annually to verify integrity. Every time a file is written to storage a checksum is generated to verify it has been correctly stored. If one copy is detected as having a checksum mismatch, e.g. due to latent corruption, then the other copy is used for repair. This active 'scrubbing' aims to maintain integrity, but if

both copies are corrupted then the asset is considered lost. The storage volume is shown in Figure 14. Storage volume starts by doubling each year. However, after time, the limited resources for checksum are insufficient to scrub the storage systems quickly. Scrubbing is given priority over ingest resulting in a backlog building up (flat areas on graph). The backlog of ingest then puts further strain on the checksum servers as checksums need generating on each new file (steep rises after flat areas).

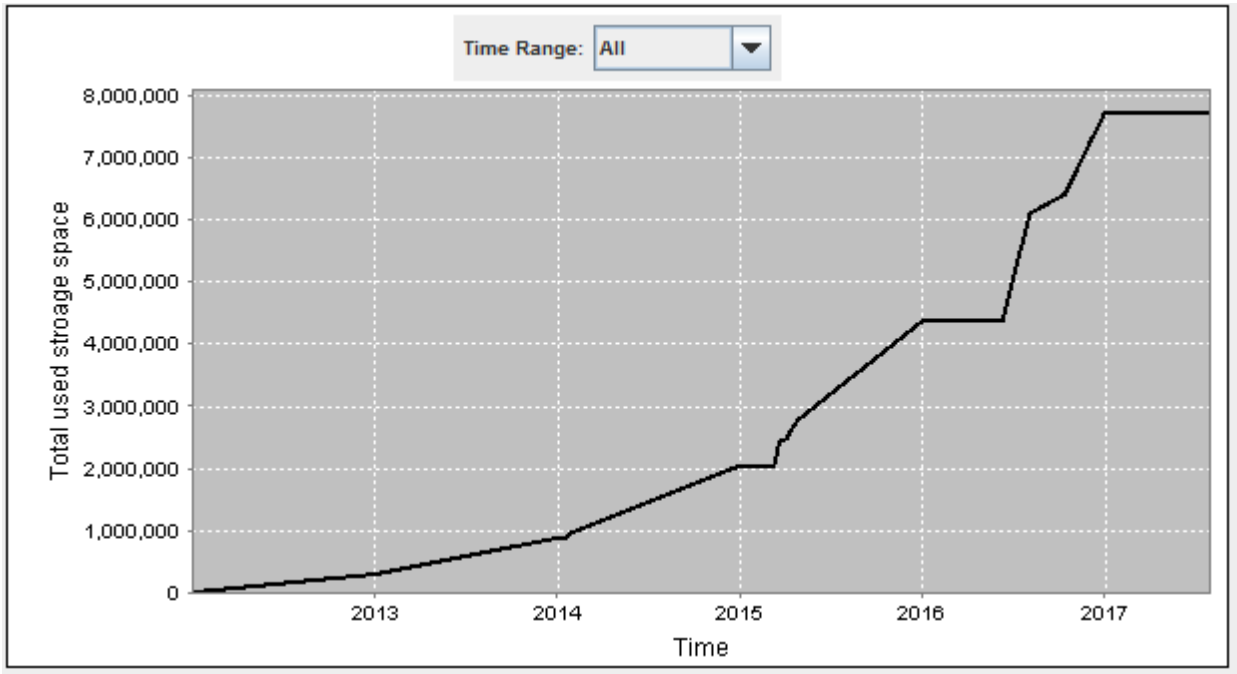


Figure 14 Archive storage volume

Figure 15 shows the number of checksum requests waiting to be processed over time. The steep rises correspond to annual scrubbing. After several years, the checksum servers are not sufficient to cope with the load of scrubbing plus ingest, and the queue of requests fails to ever fall back to zero showing server overload.

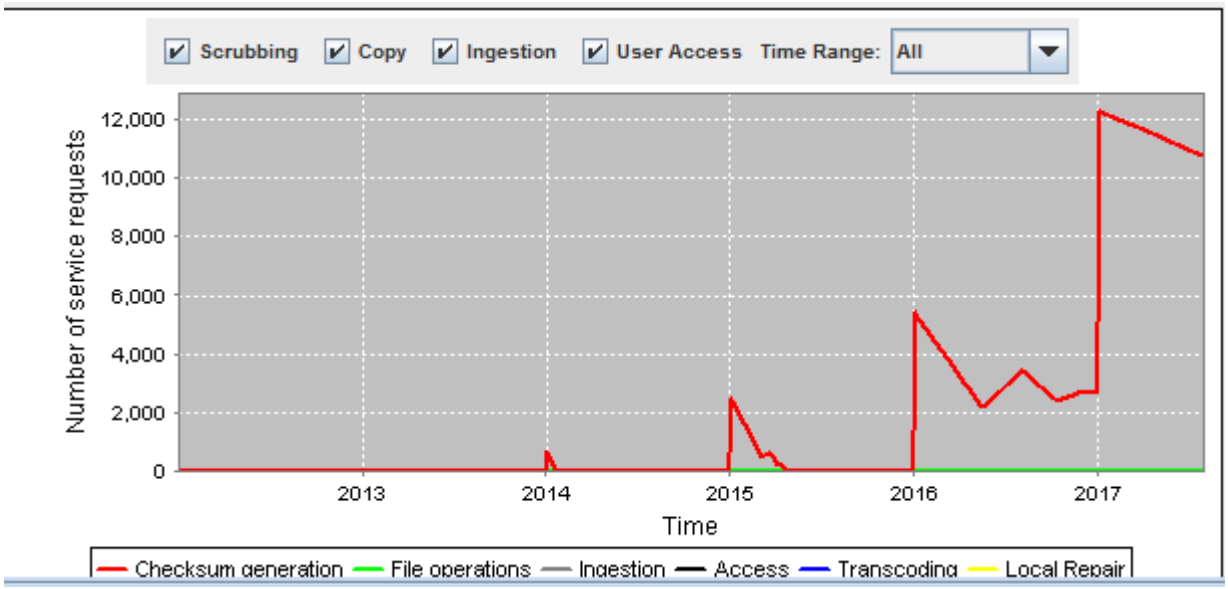


Figure 15 Scrubbing requests

The impact of having finite and fixed resources for computing integrity of the assets is shown in Figure 16 and Figure 17. The number of assets at risk is initially kept under control by the process of scrubbing (integrity check and repair) and this can be seen as the number of assets at risk falling to near zero after each annual scrubbing phase. However, when the servers for checksum generation are overloaded, the number of assets at risk goes out of control (Figure 16) with the result that the asset loss rate increases (Figure 17).

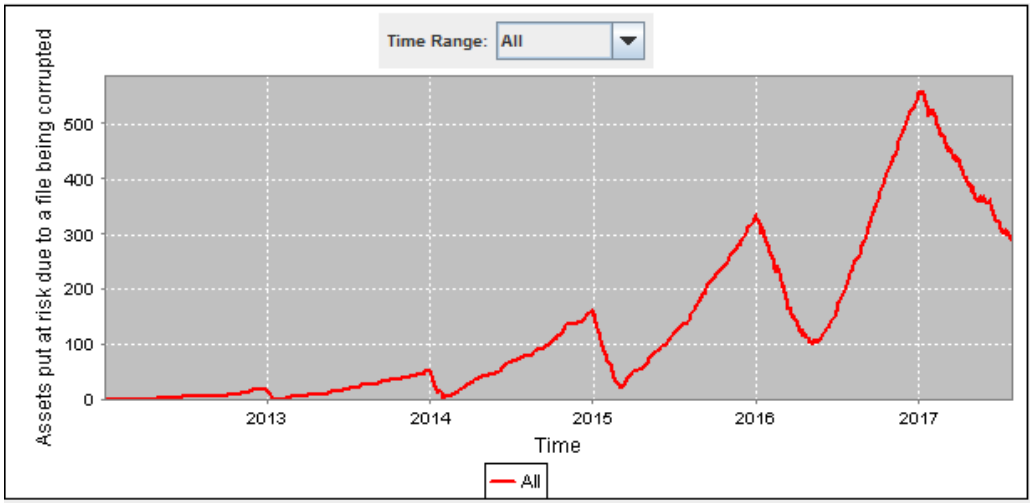


Figure 16 Assets at risk

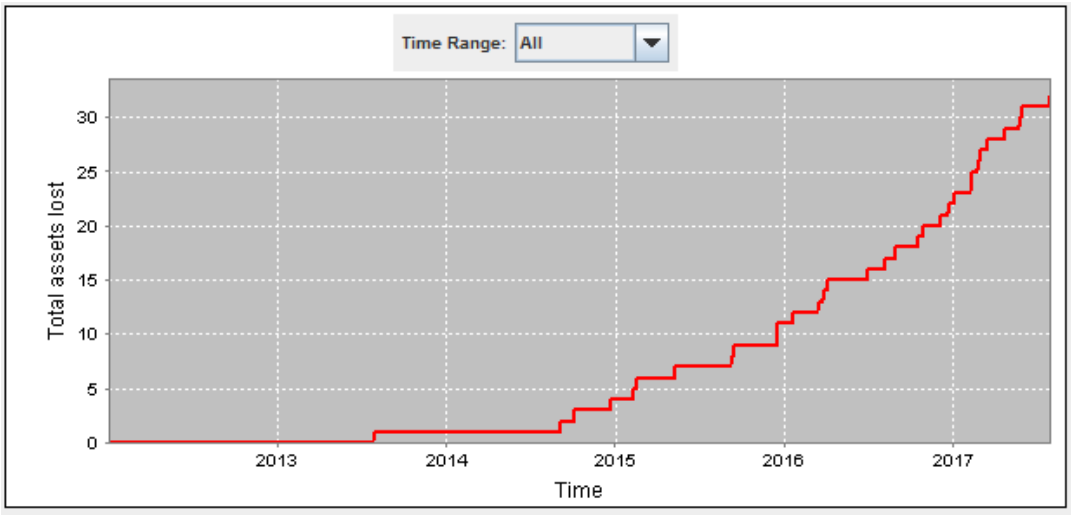


Figure 17 Assets lost

Whilst the example above is somewhat contrived, it serves to show the need to consider finite resources and contention. A similar result can arise from the contention between the requirements to provide demand-driven customer-facing services (such as access to assets), the load imposed by regular preservation activities, such as storage or file-format migration, and the load imposed by file assets that grow in size as well as number.

Resources can be shared to address the limitations of a finite resource pool. The simulation tool allows resource sharing according to several strategies: a static allocation, a greedy strategy that delivers resources where they are needed most, and a round-robin strategy that allows all services a fair use of resources. Even the simplest of these strategies can make a large difference to the ability of a resource-constrained archive to cope with the

demand of internal and external activities. The tool can be extended to include more sophisticated scheduling algorithms that can balance resources against requirements.

As well as steadily increasing demand on resources (e.g. from users), an archive can experience a sudden claims on its resources owing to catastrophic events (e.g. a DC going offline or large scale storage system failure). The tool allows us to schedule these as 'one off events' in the simulation and investigate the consequent hit on the use of resources.

3.9 Integration of iModel and service management

The iModel storage simulation tool has been integrated with another tool from IT innovation called Ting. Ting is a service monitoring and management framework. It can be used by a Service Provider to offer services to customers, and to monitor and manage their use of these services. It can also be used by customers to monitor and manage their use of resources.

The Ting service management software has in turn been integrated with other PrestoPRIME tools: the P4 preservation platform and the MServe ‘Preservation as a Service’ framework. MServe is a RESTFul Web Framework for Service Providers who want to deliver preservation using a Software as a Service model. The purpose of MServe is to provide human and machine usable interfaces to control the ingest, access, processing and manipulation of content using computing resources.

Further details are available in PrestoPRIME Deliverable ID3.4.4 “Second Version of Prototype for distributing and storing audiovisual content using federated storage and compute services”.

The rest of this section provides an overview of the integration and what can be achieved.

iModel, Ting, P4 and MServe together provide a complete solution for supporting the plan-do-check-act cycle (PDCA) - see Figure 18. PDCA is an iterative four-step management method used in business for the control and continuous improvement of processes and products.

Note that when we use the term Service Provider, this is in a broad sense. A Service Provider might be a third-party commercial service provider that is contracted by an archive to fulfil some of the archive functions, or it might be the internal IT function of an archive where the IT function provides services to the rest of the archive or wider business.

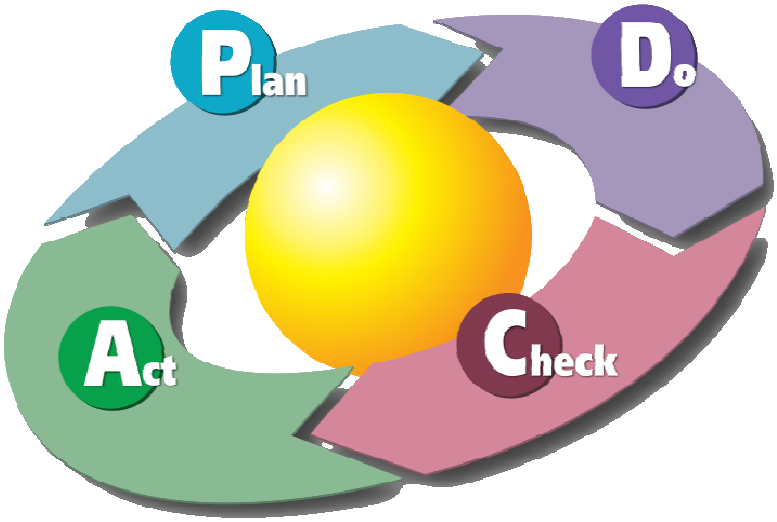


Figure 18. The plan-doc-check-act (PDCA) cycle[5]

iModel can be used to help *plan* a new storage system, it can be implemented ("*do*") using P4, MServe and the other tools developed in the PrestoPRIME project and the *check* and *act* aspects come from the interactions of MServe, P4 and Ting with Ting monitoring the SLAs and taking action when required.

To complete the cycle though we must return to the planning stage, taking into account what has been learnt in the implementation. The "Plan & Optimise" section of the Ting web GUI (Figure 19) is a decision support tool and assists in this re-evaluation process by pulling in the historical monitoring data from MServe relating to the space used by the ingested assets and file corruption events and then using this data to help parameterise iModel. This serves two purposes:

1. By overlaying the actual data collected over time with parameterised predictions from iModel for the same time period we are able to compare the two and validate the model's predictions.

2. The parameterised model can be used to predict future trends and issues, thus helping plan for the future. For instance, it may predict that with the existing configuration that a data loss incident could occur in the next year(s) and this information would be taken into account by the administrator in any re-evaluation of the storage system.

Figure 19 shows a screenshot of the long-term planning service as controlled through the Ting web GUI. This interface allows the storage simulation model to be run and to predict the performance (in terms of data safety and cost) of the archive over an arbitrary period of time. A number of simulations can be run in parallel, the results stored, retrieved and compared using graphical plots.

[5] Diagram by Karn G. Bulsuk (http://www.bulsuk.com).

Figure 19. Ting web GUI for "Plan & Optimise". A plot of used storage space is shown.

To achieve this integration, the Ting GUI retrieves historical data from the Ting service, computes parameters such as ingest and access rate and then uses these to modify a pre-defined iModel configuration. The execution of iModel is then launched on the MServe system (in the same way as it is used by P4 for data processing) and the resulting data retrieved.

Figure 20 and Figure 21 show the results of predicting one year into the future based on a simple storage configuration with no replication.

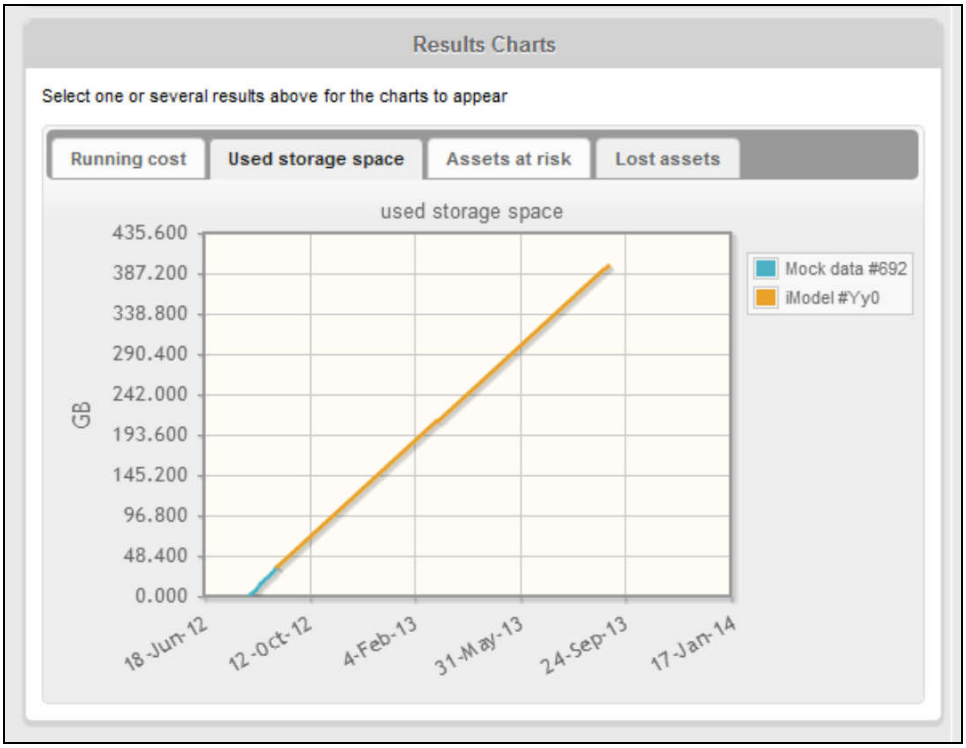


Figure 20. Prediction of used storage space for the next year (orange), generated by iModel according to parameters extracted from historical data (blue).

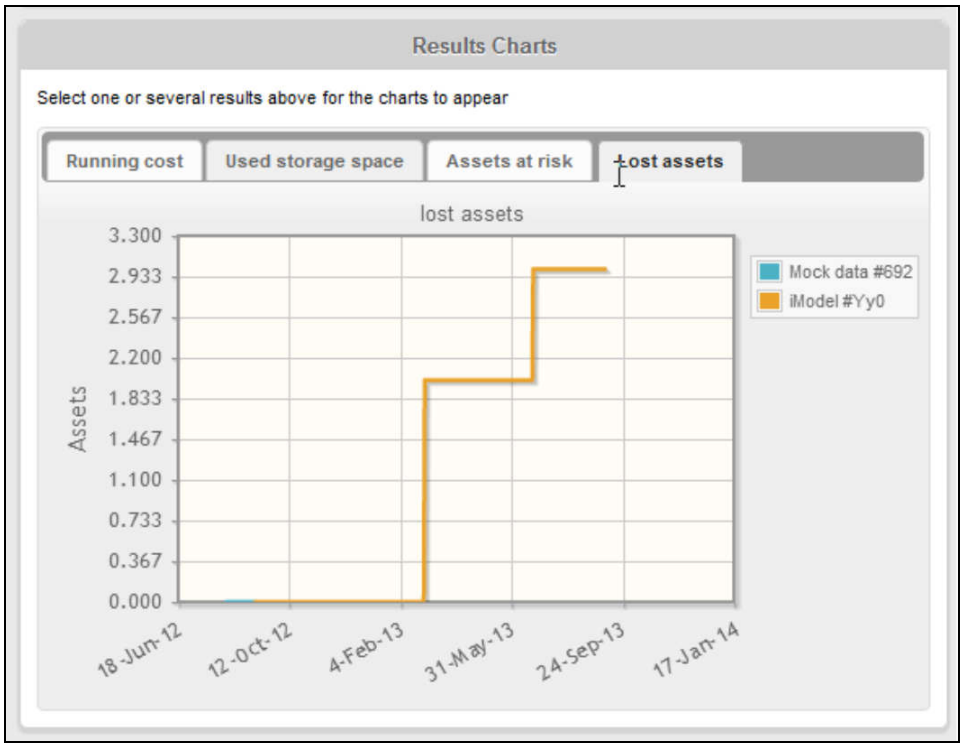


Figure 21. Prediction of the number of lost assets for the next year (orange), generated by iModel according to parameters extracted from historical data (blue).

4 iWorkflow

4.1 D3 transfer workflow

This section provides an example of work done to simulate the BBC D3 workflow. This shows both screenshots from the tool produced and also some examples of the results achieved. The objective of the modelling is to understand the compromises between cost, quality and throughput of the workflow – and how to improve them all through effective automation – whilst also keeping an eye on what resources are needed (e.g. staff, computers, head-life on D3 decks etc.).

The D3 project at the BBC is an effort to migrate video from approx. 100,000 D3 tapes into file format (MXF wrapped uncompressed video and audio) and store it on LTO data tape.

The details and data presented here are based on a simplified snapshot of the D3 process from around June 2011. The approach is to capture the characteristics of the workflow and its possible alteration for the purposes of simulation rather than providing an accurate description of the exact workflow used at the BBC. In particular, the graphs showing the results of simulations are based on representative but fictitious numbers given in the Appendix. The objective is to provide examples so that the reader understands what each parameter of the model is for and what the model produces. Actual numbers from the BBC workflow have not been included for confidentiality reasons.

The workflow starts with D3 tapes that operators load in to D3 decks and capture the resulting SDI stream to a file. These files are then written to data tape and the AV content manually inspected by QC operators at dedicated QC stations. The operators look for defects introduced during the transfer as well as already existing in the video (e.g. from previous migrations such as 2” Quad to D3).

Inputs to the simulation include the number of D3 tapes, the number and cost of D3 operators and decks, the number and cost of QC operators and workstations, the time and resources needed for each step (e.g. transfer, reviewing defects), the frequency of defects and the effectiveness of operators in detecting them, the likelihood of re-transfers being required, and the cost/capacity of the storage systems used in the workflow.

Scenarios that can be simulated include: (a) the result of reducing time spent on manual QC, e.g. time-boxing instead of a full pass of every item (b) the benefits of using automated quality analysis software to guide the QC operators, and (c) the effect of increasing resource to remove bottlenecks, or the impact of temporary loss of resources, e.g. operator illness or systems failures.

The result of a typical analysis shows the rate at which D3 tapes complete the process for different workflow configurations and the corresponding number of defects not picked up in QC. The costs of the different configurations can be compared and hence a cost/ throughput/ quality comparison done. Optimisation and sensitivity analysis can then done for each of the steps in a given workflow, e.g. by looking at queue build up and resourcing for QC.

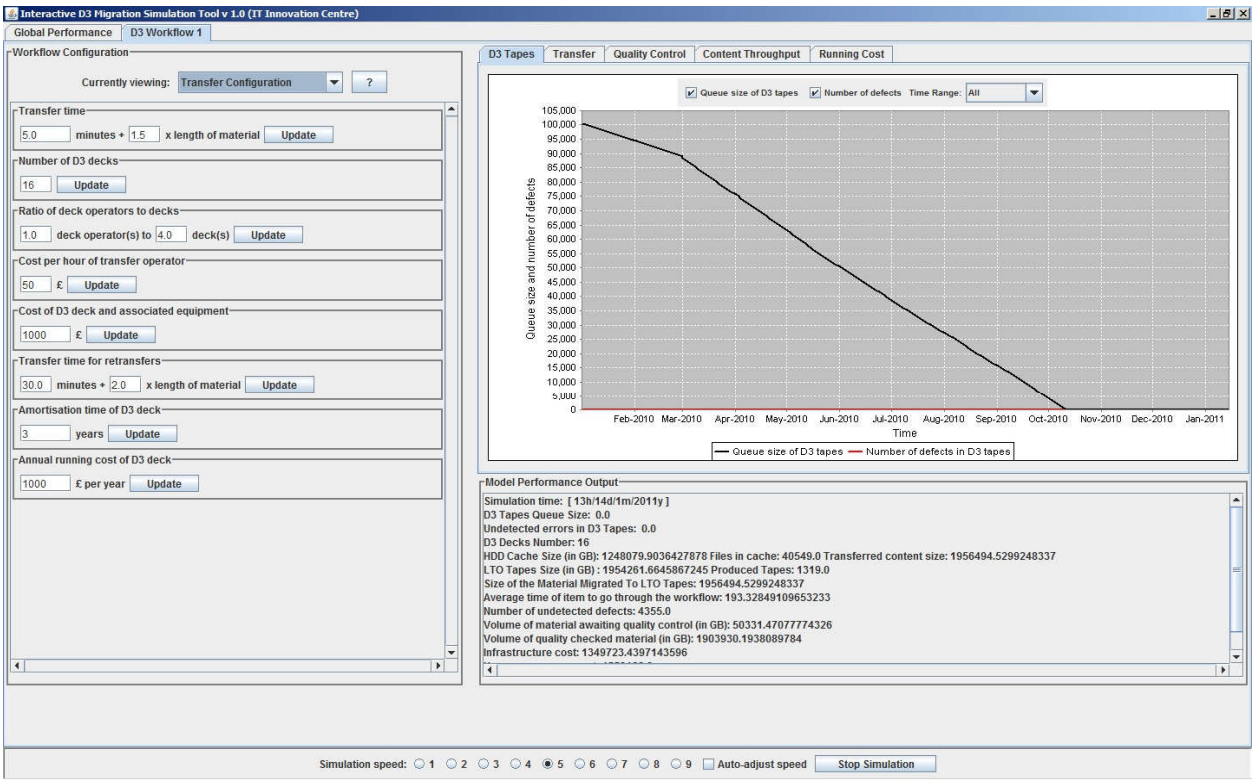


Figure 22 Configuration and simulation of the D3 transfer part of the workflow. The graph shows the number of D3 tapes that still need to be transferred as a function of time.

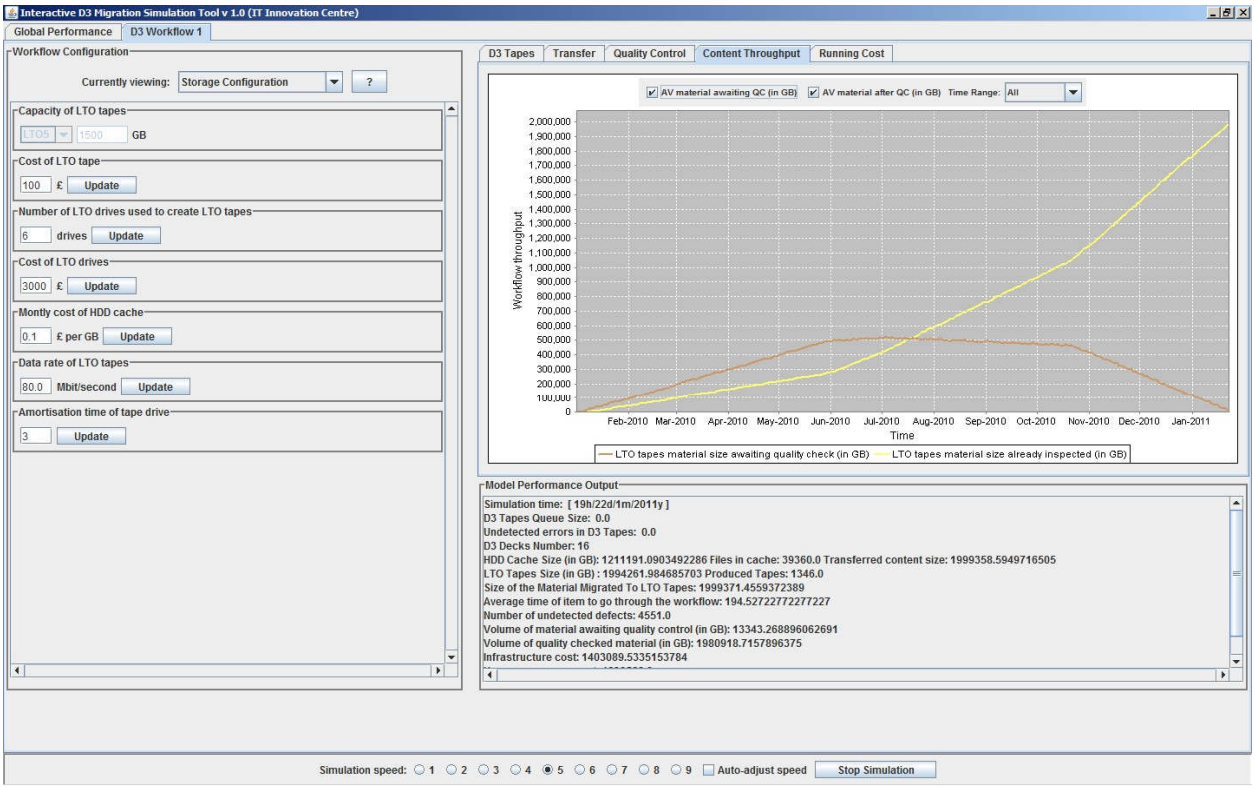


Figure 23. The LTO part of the workflow where data from D3 video tapes are MXF wrapped and written onto LTO data tape. The graph shows how many LTO tapes have been created and how many of those have passed through QC.

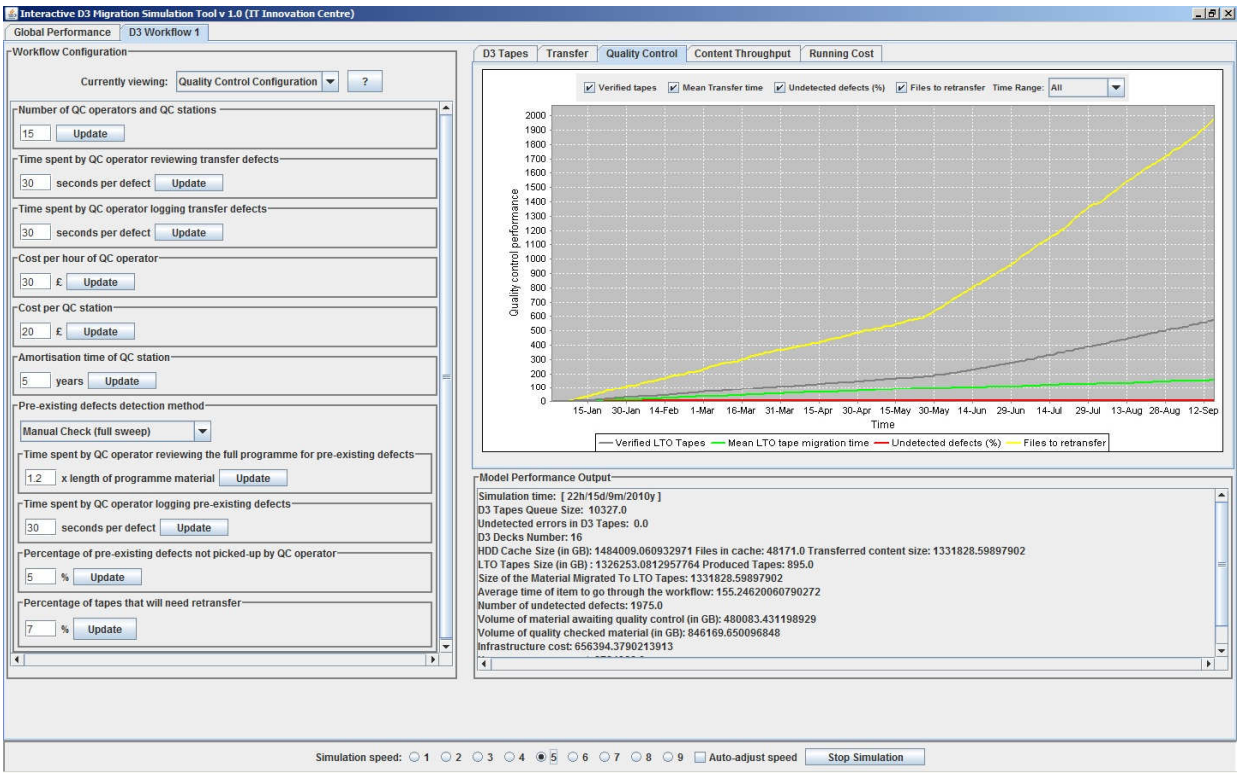


Figure 24 The QC part of the workflow allows different QC strategies to be investigated, e.g. full manual sweeps of the content or use of software defect detection tools. The graphs show how many defects pass through undetected along with the rate at which items are processed.

Running the tool multiple times with different configurations allows different workflow strategies to be investigated. An example is shown below for some hypothetical workflow where parameter values, e.g. cost and error rates, have been exaggerated to make the comparison more visual and easy to interpret.

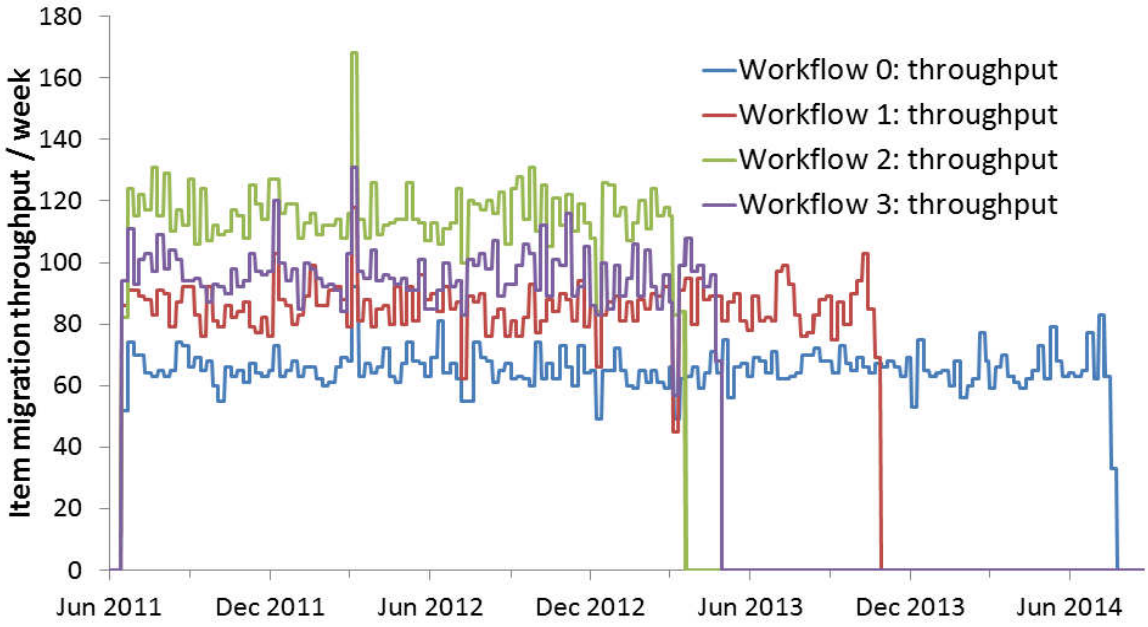


Figure 25 The throughput of 4 different workflow configurations. The higher the throughput, the higher the cost, but the faster the whole workflow completes. This allows different workflows to be assessed in terms of completion within a required project timeframe, e.g. 2 years.

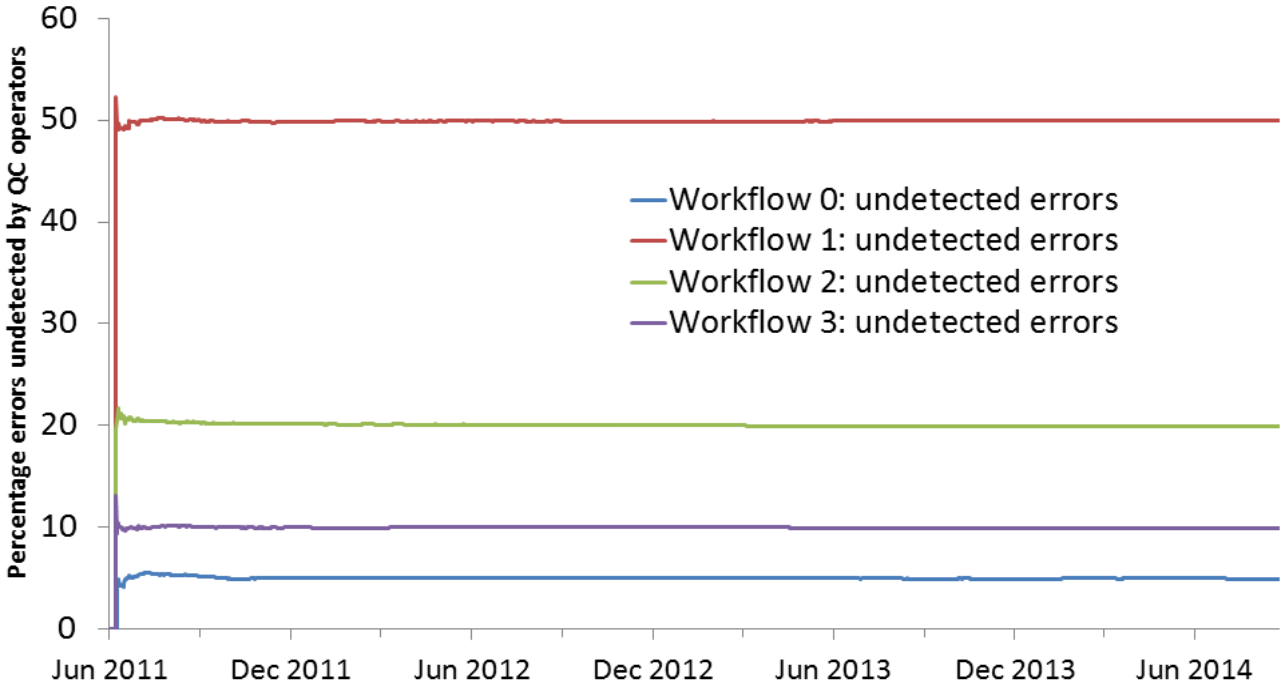


Figure 26 The effectiveness of different QC strategies used in each of the workflows is shown in this graph. Note that workflow 0 has the lowest error rate, i.e. the most effective QC, but also took the longest to execute (see previous figure) and also had the highest cost. For the other workflows, there is not a direct correlation between QC effectiveness and throughput showing that some strategies are more cost effective and efficient than others.

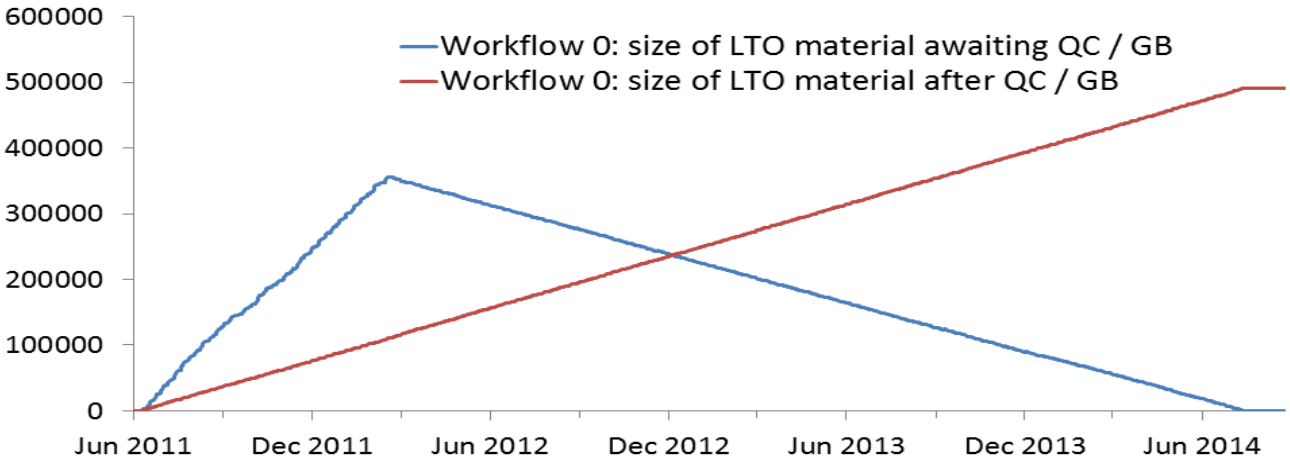


Figure 27 For each of the workflows, details are available on where the bottlenecks are in the different stages of the process. This allows further optimisation to be done.

The resources used in the workflow are also calculated or set as input parameters, e.g. amount of head life for the D3 decks, number of QC staff, and defect detection rates in the QC stage. This allows various risks to be explored, e.g. whether head life will run out, whether staff outages will compromise hitting project deadlines, or number of tapes that will not be QC'ed to a high enough standard. This is another example of how the risk management part of WP3 has been combined with the modelling work in WP2.

4.2 Costs of workflow modelling

The simulation of the BBC D3 workflow took approximately 3 months to implement. This included the time needed to: interview staff at the BBC; define the objectives of the simulation work; describe the workflow and agree this with the BBC as representative of the true workflow; implement the simulation; and present and review initial results with the BBC. The model shows the behaviour which is in-line with expectations and simple 'back of the envelope' calculations, but has yet to be fully calibrated. Calibration requires detailed analysis of statistics collected by the BBC on past performance and resources used for the D3 workflow. This analysis and calibration task is estimated at a further 1 person month of effort. The simulation itself was implemented in Java using a bespoke software framework by an experienced software developer who had previous experience of simulation and modelling work. The simulation framework was not developed from scratch and uses some components from iModel and previous work by the software developer.

In total, a reasonable estimate of the effort needed to build and calibrate a model from the ground-up and including all steps from initial specification through to final calibration and acceptance tests is approximately 4-6 person months.

For comparison, IT Innovation tasked a student to build a similar simulation using a commercial off the shelf stochastic process modelling package called Simul8[6]. The work used the same specification of the workflow as used for the bespoke software approach used by IT Innovation (iWorkflow). The student was already familiar with the Simul8 software package through a course they had done on Management Science and Operational Research, which also meant they had a background in modelling techniques. The student took approximately 2 person months to complete the modelling work, including understanding the problem. The starting point for the student was more advanced than for the IT Innovation work as the workflow to be modelled was already defined in some detail at that point. An extra person month of effort would have been required if the workflow to be modelled had not already been specified. The model developed was not as sophisticated as that created by IT Innovation and did not include some of the key features of the workflow, e.g. some aspects of resource contention. The model was also not calibrated using BBC statistics on previous execution of the D3 workflow.

A reasonable estimate of the effort needed to build an equivalent model using Simul8 from the ground-up and from a 'green-field' starting point would be 4-6 person months of effort by someone who had at least some basic training for the tool.

These are only two sample points of the effort required to build a workflow model. The effort required also depends hugely on the skills and experience of the individuals involved. However, it would appear reasonable to conclude that building detailed models of the processes involved in large-scale digital transfer projects can take effort measured in months rather than days. This has implications on the cost/benefits of investing in detailed simulation and modelling and suggests that only large projects are likely to benefit from this approach. For smaller organisations and smaller projects, simpler models are likely to provide a better return on investment, for example using simple 'back of the envelope' calculations and spreadsheet based approaches.

[6] http://www.simul8.com/

5 Relationship to other PrestoPRIME deliverables

Although the modelling work done by IT Innovation is available as 'stand-alone' tools, there are many links between this work and the rest of the project. These include:

- The initial PrestoPRIME Deliverable D2.1.2 on modelling provided some examples of costs and failure rates for storage and the trends for how these change over time.

- Deliverable D2.1.1 "Preservation Strategies" considered the different approaches to preserving AV content (migration, emulation etc.) including the costs involved.

- Failure rates seen in audiovisual archives using IT storage technologies, e.g. data tape libraries. Some examples are provided in PrestoPRIME Deliverable D3.2.1 "Threats to data integrity from the use of large scale storage management environments".

- WP3 work on JPEG2000 and the impact of data corruption provides input to iModel both in terms of error rates and the impact of errors in terms of usability of JPEG2000 content. See JPEG2000 part of ID3.1.3.

- WP6 work on cost models and the information being gathered through the current survey on costs provides inputs to our models on the cost of preservation activities. See Deliverable D6.3.1 "Financial Models and Cost Calculations"

- WP3 work on automated video quality analysis has fed into our modelling work of digitisation and transfer workflows, in particular in terms of options that can be used in quality control stages of the workflow to supplement or replace labour-intensive manual QC. See ID3.2.3.

- The tools we have developed have been integrated with the service management tools in WP3. This allows monitoring data collected from real systems to be used to calibrate the models as well as the models to be used for forecasting the possible impact of applying different service management policies. See ID3.4.4 for details of the approach taken.

- The tools we have developed fit in the Preservation Planning part of the OAIS model and in that way are available for use alongside both P4 and Rosetta. See WP5 D5.2.3 2Advanced Prototype of Open PrestoPRIME reference implementation" and D5.3.2 "ExLibris Preservation System"

- The modelling tools have been tested/evaluated in the PrestoPRIME testbeds. The reports on these testbeds present evaluation findings that show what others consider good and bad about the tools and where further enhancements can be made. See D8.1.1 "Technological Showcase" and D8.2.1 "Report on the Final Evaluation Phase".

6 More information

The tools are available online from a website hosted by IT Innovation: http://prestoprime.it-innovation.soton.ac.uk iModel is available for download in both binary format and LGPL open source. The website includes a bug and issue tracker, links to reports and publications, and some examples of the use of the tools.

The tools have been presented at the following events in 2011 and 2012.

- FIAT 2011 (PrestoPRIME workshop)
- IASA 2011 (conference presentation)
- AMIA 2011 (PrestoPRIME workshop and open-source tools session)
- IBC 2011 (conference presentation)
- Screening the Future 2011 (conference presentation)
- Screening the Future 2012 (workshop on cost modelling)
- FIAT 2012 (PrestoPRIME workshop)
- PASIG 2012 Dublin (conference presentation)
- PrestoPRIME showcase 2012

The presentations at IBC and IASA resulted in invitations to have the respective papers included in the following journals:

- SMPTE Motion Imaging Journal (Jan/Feb 2012)
- IASA Journal (Jan 2012)

Full details of journal and conference publications are below.

- Wright R. (2011). Storage Strategy Tools. Presentation at the International Association of Sound and Audiovisual Archives (IASA) 42nd Annual Conference. Frankfurt, Germany, 3-8 September 2011

- Addis, M., Jacyno, M., Hall-May, M., McArdle, M. and Phillips, S. (2011) PLANNING AND MANAGING THE 'COST OF COMPROMISE' FOR AV RETENTION AND ACCESS. In: 2011 Conference of the International Broadcast Convention, 6-11 September, Amsterdam.

- Addis M. (2011) Cost and Risk modelling. Presentation at PrestoCentre training event. 12-16 Sep 2011. INA, Bry-sur-Marne, Paris.

- Addis M. (2011) Storage and Services: Planning and managing cost, quality and risk. Presentation at the FIAT/IFTA world congress 2011 as part of the PrestoPRIME pre-conference workshop. 28 Sep 2011. Turin, Italy.

- Addis, M., Allasia, W., Bailer, W., Boch, L., Gallo, F., Schallauer, P. and Phillips, S. (2011) Digital preservation of audiovisual files within PrestoPRIME. In: 9th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods, 28-30 September, Barcelona.

- Hall-May M. (2011) Storage and Services: Planning and managing cost, quality and risk. Presentation at AMIA 2012, 18 Nov 2011. Austin, Texas.

- Addis, M., Wright, R. and Weerakkody, R. (2011) Digital Preservation Strategies: the cost of risk of loss. SMPTE Motion Imaging Journal, 120 (1).

- Addis, M., Jacyno, M., Hall-May, M. and Wright, R. (2012) Storage Strategy Tools. IASA Journal, 38.

- Addis, M., Jacyno, M., Hall-May M, Phillips S, McArdle M. (2012). PLANNING and MANAGING the 'cost of compromise' for AV RETENTION and ACCESS. SMPTE Motion Imaging Journal, January/February 2012.

- Addis, M (2012). Modelling cost, risk and loss. Presentation at Screening the Future II. 22 May 2012, University of Southern California, Los Angeles, USA.

- Addis, M. (2012) Keeping audiovisual content alive: Estimates of cost, risk and loss. Presentation at INA Experts Forum, 18 June 2012, INA, Paris.

- Addis M. (2012) Cost and Risk modelling. Presentation at PrestoCentre training event. 10 Sep 2012. INA, Bry-sur-Marne, Paris.

- Hall-May, M. Planning and Managing Automated Services for Ingest, Storage and Access. Presentation at Preservation and Archive Special Interest Group (PASIG). 18 Oct 2012, Dublin, Ireland.

- Addis, M. (2012). Cost Model for Forever Storage and Access. Presentation at Preservation and Archive Special Interest Group (PASIG). 18 Oct 2012, Dublin, Ireland.

The tools have been referenced in posts on the PrestoCentre blog:
http://www.prestocentre.org/blog/203

Appendix A: cost modelling

Long-term cost modelling is an active area of research and development. This section reviews some of the current activities by others in the field.

Many institutions are interested in the Total Cost of Preservation (TCP) over time for their assets. Existing cost modelling approaches for long-term cost estimation can be split, roughly speaking, into three main classes:

1. Empirical models based on the preservation lifecycle where costs are estimated for each of the functions at the different stages of the lifecycle. Each stage is broken down into smaller and smaller functions until specific cost estimates can be calculated.

2. Cost estimates based on previously incurred costs of similar preservation projects or activities. Data collected from past experience is extrapolated or interpolated to predict future costs.

3. Simulations of the operation of a repository based on the services provided, processes followed and resources used. Cost data is calculated, collected and aggregated as the simulation progresses.

Lifecycle cost models.

The Digital Preservation Coalition defines7 preservation as “Digital Preservation Refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary.” Preservation means enabling access, i.e. ensuring that data can be correctly interpreted and used by a designated community. Preservation involves activities across the complete content lifecycle of and hence many cost modelling approaches analyse the costs associated with each stage in the lifecycle.

Examples of this approach include:

- LIFE model developed by the British Library[8].
- KRDS model for research data developed by Neil Beagrie[9].
- California Digital Library (CDL) Total Cost Preservation (TCP) model[10].
- Danish National Archives (DNA) Cost Model Digital Preservation (CMDP)[11.]

The Digital Curation Centre (DCC) lifecycle model provides a useful checkpoint when deciding what is in or out of scope of a cost model since it highlights the broad range of activities involved in preservation beyond retention of digital objects, e.g. planning, community watch and metadata.

[7] http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts
[8] http://www.life.ac.uk/2/documentation.shtml
[9] http://beagrie.com/krds-i2s2.php
[10] http://wiki.ucop.edu/display/Curation/Cost+Modelling
[11] http://www.ijdc.net/index.php/ijdc/article/viewFile/177/246

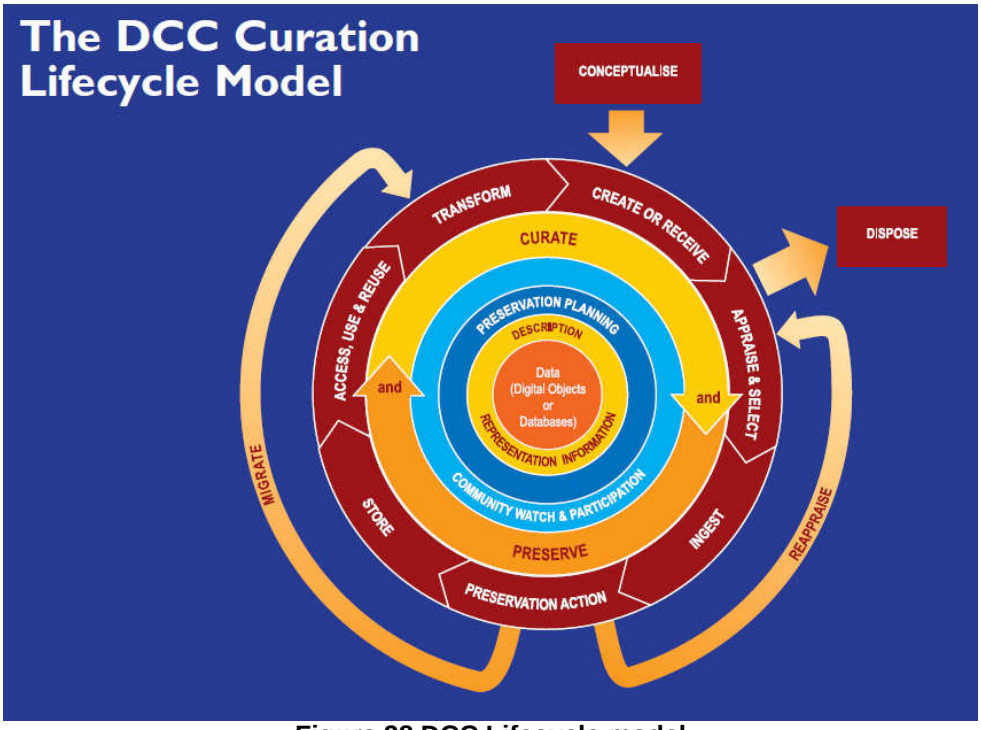


Figure 28 DCC Lifecycle model
(reproduced from the DCC website http://www.dcc.ac.uk/lifecycle-model/)

ISO16363 Trusted Digital Repositories (evolution of TRAC: trusted repositories audit criteria) is also worth a mention. Although not a cost model or lifecycle per se, ISO16363 does specify a set of criteria for assessing the trustworthiness of a repository and provides some guidelines on how to meet those criteria. In this sense it provides an additional checklist on what will be needed within a repository and hence where costs will be generated.

LIFE

A general approach to cost modelling is the work of the LIFE project[12] coordinated by the British library. This has developed a detailed lifecycle for digital preservation and then developed methods to estimate the costs of each stage in the lifecycle over time.

The basic elements of the cost model are:

$$L_T = Aq + I_T + M_T + Ac_T + S_T + P_T$$

Figure 29 LIFE cost model (reproduced from How much does it cost? The LIFE Project - Costing Models for Digital Curation and Preservation[13])

Where:

- L_T = Total cost

[12] http://www.life.ac.uk/
[13] http://liber.library.uu.nl/publish/issues/2007-3_4/index.html?000210

- Aq = Acquisition cost
- I = Ingest cost
- M = metadata cost
- Ac = Access cost
- S = Storage cost
- P = preservation cost.

The subscript T means that costs have to be calculated over the lifetime of the items being preserved.

A valuable output from LIFE are the series of case studies[14] (web archiving, e-Journals, newspapers etc.) from the partners involved that include detailed spreadsheets implementing the LIFE model and provide real-world worked examples of what the costs of preservation really are. The LIFE approach is founded on considering cost over time, e.g. for 5,10 or 20 year periods, with the result that the LIFE model and examples include activities such as migration and time varying costs such as storage.

Acquisition	Ingest	Metadata	Access	Storage	Preservation
Selection (Aq1)	Quality Assurance (I1)	Characteri-sation M1)	Reference Linking (Ac1)	Bit-stream Storage Costs (S1)	Technology Watch (P1)
IPR (Aq2)	Deposit (I2)	Descriptive (M2)	User Support (Ac2)		Preservation Tool Cost (P2)
Licensing (Aq3)	Holdings Update (I3)	Administrative (M3)	Access Mechanism (Ac3)		Preservation Metadata (P3)
Ordering & Invoicing (Aq4)					Preservation Action (P4)
Obtaining (Aq5)					Quality Assurance (P5)
Check-in (Aq6)					

Figure 30 Breakdown of cost elements in the Life model (reproduced from How much does it cost? The LIFE Project - Costing Models for Digital Curation and Preservation)

[14] http://www.life.ac.uk/2/documentation.shtml

California Digital Library (CDL) Total Cost of Preservation (TCP) model

The CDL approach is to take a OAIS and service oriented view of the cost modelling problem[15]. The elements of the model are shown below.

Preservation activities are embodied in an archival

1. ***System***; composed of various
2. ***Services*** supporting necessary and desirable functions; running on
3. ***Servers***; designed, deployed, maintained, enhanced, and utilized by
4. ***Staff***; in support of content
5. ***Producers***; *who use*
6. ***Workflows*** to submit instances of
7. ***Content Types***; which occupy
8. ***Storage***; and are subject to ongoing
9. ***Monitoring***; and periodic
10. ***Interventions***; all subject to appropriate managerial
11. ***Oversight***.

Figure 31 CDL breakdown of preservation activities reproduced from Total Cost of Preservation (TCP): Cost Modelling for Sustainable Services

The model considers the need to support Producers who submit content to be preserved, but not Consumers who subsequently need to access and use that content. Access can be a major if not dominant cost in digital preservation.

$$TCP = A + n \cdot P + m \cdot W + \ell \cdot C + k \cdot S + j \cdot M + i \cdot V + O$$

TCP	Total cost of preservation for all *Producers*.
A	Fixed cost of the baseline archival *System*.
n	Number of content *Producers*.
P	Unit cost of supporting a *Producer*.
m	Number of submission *Workflows*.
W	Unit cost of supporting a *Workflow*.
ℓ	Number of *Content Types*.
C	Unit cost of supporting a *Content Type*.
k	Number of units of preservation *Storage*.
S	Unit cost of *Storage*.
j	Number of preservation *Monitoring* activities.
M	Unit cost of a *Monitoring* activity.
i	Number of preservation *Interventions*.
V	Unit cost of an *Intervention*.
O	Fixed cost of administrative and managerial *Oversight*.

Figure 32 Cost model parameters reproduced from Total Cost of Preservation (TCP): Cost Modelling for Sustainable Services.

[15] https://wiki.ucop.edu/download/attachments/163610649/TCP-total-cost-of-preservation.pdf?version=5&modificationDate=1336402730000

The model divides costs into fixed costs and recurring costs where costs can be one off or recurring. This allows the model to include one-off capital expenditures, e.g. large items of equipment that are needed to establish a service, as well as unit costs where the total cost of the service is proportional to usage, e.g. units of storage.

To calculate the cost per Producer, the approach is to apportion fixed costs equally across all n Producers and then add the unit costs incurred by the specific Producer. This results in a Pay As You Go (PAYG) cost per Producer for a period of time, e.g. a year.

By then applying a discount factor d to the PAYG cost G, the long-term total cost is calculated over T periods.

$$G(T,d) = G \cdot \frac{1-(1-d)^T}{d}$$

Finally, the model calculates the Paid Up Price (otherwise known as an endowment) that is paid by the Producer in order to cover the total cost. This includes the interest earned by the cash whilst it is drawn down to pay the costs of the service.

$$F(T,d,r) = G \cdot \frac{e}{r} \cdot \frac{(1+e)^T-(1-d)^T}{(1+e)^T \cdot (e+d)}$$

r is the nominal annual percentage rate (APR) of investment return and e is the effective annual rate including monthly compound interest.

$$e = \left(1+\frac{r}{12}\right)^{12} - 1$$

This then leads to the ability to compare PAYG (including discounts) with Paid-Up pricing.

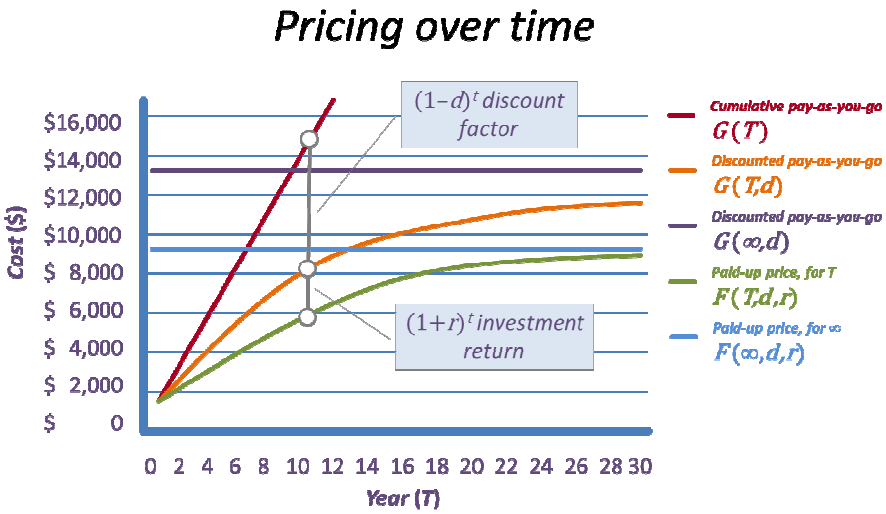


Figure 33 Total cost over as a function of time. Reproduced from Total Cost of Preservation (TCP): Cost Modelling for Sustainable Services

KRDS (Keeping Research Data Safe)

The KRDS cost model takes a lifecycle approach (see stages in table below taken from KRDS user guide) to breakdown costs into different categories. The recommendation is then to build a spreadsheet model of the total cost, inc. annual discounting for estimating long term costs. In this respect, the KRDS model is similar to LIFE and other lifecycle/spreadsheet approaches.

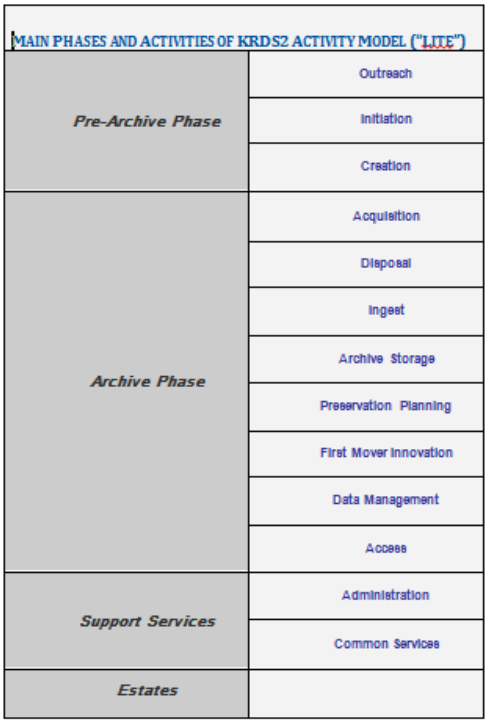


Figure 34 Main phases/activities in the KRDS model (reproduced from KRDS2 activity model http://www.beagrie.com/krds.php)

More interesting in KRDS are tools/guidelines for doing a corresponding value and benefits analysis[16] This considers internal and external beneficiaries, direct and indirect benefits and whether the benefits accrue in the short or long term. Examples are included of generic benefits for research data. By considering tangible and intangible benefits, the model is not dissimilar from earlier work done by eSpida[17] that used a Kaplan and Norton balanced scorecard approach.

[16] http://www.beagrie.com/intro_benefits%20analysis%20toolkit_0711.pdf
[17] http://www.gla.ac.uk/services/library/espida/

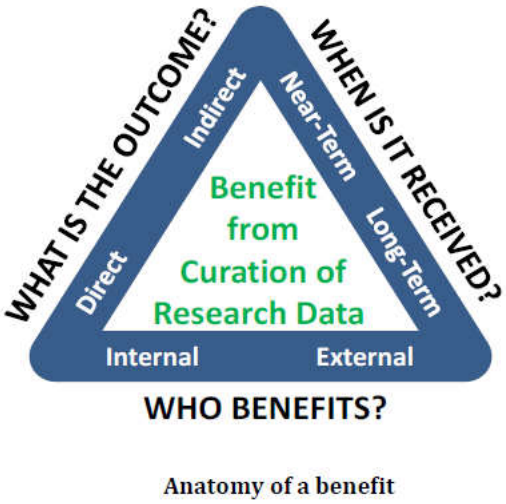


Figure 35 KRDS model of benefits of research data curation. Reproduced from KRDS toolkit guide http://www.beagrie.com/krds.php

Danish National Archive (DNA) Cost Model Digital Preservation (CMDP)

Like the TCP model from CDL model, the CMDP from DNA also bases itself on OAIS. Perhaps the most interesting part of the model is how it treats the costs that arise from the different formats that need to be handled (documents, images etc.) as shown in Figure 36.

*Format Interpretation (pw) =number of pages * time per page (min) * complexity (L, M, H)*

Format	Specifications and other relevant documentation	No. of pages	Complexity	Quality
TXT	ISO 10646 ISO 646	20 15	L L	H H
PDF/A 1.0	PDF/A (ISO 19005-1) PDF 1.4 (ISO 32000-1)	29 700	L H	M M
TIFF 6.0 LZW	TIFF 6.0 Baseline LZW (ISO 12639:2004)	121	M	H
GML 3.X	ISO 19136 2007 ISO 19100-serie (Open GIS) 19103 19104 19107 19108 19109 19111 19123 (understanding of xml, xml schema and Xlink assumed)	380 67 102 166 48 71 78 65	H	H

Table 7 Examples of how different formats' documentations (no. of pages, complexity and quality) have been evaluated as basis for calculating the Format Interpretation factor.

Figure 36 Variables used in establishing a measure of effort associated with the handling and preserving of different file formats. Cost in this sense is a 'Format Interpretation' factor that is subsequently used to estimate the effort involved in different activities involving a format, e.g. writing or supporting software to read or migrate that format. Reproduced from CMDP project report http://www.costmodelfordigitalpreservation.dk/

This is then used to calculate the cost of different preservation activities, e.g. format migration which includes development/testing of format conversion tools and use computational resources to do conversion.

For example, the migration cost in person weeks of effort is calculated to be:

Migration Cost (pw): Format Interpretation + Software Provision + Migration Processing

The DNA then compared the predictions made by their model with costs of actual preservation projects done in the past. This is shown in Figure 37 for one of the projects. The difference is significant in most areas, with the model typically estimating costs that are higher or lower than reality by 50% or more. This shows that accurate cost prediction is not easy. Indeed, getting within 50% should be considered a good result.

	Case 1		CMDP		CMDP - Case 1	
	pw	%	pw	%	Δ pw	%
IP Designs	**44**	**12**	**50**	**24**	**6**	**12**
A (1968-1998)	29	66	20	40	-9	-31
B (1999-2000)	15	34	16	32	1	6
C (2001-2004)	0	0	14	28	14	n.a.
B & C	15	34	30	60	15	50
Migration Plans	**150**	**42**	**39**	**19**	**-111**	**-74**
A (1968-1998)	105	70	15	38	-90	-86
B (1999-2000)	30	20	14	36	-16	-53
C (2001-2004)	15	10	10	26	-5	-33
B & C	45	30	24	62	-21	-47
Prototypes (Software Provision)	**164**	**46**	**116**	**57**	**-48**	**-29**
A (1968-1998)	101	62	48	41	-53	-52
B (1999-2000)	50	30	36	31	-14	-28
C (2001-2004)	12	7	32	28	20	62,5
B & C	62	38	68	59	6	9
Migration Package (total)	**358**	**100**	**205**	**100**	-153	**-43**

Figure 37 Comparison of actual costs (expressed as person weeks of effort) for a preservation project (Case 1) with the predicted costs from the CMDP model. Reproduced from CMDP project report http://www.costmodelfordigitalpreservation.dk/

Cost models based on historical data

The most relevant example here is the NASA Cost Estimation Toolkit (CET)[18]. This is used to calculate mission costs and has limited detail on the long-term retention and access to data from a mission. CET is a useful reminder that probably the best way to estimate future costs is from past experience – but only if there is enough data to hand – which in turn requires a proactive effort to ensuring the right data is collected from the outset. It is unlikely that this data is available for AV preservation and therefore the CET type approach was ruled-out. However, in the long-term as a AV preservation service becomes established and operated, this approach should become increasingly viable.

[18] http://opensource.gsfc.nasa.gov/projects/CET/CET.php

Cost models based on simulation

This is the approach currently taken by David Rosenthal (Stanford University) for his long-term cost modelling work, much of which is described on his blog[19]. David's current focus is on the effects of uncertainty or variability of the inputs to a cost model (e.g. interest rates) on long-term costs through use of Monte Carlo techniques[20].

A simulation approach is also taken by IT Innovation to cost modelling for audiovisual (AV) preservation. Our approach is to model preservation processes and their associated costs using a Discrete Event Simulation with a stochastic approach to event generation (e.g. ingest or access workloads, data corruption).

The main benefits of Monte Carlo or stochastic approaches are to generate a probability distribution for costs over time. This allows actuarial analysis of long-term costs, e.g. what is the probability that costs will not go above a given limit. For example, this can be important when considering 'endowment models' where the question is what one-off sum of money needs to be invested today to secure the long-term preservation of data, e.g. over decade or century timescales.

[19] http://blog.dshr.org
[20] http://blog.dshr.org/2011/09/modelling-economics-of-long-term-storage.htm

Appendix B: D3 workflow

This section describes the specification of the workflow simulation developed for the BBC D3 project. The purpose is to show the level of detail needed to model a real-world transfer workflow. This level of detail is significant and the model developed is still a simplification of exactly what happens in the real-world. The section concludes with some possible variations that could be made to the existing BBC workflow as hypothetical examples of candidate improvements. The objective is to see whether these could easily be coded into the simulation so that the simulation could then be used to do a cost/benefit analysis of the different options. The details are based on a simplified snapshot of the D3 process from around June 2011. In particular, some of the numerical parameters have values that are representative but fictitious. The objective is to provide examples so that the reader understands what each parameter of the model is for and what the model produces. Actual numbers from the BBC workflow have not been included for confidentiality reasons.

D3 description and workflow options

The D3 project at the BBC is an effort to migrate video from approx. 100,000 D3 tapes into file format (MXF wrapped uncompressed) and store it on LTO data tape.

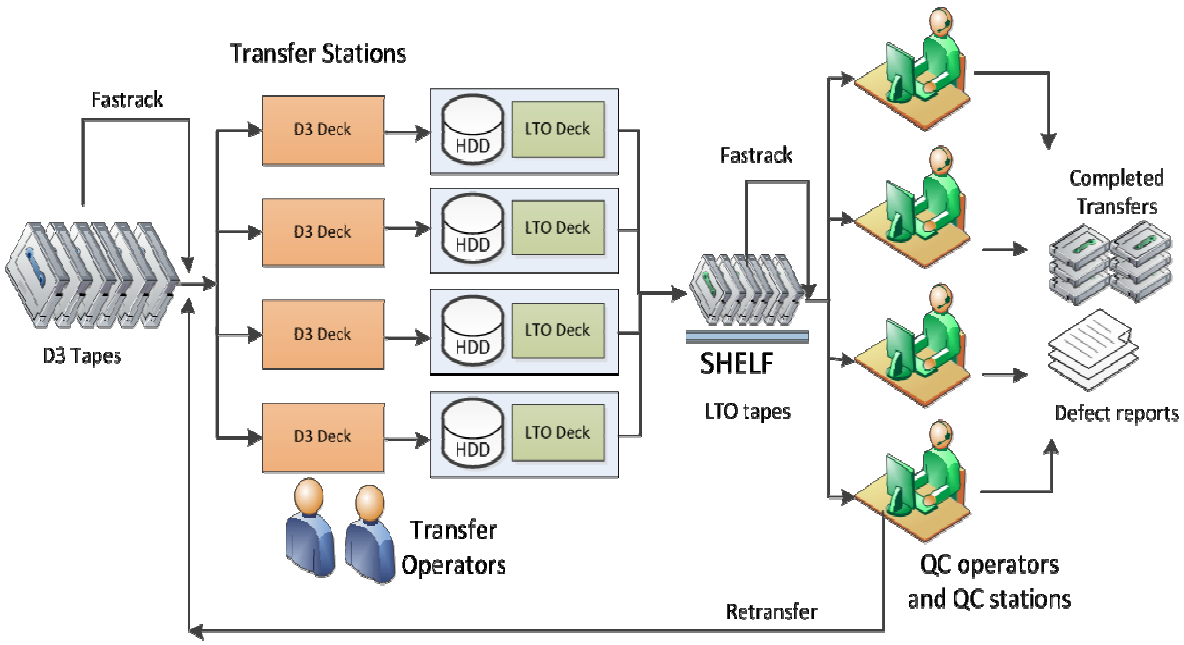
Technical details can be found in the BBC whitepaper 155:

http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP155.pdf

In particular, see the sections on ‘Archive and Preservation’ and ‘Use in an archive environment’.

Current D3 workflow

For the purposes of simulation/modelling, a shorthand description is below.



- Workflow starts with a pile of D3 tapes (e.g. 100,000). Each tape contains 1 TV 'program'. The program length is variable. Average program length is approximately 45 minutes. SD video (and audio etc) uncompressed at approx. 170Mbit/sec = 77GB per hour, so average file size is approx. 55GB.

- Each tape is loaded into a 'D3 deck' (in reality this is a D3 deck augmented with other equipment e.g. PAL transform decoder) that produces a real-time data stream (SDI). The deck also generates markers (metadata) that indicate data read problems. Real-time means that a 20 min program will take 20 minutes to playback and data to be recorded.

- A DigiBeta copy is created at the same time as the D3 transfer (essentially tape to tape copy). This goes back to the archive. A 'browse' quality proxy file and other data files are also generated.

- If a D3 tape is flagged for re-transfer (e.g. because of errors picked up in QC) then it goes to the front of the D3 transfer queue.

- When a D3 transfer is repeated then a new DigiBeta tape is created. This goes to the archive and supersedes the previous DigiBeta tape.

- The SDI stream from each D3 deck is recorded into a file with a MXF wrapper (using ingex). This is done by a dedicated machine attached to each D3 deck. The output goes onto a local HDD.

- The file from a D3 transfer may be 'chopped up' into smaller files where each of the smaller files corresponds to a specific programme or part of a programme. Where to chop is defined by catalogue entries in infax (BBC catalogue system).

- When the HDD cache in the ingex machine is full, the MXF files are written to data tape (LTO3). The disk cache copy is then deleted. An LTO3 cartridge can store several hours of D3 programme material, so LTO writing is initiated manually by an operator when the HDD cache is full, e.g. once a day.

- The tapes are put on shelves pending Quality Control (QC). They can be there for up to 6 months.

- High priority material can be 'fast tracked'. This could occur by putting it at the front of the queue for D3 transfer and then taking the result straight for QC, or it could be a D3 transfer that has already been done and fast-tracking the LTO tape straight to QC. Approx 15 MXF files are fast tracked through QC per day. This is out of the up to one hundred files that the QC workflow will normally process each day.

- One item per D3 channel per day is fast tracked. This ensures that any problems with the D3 workflow is picked up early, e.g. it won't be 6 months before a problem is detected and hence all the output from a given D3 deck has to be repeated.

- When an LTO is to be QC checked, it is placed in a red bag along with all the D3 tapes that were used to produce the MXF files on that LTO. The relevant paperwork and an empty green bag are also placed inside each red bag.

- Red bags are queued up in the QC area. Each Primary QC operator will work from one red bag at a time. If they complete the QC of a red bag before the end of their shift then they move onto another red bag. If they do not finish the QC then the red bag is left at the workstation for the next operator – this is because the contents of the LTO in the red bag will have been (partially) extracted onto that particular workstation.

- A QC operator takes the LTO tape from the red bag and loads it into a QC workstation, (partially) extracting the files onto the workstation's local HDD as required. Files can be extracted from the LTO while other files are being viewed, but only files that are 100% extracted can be viewed. The files are viewed and the operators check for defects. They do this in two stages.

    - The first stage is checking the problems flagged up by the D3 deck. They only look at the specific parts of the AV flagged as problematic. The operator can take the original D3 tape and use a communal D3 deck in the QC area to look at areas of concern. This phase also includes global checks e.g. on file timecode continuity and MXF file metadata.

    - The second stage is a full end-to-end manual check of video and audio to pick up any other defects – for example, because the D3 tape is actually the output of a previous migration, e.g. 2" Quad or 1" video tape, and defects exist because of the previous migrations or carriers.

- The video is viewed in real time. There may be cases where the video is paused, e.g. for notes to be added.

- The result of the Primary QC process is a defect report (typically referred to as a 'QC Report') on each MXF file. The defect reports are created electronically on the QC workstation. A paper copy is also printed and attached to the tape that has been QC'ed. The electronic QC reports are harvested from the QC workstations and used to build a database of information on QC.

- As the Primary QC of an LTO is completed, the contents of the red bag are transferred to the green bag provided. When QC of the LTO is complete the red bag will be empty and the green bag will be ready for collection by the logistics team.

- The Primary QC operators must delete files from their workstations. When they complete the QC of an LTO then all the MXF files extracted from that LTO must be deleted.

- The logistics team process the green bags: the original D3s are still kept, the LTO goes back to the archive, and the paperwork gets barcode scanned then filed. In effect, this barcode scanning 'activates' the relevant QC Reports that have been harvested i.e. a QC Report will not be 'visible' in the database (e.g. to Secondary checkers) until both: a) its electronic version has been harvested; and b) its paper version has been barcode scanned.

- Tapes that fail QC go back for re-transfer. The paper QC report is kept with the tape so it can be inspected when re-transfer is attempted.

- re-transfer request will only come after the tapes has gone through the secondary checker.

- The two stages of Primary QC are performed by the same Primary QC operator. Note that the operator may 'fail' an MXF file prior to completing both stages and abort the QC process (see the operators' manual for details of this).

- Spot checks are done on the quality of work by the QC operators. This is called secondary QC. This is done by selecting a batch of tapes that have gone through a given QC operator and checking them against the defect reports. The tapes that have passed QC are the target of QC checking, i.e. the contents of a 'green bag'. The QC of the QC operators lags behind by as much as a couple of weeks (there are only a few operators who do the spot checks). The contents of 'green bags' won't go back to the archive if they are awaiting a spot check.

- Secondary QC operators' are responsible for approving or overturning all the MXF file PASS / FAIL decisions made by Primary QC operators. All tapes passing through primary QC will also undergo secondary QC.

- Secondary QC operators also perform spot checks on MXF files (additional to reviewing the PASS/FAIL decisions). Spot checks are not done on all MXF files.

- There is a queue of content (although not a physical queue as elsewhere in the process) between the Primary QC output and the Secondary QC input.

Notes:

- D3 transfer

    - There are human operators of the D3 decks. Typically 1 operator can monitor and manage 4 D3 decks (e.g. loading/unloading tapes, checking signal levels etc, logging transfers). There are currently 8 decks.

    - If a D3 tape contains x minutes of programme material, then end-to-end transfer will typ. take longer (time needed to load/unload tape, check levels, do some housekeeping etc.). This can be modelled as a fixed overhead (e.g. 5 minutes) plus a multiplied of the programme material length (e.g. 1.5x)

    - D3 decks can be modelled as having a limited headlife (e.g. 10,000 hrs – BBC to confirm the right value). As tapes go through the deck the remaining headlife is decremented. If the headlife is used up then a replacement head/deck is needed. Therefore, an output of a simulation should be number of head hours used (inc. re-transfers etc.) and frequency of head expiry.

    - D3 transfer is not a 24x7 operation. It takes place in shifts. The working hours for D3 transfers and QC are not the same.

    - The use of one LTO drive per D3 deck is potentially inefficient. D3 transfer creates files at approx. 170Mbit/sec. LTO3 data write rate is 80MB/sec, i.e. 640Mbit/sec. This provides a potential optimisation opportunity, e.g. if the BBC move to LTO5, then the write rate would be 140MB/sec (over 1

Gbit/sec) and hence 1 drive attached to a shared SAN across all the D3 decks could cope with an entire day of D3 transfer.

  - Errors that can occur during transfer including failing to capture the SDI stream properly to file (e.g. dropped frames) and failures in PAL transform done on the video content. Currently these are detected during QC, but the BBC are working in an automated checking tool that will pick up problems.

- The QC operators are separate from the D3 transfer operators. The QC operators can only process one LTO tape at a time. They have their own dedicated workstation (LTO drive, playback computer, monitor). If the length of a video item is x minutes then it typically takes 2 or maybe 3 times as many minutes to QC the item.

- QC takes considerably more time and effort than initial transfer. The limits on QC staff and workstations means that substantial numbers of items can get queued up for QC, e.g. an item can be in the queue for QC for several months.

- If an item fails QC then re-transfer may be attempted, e.g. if the defects come from the D3 deck (e.g. because tape or deck have problems etc).

  - It is not possible to predict this from the number of defects flagged up by the D3 deck during original transfer. The percentage of tapes that need to be re-transferred is between 5 and 7%.

  - The time taken to do a second transfer is not known, but we know the general rule from the BBC that if a tape doesn't playback easily first time, then it needs 5x the effort second time round (e.g. due to specialist preparation, equipment, monitoring etc.). Not clear whether this rule holds for D3 as it is more automated than most transfer chains, but it's a good enough default. So, assume that transfer time is x2 programme material time + 0 minutes overhead. Assume the operator is dedicated to the transfer (as opposed to handling multiple concurrent transfers for first attempt). This means that if an operator is diverted to do a re-transfer then the set of normal D3 transfers they would otherwise be doing has to halt.

  - The success rate of a second transfer is not known. We can assume 100% until the BBC tell us otherwise.

  - The D3 decks have limited head-life. Re-transfer is not desirable due to the extra wear on the D3 heads.

- LTO tapes

  - The LTO tapes used by the QC operators occasionally have problems, which require a re-transfer of the relevant D3 content. BBC report that 8 out of 4500 tapes have had errors.

  - More than one programme is stored on an LTO3 tape. For example, LTO3 = 400GB capacity but it's not possible to use 100% without splitting files across tapes. Each MXF file is about 55GB on average and there are an average of about 6 MXF files per LTO. An operator will QC all the previously unchecked

MXF files on an LTO tape prior to returning it to the logistics team. This affects the average time that an item will take to go through the chain because it will be part of a batch of items that all have to be processed before any one item in the batch is finalised (i.e. an entire green bag must be returned to logistics and have its barcodes scanned before any of the MXF files are available in the database for Secondary QC – the QC of an MXF file is not 'finalised' until it has passed through Secondary QC).

- LTO drive problems are far more frequent than LTO media problems. This affects the QC workstations since if a drive fails then the workstation can no longer be used for QC. Any data files generated on that machine have to be moved to another machine, and any MXF files extracted have to be extracted again on a different workstation i.e. the QC of the affected LTO is delayed.

Input variables to the model include:

- D3 tapes
  - Number of D3 tapes (e.g. 100,000).
  - Length of programme material on each D3 tape (e.g. between 10 and 60 minutes with 20 minutes as the average).
  - Distributions of defects (a) from D3 transfer, (b) previously existing defects in the tapes. Assume it is linear with programme material length and independent of age of tape etc. (e.g. 2 defect per minute for transfer, 1 defect per minute already existing)
- Transfer
  - Transfer time as an overhead + multiplier of the length of programme material on D3 tape (e.g. 5 minutes + 1.5x programme length)
  - Number of D3 decks (e.g. 8)
  - Ratio of deck operators to decks (e.g. 1 to 4)
  - Cost per hour of transfer operator (e.g. £50)
  - Cost of D3 deck and associated equipment (e.g. £10k)
  - Transfer time for re-transfers, i.e. tapes returned from QC, e.g. 30 minutes + 2x length of programme material, dedicated operator.
  - Headlife of a D3 deck (e.g. 10,000 hrs)
  - Annual running cost of D3 deck, i.e. maintenance (e.g. £1k per year)
- Storage
  - Capacity of LTO tapes (e.g. LTO3=400GB,LTO4=800GB,LTO5=1500GB)
  - Cost of LTO tape (for each generation)

- Number of LTO drives used to create LTO tapes.
- Cost of LTO drives (e.g. LTO5 = £3k)
- Cost of HDD cache (£ per GB)
- Data rate of LTO tapes (for each generation)
- Amortisation time of tape drive (e.g. 3 years)
- Occurrence and impact (e.g. amount of D3 transfer station downtime) of LTO drive faults

- QC
  - Number of QC operators and QC stations (e.g. 10, noting that there is a 1 to 1 ratio).
  - Time spent by QC operator reviewing transfer defects (e.g. 30 sec per defect)
  - Time spent by QC operator logging transfer defects (e.g. 30 sec per defect)
  - Time spent by QC operator reviewing the full programme for pre-existing defects (e.g. 1.2x length of programme material).
  - Time spent by QC operator logging pre-existing defects (e.g. 1 minute per defect)
  - Time spent by QC operator waiting for files to extract from LTO in situations where there is not already at least 1 file 100% extracted onto their workstation (e.g. 15 mins per day)
  - Time spent by QC operator waiting for MXF files to be deleted from their workstations (e.g. 25 mins per day)
  - Probability distribution for how long a workstation is likely to have a red bag allocated to it yet no QC operator
  - Occurrence and impact (e.g. amount of wasted QC operator time and delay to LTO QC progress) of LTO drive faults
  - Cost per hour of QC operator (e.g. £30)
  - Cost per QC station (e.g. £20k)
  - Amortisation time of QC station (e.g. 5 years)
  - Percentage of D3 tapes that will need re-transfer (e.g. currently 5-7%)
  - Percentage of pre-existing defects not picked-up by QC operator (e.g. 5%).

Some ‘internal’ parameters of the model:

- Size of HDD cache (used in-between D3 transfer before LTO3 writing), e.g. 500GB
- QC queue length (number of tapes, length of time in queue) of LTO tapes pending QC

Outputs of the model:

- Throughput, i.e. rate at which items pass through the entire chain.
- Number of defects likely to remain undetected in output tapes.
- Average time for item to go through the workflow (bearing in mind queues).
- Annual total cost of whole set-up, i.e. sum over TCO over all parts inc. amortisation
- Cost per item (i.e. annual total cost for set-up divided by annual throughput)
- Headlife used and headlife remaining across the D3 decks.

Variants to the workflow

This section describes some possible variations to the workflow that could be useful to model using the tool to allow their value to be assessed before implementation.

QC Cache

Changes already underway, but not yet deployed, include a ‘QC cache’. This replaces the need for LTO drives in each workstation. Instead a shared network-accessible file store is accessed by each QC station and used to hold all the files in the QC process. For the purposes of modelling, we can say that LTO tapes are loaded overnight onto the file store and then the next day the QC operators can do QC directly against the file store.

For the purposes of modelling, the QC cache can be represented by X TB of HDD storage and Y LTO drives. A batch of LTO tapes are loaded into the drives and content transferred to the HDD store overnight. The number of tapes that are loaded in each night is the number needed to keep the store ‘topped-up’; however, the number of tapes per night is capped at a predefined value. Files are removed from the QC Cache manually by the Secondary QC operators when the whole QC process has been completed; however, each file will ‘timeout’ (and be automatically deleted) Z days after that file is Primary QC’ed.

In addition, the QC Cache system will be accompanied by a means to remove the barcode scanning of paper QC Reports from the critical path (this barcode scanning relies on a complete LTO being checked and its green bag being collected by logistics, hence introducing delays).

The benefits of the QC cache are:

a. No need for LTO drives in each QC station, which reduces down time due to drive failures

b. No handling of LTO tapes by QC operators, which reduces risk of damage, loss, contamination of tapes.

c. Fewer LTO drives.

d. Easy access to files between operators, e.g. by those doing quality spot checks on the PASS files from the QC operators.

e. Files for a given TV programme or set of programmes from one or more D3 tapes could be split across QC operators for accelerated fast tracking of urgent QC

f. Filestore can be used to make additional LTO copies (eventually there will be no DigiBeta, so current approach of 1 LTO3 + 1 DigiBeta will be replaced by 2xLTO4).

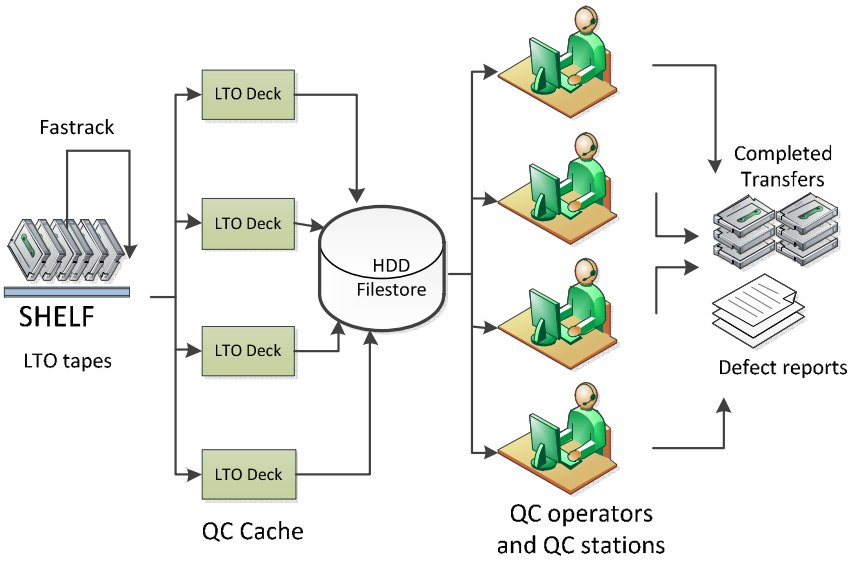
g. All the LTO tape extractions are handled by a central, dedicated system and so can be performed more efficiently.

h. It provides centralised control over the allocation of LTO tapes to Primary QC operators (currently operators can pick any red bag to work on).

Challenges:

- There is a tradeoff between:
    - Filestore capacity and its associated cost
    - Automatic ‘timeout’ of MXF files: the shorter this is, the more chance there is that the Secondary QC operators will not complete QC of an MXF file prior to it being automatically deleted.
    - The size of the ‘buffer’ of MXF files that have not been Primary QC’ed. The larger this is, the more protection there is against the Primary QC operators running out of MXF files to work on. This scenario could arise if they are more productive that usual on a particular day, or if the overnight extraction to the store failed for some reason.

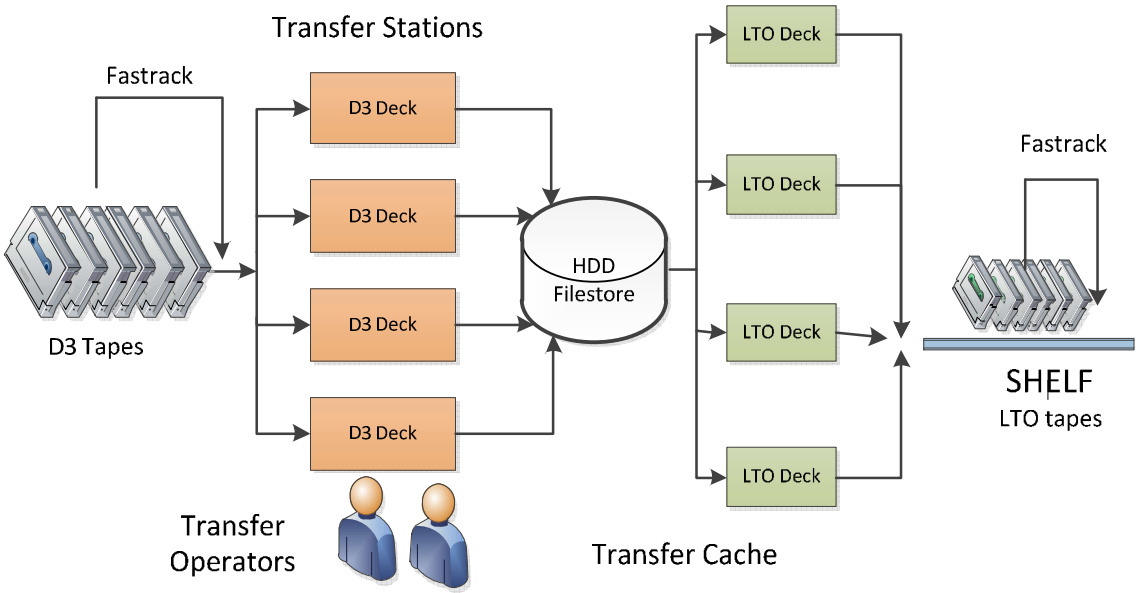
The use of a QC cache in the workflow is shown below:



Transfer cache

A shared HDD cache and LTO drives could be used instead of current set-up of one LTO drive and one HDD cache attached to each D3 deck. This is effectively the way that the model works currently, so we've already done this part!

Ultimately one HDD Filestore could be used for both the Transfer and QC processes, and LTO production need not happen until MXF files have passed the QC process. However, this would mean that the transfer and QC processes would be tightly coupled.



Automating the Primary QC Process

There are probably two phases to increasing the automation of the Primary QC process:

1. **Ensure the quality of the transfer.** Currently we know about all the points in the content where the output of the tape player may not have matched what was on the tape: these points are marked by 'flags' in the MXF file. A hardware device is currently being developed that will check the output of the tape player against what is recorded in the MXF file: this device will add additional 'flags' into the MXF file to highlight points at which problems may have occurred. When this hardware device is in service the Primary QC operator only needs to review the parts of the content that have been flagged – if they do not detect any problems with the content then we can be confident (to some unknown degree) that the transfer from tape to MXF file was successful.

2. **Detect historical faults.** If the Primary QC operator is no longer reviewing the entire length of the content then historical faults will not be detected. Software video quality analysis could be useful in detecting these historical faults e.g. the software could be used to add additional flags to the MXF file and these could be reviewed by the Primary QC operators.

    - What increase in undetected defects does this result in? (e.g. 20% more undetected defects with the software compared to a Primary QC operator reviewing the entire content.) Note that the software may well detect some types and instances of defect that a human could easily miss (e.g. 'stuck bits').

    - This requires a set of parameters on the JRS tool, e.g. item processing time, defect detection rate, cost of software, cost of machine needed to run the software. We could ask JRS for these or simply invent some parameters and default values.

    - There are several places where the tools might be used: (a) on the QC workstation, (b) immediately after the D3 decks in the transfer part of the chain, e.g. to extend the D3 transfer report to include software detected defects, (c) as a dedicated step in-between transfer and QC where further operators take the LTO tapes and run them through a defect detection process in advance of human QC

    - This software QC analysis need not be part of the 'Transfer & QC' workflow but could be performed at a later date. In this scenario, the question is: if this detection of historical faults is performed later, will tapes be discovered for which a transfer from an alternative source tape would be desirable (assuming that it would contain different historical faults), and would this be possible given the number of head hours remaining on the tape players at that time?

There are parameters that are important for modelling the benefit of both phases 1 and 2:

- False positive rate. Each false positive will require review by an operator. This increases operator time.

- False negative. These are real video defects that aren't detected by the software and because the QC operator no longer does a full scan of the video, they slip through the net.

Timeboxed QC

Reducing time spent on manual QC would save cost, but with consequent increase in defects that are undetected. This could be achieved in several ways, e.g. time-boxing the QC time for each item.

For example, if QC is dropped to 0.5x programme time for pre-existing defects then only 75% of defects might be detected rather than say 95%.

The question here is what fraction of a programme needs to be sampled to have a given confidence that at least Y% of defects will on average be detected. For example, if no defects have been found after viewing 25% of a programme at random, then it might be possible to assert that there is 90% confidence that no defects exist on the rest of the tape.

This approach requires statistical analysis of defect rates in tapes.

Any findings from this analysis would also benefit the process used by the Secondary QC operators and the spot checks that they perform.

Appendix C: storage cost-curves

Kryder's Law is often used to estimate future costs of storage. However, the cost per TB for storage is dependent on the total number of TB being stored. Therefore, the cost per TB in a long-term cost projection has to take into account the data volume being stored,

A simple example of falling cost of storage is shown below in Figure 38. This shows the cost per TB of storage for three IBM tape library solutions. The costs include library hardware, drives, media and slot licensing. Assuming the library price remains constant but the cost of media falls and capacity increases according to the LTO roadmap, the cost per TB can be estimated in the future. An entry level library capable of storing 10s of TB today is able to store PBs of data within a couple of decades. Or put another way, a fixed volume of data over a 20 year period that requires a large scale enterprise solution today will only require an entry level solution in 20 years.

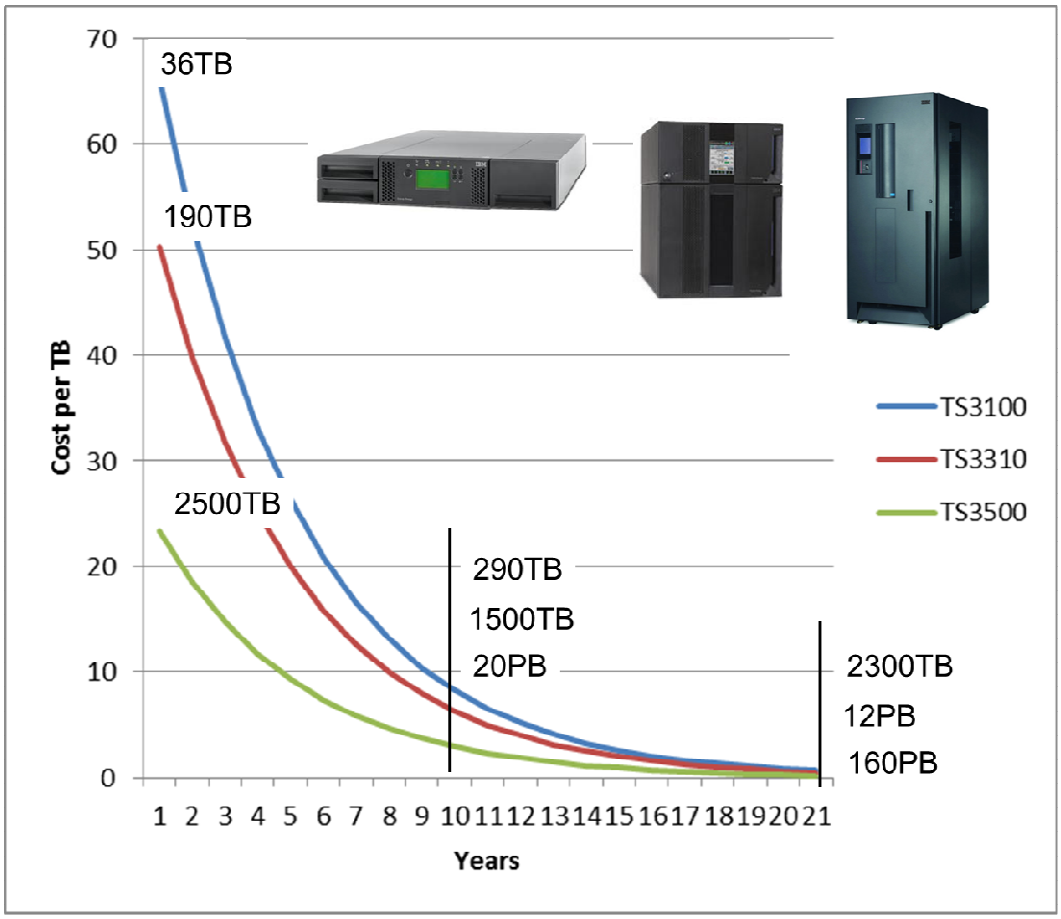


Figure 38 Projected cost per TB of storage for three models of IBM tape library

The key point is that the large scale solutions provide the lowest cost per TB and hence unless data volumes are growing an archive won't be able to continue to take advantage of this over time and will end up switching to ever small storage solutions. An archive will only be able to follow a given cost-curve if they store ever more data in order to stay on that curve.

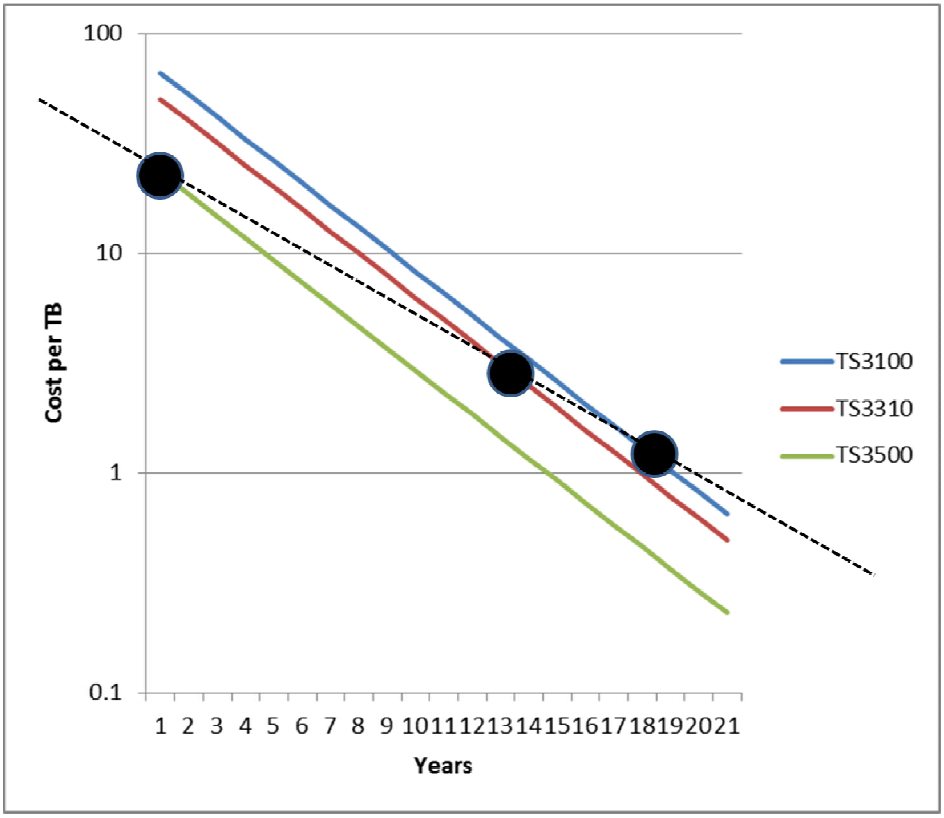


Figure 39 Jumping between cost curves

Figure 39 shows how the same cost curves as Figure 38 but on a logarithmic scale. The black dots and the dotted line that joins them is a fixed 2.5PB storage requirement. This requirement is initially met by the large-scale enterprise storage solution (TS3500) but over time can be met by the mid-range (TS3310) and then entry level solution (TS3100). These solutions are more cost effective at this data volume than sticking with the enterprise solution and only having it partially filled. The dotted line is the effective rate at which the cost per TB falls. This rate is lower than the rate that would be achieved if the data volume in the archive was increasing so that a fully utilised enterprise library was always the best option. In other words, the cost per TB hasn't fallen as fast as would have been calculated by taking the initial cost per TB and then applying Kryders law.

Appendix D: example iModel test case

Test1a: One copy model, latent corruption only

Naïve distribution of corruption

If corruption is not stochastic and the probability that a file is corrupted is proportional to the time spent in storage, then the questions are:

(a) what is the probability that the file is first lost in year n?

(b) what is the cumulative probability that the file is lost in year n or before?

Basically, this is survival analysis[21] for the file including look at the first hitting time of file loss[22]

P	Probability a file is corrupted each year
(1-p)	Probability a file is not corrupted in a year
$(1-p)^n$	Probability a file is not corrupted after n years
$1-(1-p)^n$	Probability a file is corrupted at least once after n years

If we set-up a 1 copy model with latent corruption where 1 in 10 files corrupted in 12 months, then after 10 years the probability of a file surviving is $0.9^{10} = 0.3487$

100,000 files over 10 years, we'd expect to see 65132 files corrupted when using iModel (or thereabouts give stochastic nature of corruption in the simulation).

Poison distribution of corruption

If the distribution of corruption events is Poisson[23] then the probability of k corruptions in Δt is given by:

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$ where λ = average number of corruptions per Δt

Therefore, the probability that at least one corruption takes place in Δt is:

$$a = 1 - e^{-\lambda}$$

If the *rate* of corruption is 1 in 10 files in 12 months, then the probability of at least 1 corruption in 12 months is 0.095162582.

Using this in the formula above, gives probability of file being corrupted at least once after 10 years as 0.63212 [24]. Therefore, for 100,000 files we would expect to see 63,212 corrupted.

[21] http://en.wikipedia.org/wiki/Survival_analysis
[22] http://en.wikipedia.org/wiki/First_passage_time
[23] http://en.wikipedia.org/wiki/Poisson_distribution
[24] Note that we get exactly the same result if we changed λ to be over a decade rather than 1 year.

iModel test inputs:

File size	100GB
Storage size	10000TB
Latent corruption	10 in 100 files every 12 months
Access corruption	0
Scrubbing	0 (i.e. turned off)
Ingest	none
Access	none

Expected output:

10 year simulation

10,000TB with 100GB files = 100,000 files

Files lost: (a) naïve distribution: 65132 files

(b) Poisson distribution: 63212 files

Observed result:

File lost: 63016 (Poisson distribution)

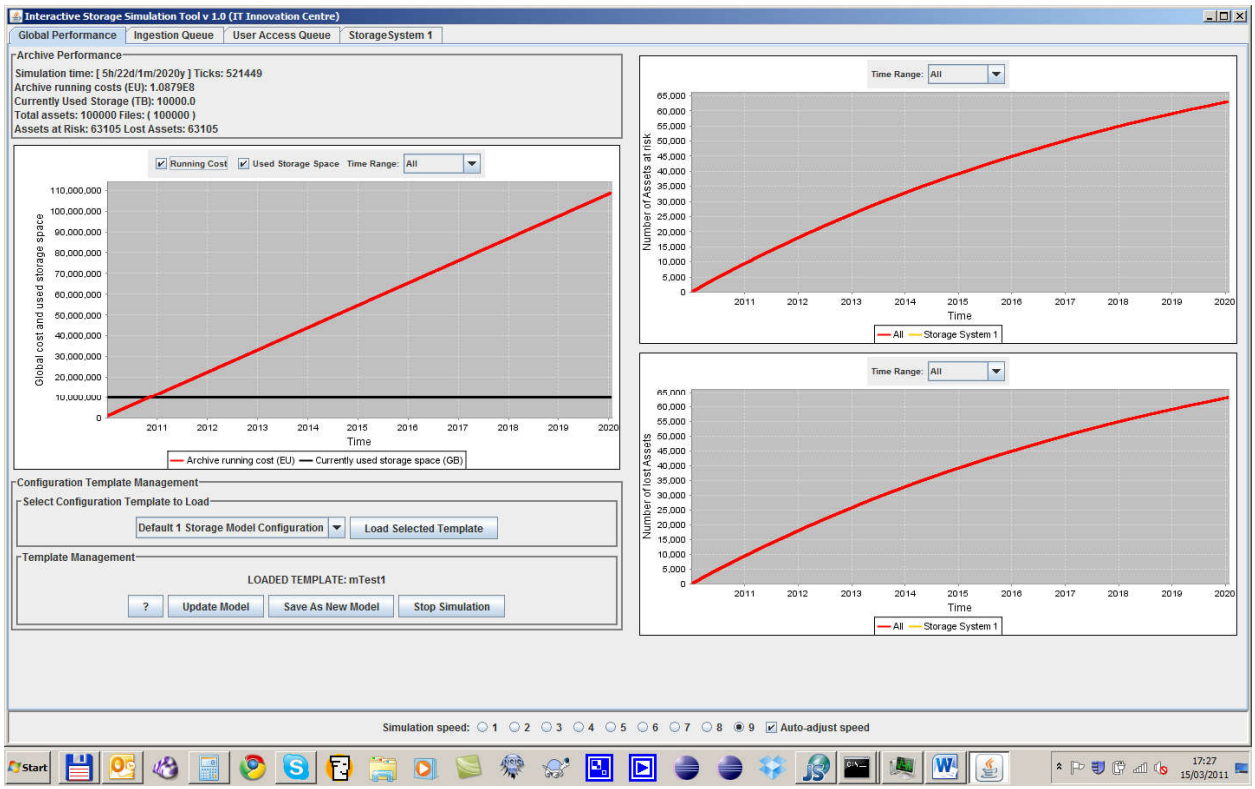


Figure 40 iModel test of single copy model (Test1a)

Test: Two copy model, with/without scrubbing

This test case considers a two-copy model with identical storage systems where a backup copy is made of files on storage system 1 onto system 2 and vice versa.

We want to calculate the probability of loss of a file over time (density function and cumulative distribution) with and without 'scrubbing'.

Suppose that each copy of a file is stored on a storage system where the probability of the copy being corrupted in a given year is p and that this probability is constant for all years of the simulation.

If both copies of the file become corrupted then the file is 'lost' i.e. its contents are irretrievable.

If only one copy of the file is corrupted and then 'scrubbing' takes place, then we return to the state of having two good copies, i.e. corruption of both copies is zero.

A copy may become corrupted more than once. We assume that this can take place, but has no influence on whether a file is lost or not, or whether scrubbing is successful or not, i.e. as soon as both copies of the file are corrupted once then the file is lost, and provided that one copy of a file has no corruption then the other copy can be repaired irrespective of how many corruptions it has suffered.

Suppose that scrubbing takes place annually and at the end of the year. In other words, we start year 1 with both copies with no corruption, we calculate the probability of both being corrupted in the year, and then at the end of the year, if one or none of the copies are corrupted then we start the following year back in the state of neither copy being corrupted.

SCRUBBING ON

p^2	Probability both copies corrupted in same year
$(1-p^2)$	Probability both copies not corrupted in same year
$(1-p^2)^n$	Probability both copies not corrupted in same year for n years in a row (probability file survives to end of year n)
$1-(1-p^2)^n$	Probability both copies corrupted in same year happens once or more during n year period (cumulative probability that file is lost up to and including year n)
$p^2(1-p^2)^{n-1}$	Probability both copies corrupted in year n and not before[25] (probability that file is **first** lost in year n)

The case of no scrubbing, we don't have a repair operation, so probability of corruption accumulates over time.

[25] This is the probability of both copies not being corrupted for the first n-1 years multiplied by the probability that they are then both corrupted in year n

NO SCRUBBING

$(1-(1-p)^n)^2$	Probability that both copies have at least one corruption each after n years (cumulative probability that file is lost upto and including year n)
$1-(1-(1-p)^n)^2$	Probability that there isn't a corruption of both copies after n years (probability that file survives to end year n)
$p(1-p)^{n-1}(2+(1-p)^{n-1}(p-2))$	Probability that both copies aren't corrupted until year n[26] (i.e. probability that file is first lost in year n)

We can plot the density and cumulative functions. Using p = 0.1 as an example:

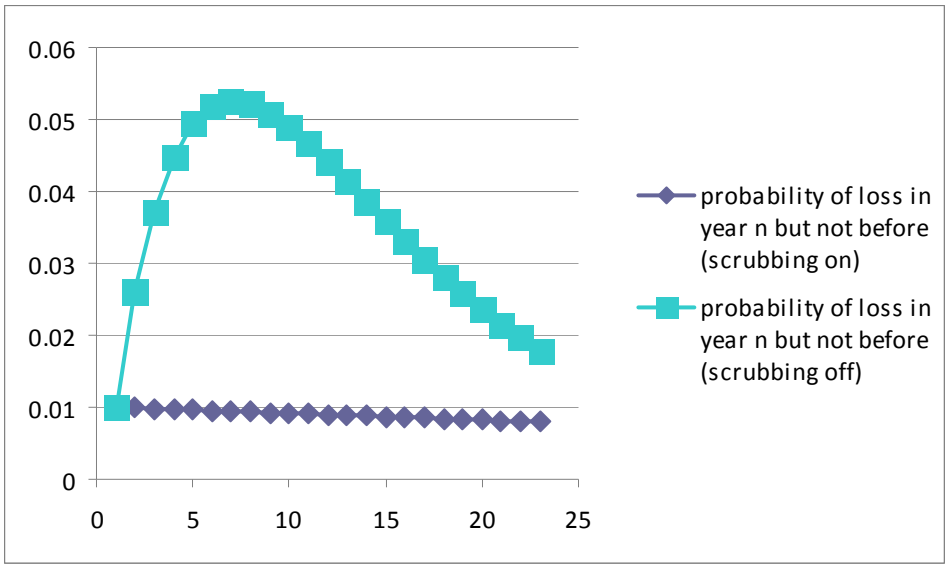


Figure 41 Probability of loss in year n

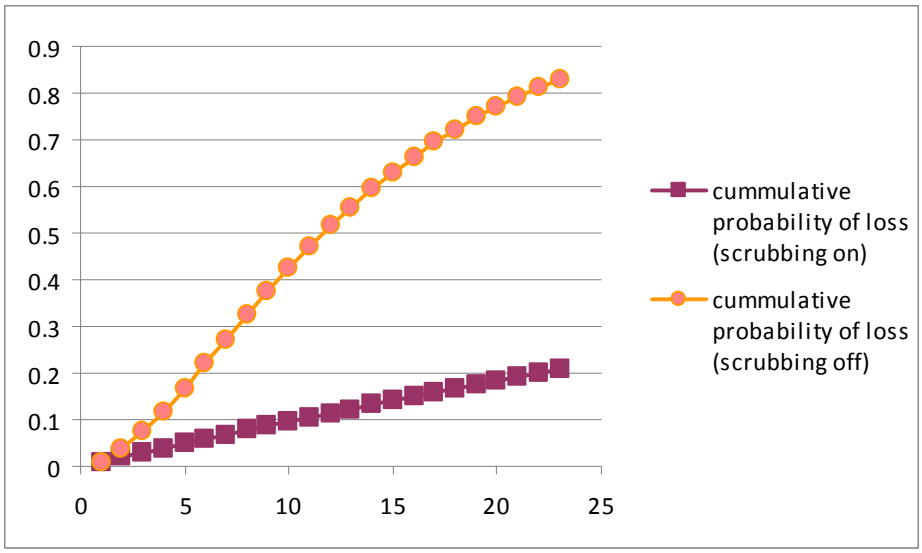


Figure 42 Cumulative probability of loss

[26] This is calculated by subtracting the probability that file is lost upto and inc. end year n-1 from the probability that the file is lost upto and inc. end year n-1, i.e. giving the probability that loss takes place inside year n

This shows clearly the benefits of scrubbing. In the case of no scrubbing, the probability of loss of the file in year n tails off for large n because there is a very good chance that the file has already been lost by this point. The analytic solution above provides a simple test case for a more complex model which should degenerate when parameters are set appropriately.

Test2a: details (scrubbing off)

Test inputs:

File size	100GB
Storage size	10000TB
Latent corruption	10 in 100 files every 12 months
Access corruption	0
Scrubbing	0 (i.e. turned off)
Ingest	none
Access	none
Replication	2 copies of each asset

Expected output:

10,000TB with 100GB files = 100,000 files

Files lost (scrubbing off)

(a) naïve distribution: 42421 files

(b) Poisson distribution: 39958 files

Observed output:

10,000TB with 100GB files = 100,000 assets

Files lost (scrubbing off) 39954

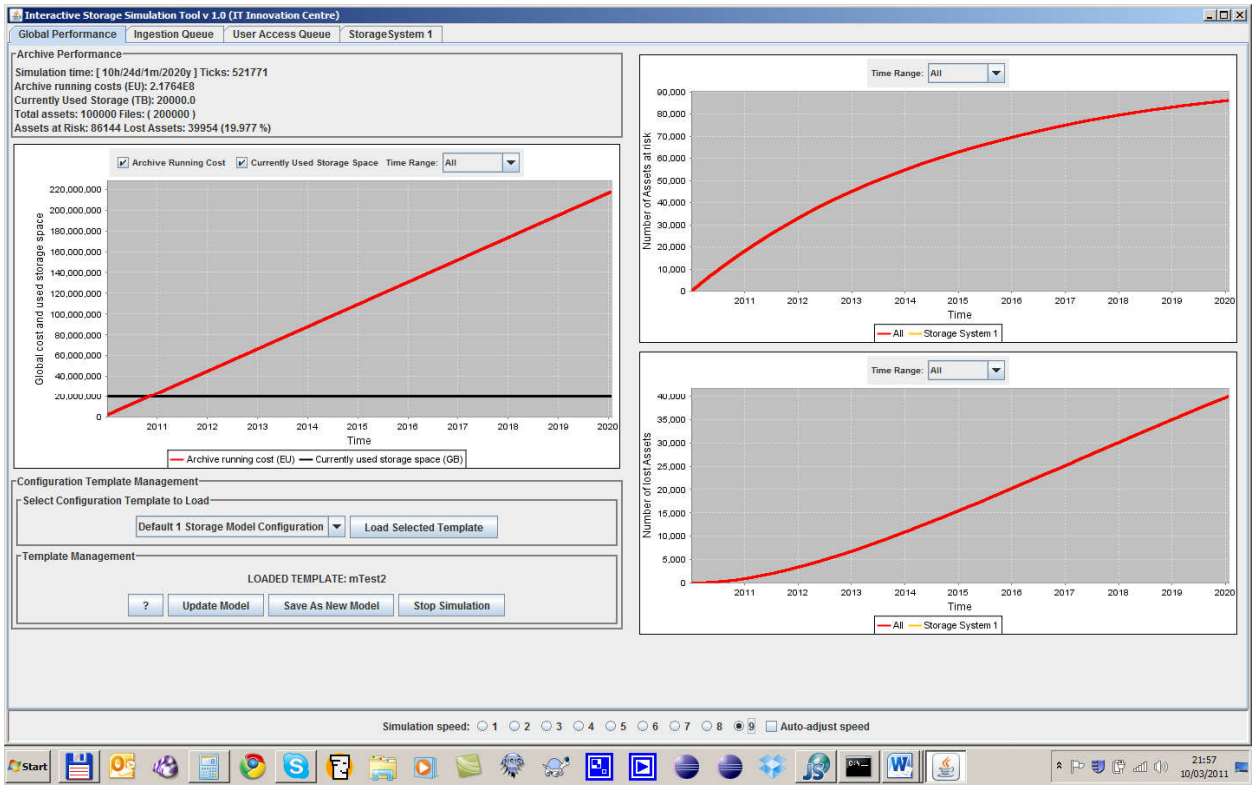


Figure 43 iModel Test 2a (scrubbing off)

The same set of input parameters to the web storage planning tool produces the same result.

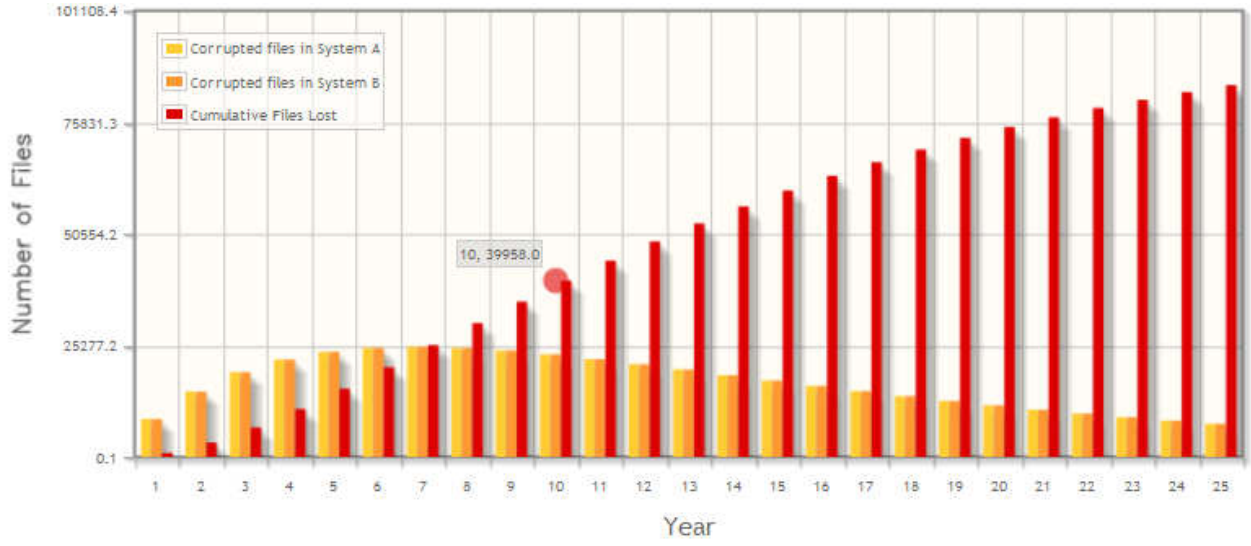


Figure 44 Results for test 2a using web tool. Files lost after 10 years (small red dot highlighting year, files lost) is as expected from Poisson distribution.

Test2(b) details (scrubbing on)

Test inputs:

File size	1GB[27]
Storage size	100TB
Latent corruption	10 in 100 files every 12 months
Access corruption	0
Scrubbing	12 months
Ingest	None
Access	None
Replication	2 copies of each asset

Expected output:

100TB with 1GB files = 100,000 files

Files lost (scrubbing off)

(a) naïve distribution: 9561 files

(b) Poisson distribution: 8695 files

Observed output:

100TB with 1GB files = 100,000 assets

Files lost (scrubbing on) approx. 8643

[27] Note that the file size is lower than in Test2(a)

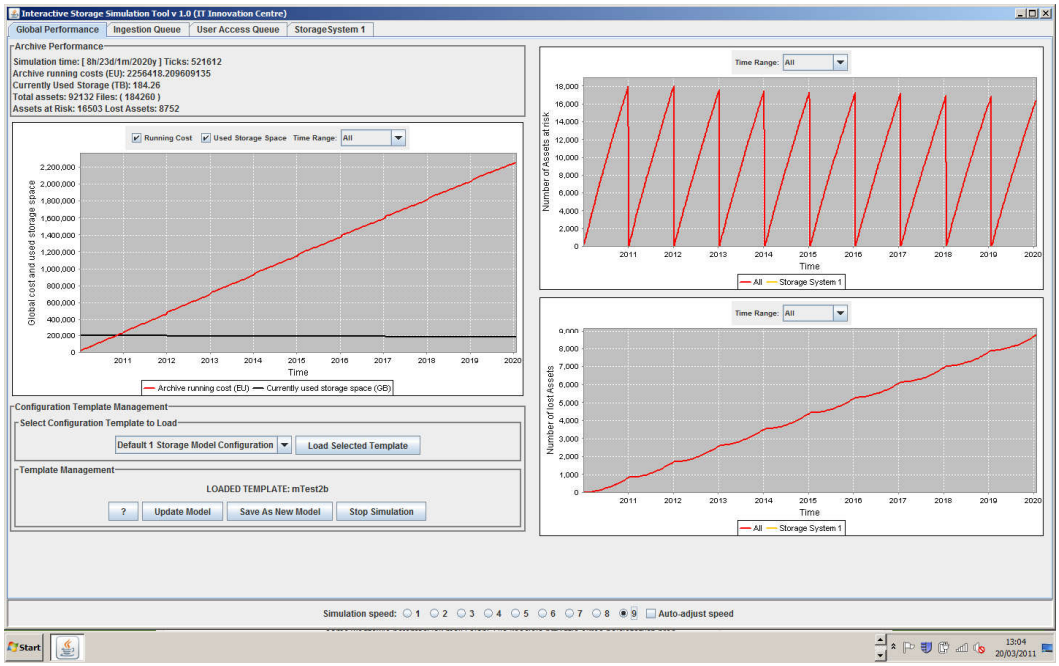


Figure 45 iModel Test2b (scrubbing on)

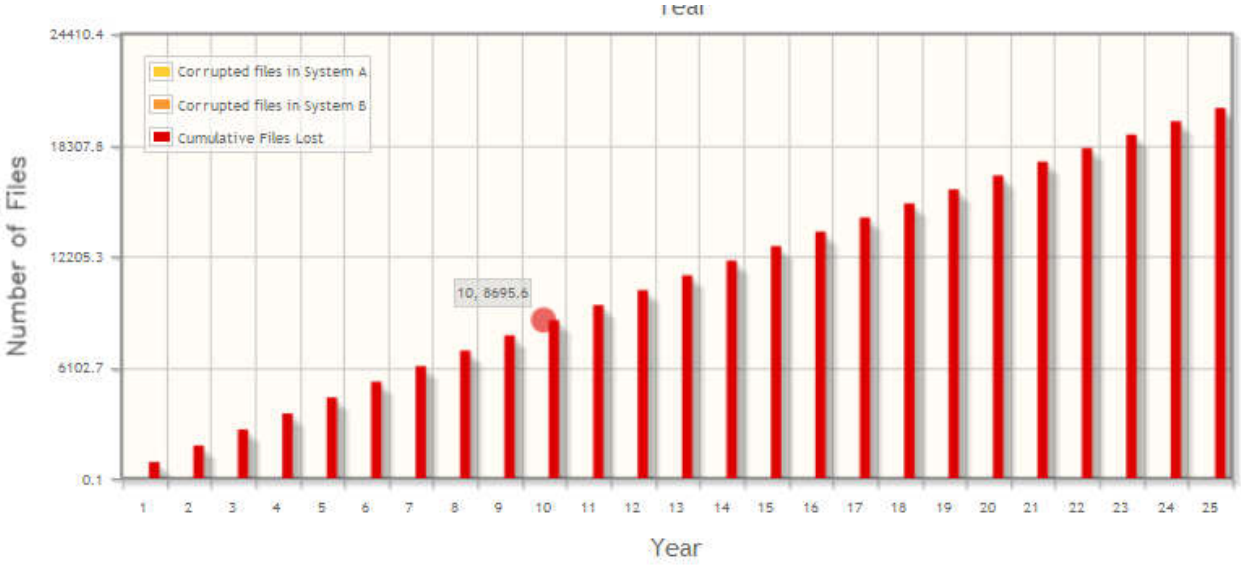


Figure 46 Results from Test2b using the web tool. Files lost after 10 years is as expected from Poisson distribution.

References

[1] Addis, M. et al. (2010). “Threats to Data Integrity from Use of Large-Scale Data Management Environments,” PrestoPRIME Deliverable ID3.2.1, http://www.prestoprime.eu/

[2] Barroso, L. A. and Holze, U. (2009). The Datacenter as a Computer: An introduction to the design of warehouse-scale machines. Google Inc. Synthesis Lectures on Computer Architecture no. 6. Published by Morgan and Claypool.

[3] Moore, R. L., D’Aoust, J., McDonald, R. H. and Minor, D. (2007). Disk and Tape Storage Cost Models. In Archiving 2007

[4] Elerath, J. (2007). Hard Disk Drives: The Good, the Bad and the Ugly!, Queue 5, 6, p28-37. http://doi.acm.org/10.1145/1317394.1317403

[5] Jiang, W. et al. (2008) Are Disks the Dominant Contributor for Storage Failures? A Comprehensive Study of Storage Subsystem Failure Characteristics. FAST '08. http://www.usenix.org/events/fast08/tech/jiang.html

[6] Addis, M. et al. (2010). Audiovisual Preservation Strategies, Data Models and Value-Chains. PrestoPRIME Deliverable D2.2.1 http://www.prestoprime.eu/

[7] Addis, M. et al (2011). “Digital Preservation Strategies: the cost of risk of loss for AV Content”. Jan/Feb 2011 edition of the Motion Imaging Journal of the Society of Motion Picture and Television Engineers (SMPTE).

[8] Rosenthal, D (2011). “Paying for Long-Term Storage”. CNI Fall 2011 Membership Meeting December 12-13, 2011. Arlington, VA. http://www.cni.org/topics/digital-preservation/paying-for-long-term-storage/

[9] Mark Peters (2011). A Comparative TCO Study: VTLs and Physical Tape. With a Focus on Deduplication and LTO-5 Technology February, 2011. Enterprise Strategy Group (ESG). http://www.lto.org/pdf/ESG_WP_LTO_TCO.pdf

[10] David Reine and Mike Kahn (2010) In Search of the Long-Term Archiving Solution — Tape Delivers Significant TCO Advantage over Disk. Dec 23, 2010. Report from the Clipper Group. http://www.lto.org/pdf/2010_December_Archive%20TCO.pdf

[11] Serge J. Goldstein Mark Ratliff (2010). DataSpace: A Funding and Operational Model for Long-Term Preservation and Sharing of Research Data. August 27, 2010. http://dspace.princeton.edu/jspui/bitstream/88435/dsp01w6634361k/1/DataSpaceFundingModel_20100827.pdf

[12] David Cavena et al (2007). “Archiving Movies in a Digital World”. Office of Information Technology, Princeton University. SUN Microsystems report, VERSION 2.1, June 8, 2007; 29pp http://wikis.sun.com/display/SunMediaSpace/2007/11/05/Archiving+Movies+in+a+Digital+WorldAMPAS

[13] “The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials" Academy of Motion Picture Arts & Sciences, 2007; 74pp available from the academy: http://www.oscars.org/contact/council.html

[14] David Cavena et al. (2011). “Archiving Movies in a Digital World”. SUN Microsystems report, VERSION 2.1, June 8, 2007; 29pp http://wikis.sun.com/display/SunMediaSpace/2007/11/05/Archiving+Movies+in+a+Digital+World

[15] David Rosenthal (2011). Progress on the Economic Model of Storage. http://blog.dshr.org/2011/11/progress-on-economic-model-of-storage.html

[16] Eduardo Pinheiro, Wolf-Dietrich Weber (2007). Failure Trends in a Large Disk Drive Population. Google. Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST’07), February 2007

[17] Péter.KELEMEN (2007) Silent Corruptions, CERN IT. LCSC 2007, Linköping, Sweden.

[18] Mary Baker, Mehul Shah, David S. H. Rosenthal, Mema Roussopoulos, Petros Maniatis, TJ Giuli, Prashanth Bungale. (2006) A fresh look at the reliability of long-term digital storage. Proceeding EuroSys '06 Proceedings of the 1st ACM SIGOPS/EuroSys European Conference on Computer Systems 2006

[19] Rosenthal, D. S. H. (2010). Bit preservation; a solved problem? International Journal of Digital Curation 1(5).

[20] Alina Oprea, Ari Juels. (2010) A clean-slate look at disk scrubbing. FAST'10 Proceedings of the 8th USENIX conference on File and storage technologies. USENIX Association Berkeley, CA, USA

[21] Kevin M. Greenan, James S. Plank, and Jay J. Wylie. (2010). Mean time to meaningless: MTTDL, Markov models, and storage system reliability. In Proceedings of the 2nd USENIX conference on Hot topics in storage and file systems (HotStorage'10). USENIX Association, Berkeley, CA, USA, 5-5.

[22] Hakim Weatherspoon and John D. Kubiatowicz (2002). Erasure Coding vs. Replication: A Quantitative Comparison.. Computer Science Division University of California, Berkeley.

Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02). http://oceanstore.cs.berkeley.edu/publications/papers/pdf/erasure_iptps.pdf

[23] Bianca Schroeder, Sotirios Damouras, and Phillipa Gill. (2010). Understanding latent sector errors and how to protect against them. Trans. Storage 6, 3, Article 9 (September 2010), 23 pages.

[24] S. Chen and D. Towsley. (1993)The design and evaluation of RAID 5 and parity striping disk array architectures. Journal of Parallel and Distributed Computing, 17(1-2):58–74,1993.

[25] A. Dholakia, E. Eleftheriou, I. Iliadis, J. Menon, and K. Rao.(2006) Analysis of a new intra-disk redundancy scheme for high-reliability RAID storage systems in the presence of unrecoverable errors. In Proceedings of the joint international conference on Measurement and modelling of computer systems, pages 373–374. ACM New York, NY, USA,2006.

[26] Adam Leventhal. (2009). Triple-Parity RAID and Beyond. Queue 7, 11, Pages 30 (December 2009), 10 pages.

[27] M. W. Storer, K. M. Greenan, E. L. Miller, and K. Voruganti. (2008) Pergamum: Replacing tape with energy efficient, reliable, disk-based archival storage. In Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST), Feb. 2008.

[28] David S.H. Rosenthal. (2010) "LOCKSS: Lots Of Copies Keep Stuff Safe", presented to the NIST Digital Preservation Interoperability Framework Workshop, March 29-31, 2010

[29] Ian Adams and Ethan L. Miller and David S.H. Rosenthal. (2011) Using Storage Class Memory for Archives with DAWN, a Durable Array of Wimpy Nodes. Technical Report. UCSC-SSRC-11-07. University of California, Santa Cruz. Oct, 2011.

[30] Zhang, Y., Rajimwale, A., Arpaci-Dusseau, A. C., Arpaci-Dusseau, R. H. (2010). End-to-end data integrity for file systems: a ZFS case study. In 8th Usenix Conference on File and Storage Technologies.

[31] http://www.theirm.org/publications/documents/Risk_Management_Standard_030820.pdf

[32] CERT: http://www.cert.org/work/organizational_security.html

[33] TRAC: http://www.crl.edu/PDF/trac.pdf

[34] DRAMBORA: http://www.repositoryaudit.eu/

[35] Nestor: http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf

[36] ISO27001 http://www.27001-online.com/

[37] ISO16363 http://public.ccsds.org/publications/archive/652x0m1.pdf

[38] Matthew Addis et al (2010). Threats to data integrity from use of large-scale data management environments PrestoPRIME Deliverable ID3.2.1 http://www.prestoprime.eu/

[39] Addis, M., Wright, R. and Weerakkody, R. (2011) Digital Preservation Strategies: the cost of risk of loss. SMPTE Motion Imaging Journal, 120 (1).

[40] Heydegger, V (2009) Just One Bit in a Million: On the Effects of Data Corruption in Files. Research and Advanced Technology for Digital Libraries, ECDL 2009, LNCS 5714

[41] Addis, M., Wright, R. and Weerakkody, R. (2010) DIGITAL PRESERVATION STRATEGIES FOR AV CONTENT. In: 2010 Conference of the International Broadcasting Convention (IBC 2010), 9-14 September, 2010.

[42] M. Baker. (2011) Preserving Breadcrumbs. Keynote at the 27th IEEE Symposium on Mass Storage Systems and Technologies (MSST), held in Denver, Colorado, in May 2011.

[43] Wright, R; Matthew Addis; Ant Miller (2008) The Significance of Storage in the 'Cost of Risk' of Digital Preservation. Proceedings of iPRES 2008: http://www.bl.uk/ipres2008/ipres2008-proceedings.pdf

[44] http://prestoprime.it-innovation.soton.ac.uk/imodel

[45] iGrafx process simulation and analysis tool http://www.igrafx.de/

[46] Simul8 Simulation Software http://www.simul8.com/

[47] PRISM probabalistic model checker http://www.prismmodelchecker.org/

[48] SimEvents discrete event simulation engine (part of Simulink from Mathworks) http://www.mathworks.co.uk/products/simevents/index.html

[49] The Modelling System Reliability For Digital Preservation: Model Modification and Four-Copy Model Study. Yan Han, Chi Pak Chan The University of Arizona Libraries. iPRES 2008. http://www.bl.uk/ipres2008/presentations_day2/44_Han.pdf

[50] Constantopoulos, P., Doerr, M., and Petraki, M. 2005. Reliability modelling for long term digital preservation. http://delos-wp5.ukoln.ac.uk/forums/dig-rep-workshop/constantopoulos-1.pdf

[51] M-Disc optical storage from Millenniata. http://millenniata.com/

[52] Cinevator Keeper product for storing digital data on film stock. http://cinevation.net/portfolio/cinevator/cinevator-keeper/

[53] Lakshmi N. Bairavasundaram, Garth R. Goodson, Bianca Schroeder, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau. An Analysis of Data Corruption in the Storage Stack Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST '08)

[54] Lakshmi N. Bairavasundaram, Garth R. Goodson, Shankar Pasupathy, Jiri Schindler. An Analysis of Latent Sector Errors in Disk Drives Proceedings of the International Conference on Measurements and Modelling of Computer Systems (SIGMETRICS'07)

[55] Haryadi S. Gunawi, Cindy Rubio-González, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, Ben Liblit. EIO: Error Handling is Occasionally Correct Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST '08)