

Managing the Value at Risk in Digital Asset Collections

Michael F. Bellacosa, MIA, MLS
Principal, Bellacosa Risk Solutions

I would like to thank Ann Green and Louis King of Yale University's Office of Digital Assets and Infrastructure who were helpful to me in articulating the major issues, pointing me to some of the initial sources and providing me with useful editorial comments. Meg Bellinger, Director of ODAI, also gave me helpful and encouraging feedback.

Introduction

The preservation of digital assets is fundamentally a risk management problem. In financial markets, risk management is the practice of identifying risks to the value of financial assets, assessing those risks and then devising and implementing the most efficient strategies for reducing those risks to tolerable levels; the economic trade-off from the risk reduction is the cost of the strategy for hedging the risks. When considering digital asset preservation, one can simply replace "portfolio of financial assets" [e.g. stocks, bonds, derivatives, etc.] with "collection of digital assets" [e.g. text, images, sound, data, etc.] and begin to look at the analogies. Indeed, many concepts from financial risk management are directly or analogously applicable for developing and describing a process for preserving the value of vulnerable digital assets. Of course, there are important differences between the two worlds.

Unlike a stock or bond, the value of digital assets is not [usually] determined transparently throughout the day by open market forces; rather, their value will be based on subjective judgments of their potential uses over long periods of time. This is relatively more difficult for academic/cultural heritage assets compared to intellectual properties in the commercial space. In addition, while some of the risk dimensions between the two worlds can overlap [e.g. legal or regulatory risks], quantifying the probabilities of occurrence of "bad" events and the severity of their impacts can be quite challenging in the case of "intangible" digital assets. Finally, strategies for hedging the value of the digital assets may involve complex technology-oriented decisions envisioning relatively long time horizons which can themselves encompass the evolution of the technologies that mitigate some risks and create others.

A great deal of research and development has been conducted on risk assessment and risk management related to digital preservation. However, this work has been mainly from the perspective of digital repositories seeking "Trustworthy Digital Repository" status through an audit process [e.g. Portico, HathiTrust]. Although audit processes such as DRAMBORA and TRAC are useful for assessing particular repositories as digital preservation options, this paper will examine the risk management of digital assets from the perspective of the stewards of the assets whose value is at risk. These content stewards are analogous to portfolio managers hired

by investors to manage their finances; here, these content stewards are responsible for risk management processes which lead to strategic decision-making and the taking of proactive steps to bring collections of digital assets to a tolerable level of risk on behalf of the content owners.

The paper will first examine each of the three legs of the risk management stool: value, vulnerability, and cost. Then, a model will be introduced which will show how a graphic display of value and vulnerability reveals the risk profile of a digital asset collection to content stewards, content owners and preservation funders. [Note on terminology: stewards and owners can sometimes overlap as can owners and funders, but stewards and funders rarely overlap since, in that case, digital preservation business cases would be self-evident and always funded.] This model can be easily calibrated to an articulated risk tolerance. Assuming that the initial risk profile for the assets under examination does not naturally fall within the stated risk tolerance, hedging strategies can be incorporated to associate the economic cost of certain risk management activities with the benefit of risk reduction. Content stewards, content owners, and preservation funders will be able to see the cost of creating a desired risk profile or, alternatively, the risk profile that must be tolerated at a desired cost level. Finally, a cyclical risk management decision-making process for using the model interactively with content stewards will be outlined which will facilitate periodic assessments of the cost-effectiveness of the risk management strategies implemented. It will also provide the opportunity over time to modify the strategy in light of evolving value and vulnerability assessments as well as newly available technologies.

Assessing and Articulating the Value of Digital Assets

To paraphrase one of the salient points from the Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2008): the value of digital assets is fundamentally about their future use to generate streams of benefits and digital preservation is the process that makes those benefit streams possible. It may seem crass but, in making a business case, digital assets cannot be said to have purely intrinsic value; rather, they must imply potential future streams of benefits that can be persuasively articulated as aligned with and supportive of the strategic objectives of the funders of digital preservation programs.

In the corporate world, investors contribute capital to a business which is used to buy and build assets and infrastructure, hire and train staff, manage production processes [including managing risk to ensure product quality and delivery] and market and sell its products to potential customers. This produces a stream of future cashflows which is paid to those shareholders as cash dividends. Analogously, in the world of digital assets, the people, departments and organizations that support digital asset preservation programs [i.e. the “funders”] play the role of investors in a preservation infrastructure and process [including managing risk to ensure product quality and delivery] which produces a stream of use-generated outcomes of value to the funders as well as other stakeholders.

To push the analogy a bit further: the “product” in the digital assets case is access to content stored in electronic form and this content must be authentic, understandable and, above all, usable. A “customer” uses that access to that content [for a fee or not] in a way which generates value for the preservation program’s “investors” [which may or may not include direct revenue generation]. These periodic value-generating outcomes are analogous to the periodic

dividend payments received by the holders of corporate stock. Just as investors can value a stock based [in this simplified corporate finance example] on the likelihood of receiving a stream of dividend payments, funders of digital asset preservation programs can value the program and the underlying assets based on the potential benefit streams that they could produce. Since many of these benefits will be non-financial, they need to be understood and articulated in the context of and in alignment with the funder's strategic objectives. Digital assets which are seen to be the source of benefit streams most closely aligned with and supportive of a funder's strategic objectives will be assessed to be the most valuable; they will demand the most rigorous [and possibly most expensive] stewardship and risk management.

In the final report Keeping Research Data Safe 2 [KRDS2], Beagrie, Lavoie and Woolard (2010) presented a "benefits taxonomy" as a way of organizing the value-generating outcomes flowing from digital assets and their preservation. Following the corporate finance analogy from above, these are the "dividend payments" that the investment in preservation generates. The taxonomy is divided into three two-part categories: direct benefits vs indirect; near-term benefits vs long-term; and private benefits vs public.

The report lists several specific examples of each, for example: the direct benefit of increasing scholarly communications/stimulating new collaborations and the indirect benefit of re-purposing data for new audiences; the near-term benefit of the availability of data underpinning journal articles and the long-term benefit of adding value over time as a collection grows to critical mass; and, the private benefit of reputational value to the sponsor/funder of an archive and the public benefit of motivating new research. Collections of digital assets can be evaluated and their value-generating outcomes organized using this framework with reference to the several examples of each; other categories and examples may surface during the assessment process. The output of a structured discussion among stewards, owners and funders is an organized list of the various value statements associated with the given collection of assets.

The JISC-funded *espida* project (2007) presented a framework for articulating the benefits from a proposed digital asset preservation program in alignment with a funder's strategic objectives. Consistent with the corporate finance analogy above, the *espida* approach frames the business case discussion as a dialogue between content stewards and funders about the cost, risk and potential return of a digital preservation investment opportunity. Six crucial questions from potential funders frame the dialogue:

- How much do you want [money or other resources]?
- What do I get for it?
- How will I know that I've got it?
- How likely am I to get it?
- What determines success or failure?
- How will you manage the program for success?

The *espida* approach adapts a well-known strategic management tool called the Balanced Score Card [BSC] introduced by Kaplan and Norton (1992) at Harvard Business School which enables the viewing of an organization's strategic objectives through four perspectives [one financial and three non-financial]. In the *espida* approach, these four perspectives provide useful

guidance for content stewards proposing digital preservation programs as they seek to articulate their value propositions to potential funders in ways that are aligned with and supportive of the latter's strategic objectives. [It should be noted that it would be rare for any given project proposal to offer significant contributions to an organization's strategic objectives along all four of these dimensions.]

As interpreted by the key *espida* project staffers (Currall and McKinney 2006), the BSC's four perspectives are: 1) customer and external stakeholders; 2) internal business process; 3) innovation and development; and, 4) financial. The first perspective covers outcomes that are valuable to the organization's beneficiaries such as the delivery of high-quality products or services which results in highly satisfactory customer experiences. The second one covers outcomes that are valuable to the organization's operations in the sense of increasing efficiency, productivity, flexibility and effectiveness. The third covers outcomes that are valuable to the organization's human and intangible assets such as the enhancement of employee skills and the creation and growth of intellectual capital. The fourth perspective is about positive economic value from either revenue generation or cost savings. The usefulness of the BSC framework in articulating the strategy-aligned value of intangible digital assets and their preservation lies in the relatively lower proportion of analytical space focused on purely monetary considerations. Moreover, the BSC is an internationally recognized tool for analyzing, designing and articulating an organization's strategic plan which may well be familiar to members of a funding organization's senior management team.

The KRDS2 benefits taxonomy can be used to organize the set of value statements drawn from a typical digital preservation proposal and the BSC approach can be used to express them through the four perspectives in line with the funding entity's strategic objectives. For example, the customer/stakeholder perspective can point to the basic public benefit of preserving the cultural heritage as well as the long-term benefit of the increased use of the materials due to better and easier access and the discovery/creation of new audiences via new technological outlets. The internal business process perspective can illuminate the near-term and long-term benefit of maintaining access to materials that otherwise might require re-creation or substitution of lost originals; a related indirect benefit would be the efficiency and productivity gains from simplifying the re-purposing and re-use of materials. The innovation perspective can point to the direct benefit of encouraging multi-media literacy as well as the private benefit [to the funder's reputation] of being seen as a significant contributor to the state-of-the-art of digital preservation practice. Finally, the financial perspective can include the direct benefit of the possibilities for marketing of various digital assets resulting in payments to copyright holders as well as reducing the operating costs for satisfying increasing demand for such content.

In developing these value statements, it can also be helpful to refer to eight types of intangible value set forth by Wired editor Kevin Kelly (2008) in a blog posting which are so valuable that people are willing to pay for them even if they could find and get the content for free: immediacy [timely access]; personalization [customizable delivery]; interpretation [guidance and support]; authenticity [reliability of content]; accessibility [worry-free availability]; embodiment [high-quality content experience]; patronage [customer loyalty]; and findability [amongst the mass of free content]. These can underpin the generation of specific value statements about particular assets and collections.

In an idealized value assessment process [preparatory to the value/vulnerability analysis that feeds into the final risk management argument and proposal], the BSC tool would first be used by content stewards and owners to understand the funder's strategic objectives. They would use the benefits taxonomy to generate an organized set of value-generating outcomes that could result from the proposed project. Then, the BSC tool would be used as a framework for matching up the set of value propositions to the funder's strategic objectives. The relative valuations of digital assets or collections to be preserved would be based on the degree to which their associated potential benefit streams are in close and clear conceptual and linguistic alignment with the strategic objectives of the funding entity. To avoid pure self-assessment by the digital asset collection's stewards or owners, a dialogue with the funders would be the mechanism for assessing how relatively valuable the assets are relative to the funder's strategic objectives. In the end, it is up to the content stewards and owners to make their case for value.

Risk and Vulnerability of Digital Assets

By building on analogies from risk management in financial markets, one can lay out a conceptual framework for the risk and vulnerability of digital assets. In financial markets, one of the most important kinds of risk is credit risk: the risk that an event may occur with some probability [e.g. a borrower defaults on a loan] resulting in an impact whose severity could range from a total loss of value to a full recovery. Indeed, the vulnerability of digital assets and collections can be defined as the probability of a "bad" event occurring multiplied by the severity of the impact of the event. [Proximity in time should be part of this formula; but modeling the evolution over time of a digital asset's value and risk is beyond the scope of this paper.]

A fully developed risk management process begins with establishing the strategic context of the organization's risk environment followed by identifying the specific risks, analyzing them, assessing their severity of impact, and managing them. Drawing from a menu of techniques, risks and their potential impacts may be avoided, reduced, transferred, deferred, or retained. Finally, any risk management process needs to include ongoing and iterative review, assessment, and communication with and among the content stewards and owners as well as the funders of the program. JISC (2009) has published a useful guide on the risk management process.

In 2009, Barateiro, Antunes and Borbinha introduced a "taxonomy of vulnerabilities and threats to digital preservation" which can be used in the identification phase of the risk management process for digital assets. A graphic chart lays out seven dimensions of risk with examples in the accompanying text of the type of "bad" events that correspond to each one, as follows: 1) data may be exposed to media faults through partial or total failure of storage media or through "bit rot"; or, obsolescence of format may make the bit stream unrenderable even if the zeros and ones are preserved; 2) infrastructure may be exposed to temporary or permanent hardware problems, hardware obsolescence, or problems with communications or network services; 3) software-dependent processes may be exposed to software bugs or obsolescence; 4) natural disasters and unintentional human error can cause data loss due to hardware damage, software damage/alteration, or from accidental deletion of files; 5) deliberate attacks on system components or data can originate outside the organization such as from malicious hacking, or from inside such as from a disgruntled employee; 6) management can fail to maintain the

organization's viability either following unforeseen shocks or because of incompetence; and, 7) legal and regulatory compliance can cause problems for the use of the otherwise well-preserved digital assets. One additional risk dimension not included in the taxonomy relates to metadata: the potential inability to find the digital assets due to inadequate cataloging processes; the potential inability to understand [and thus use] the digital assets due to a lack of contextual, descriptive metadata; and the potential failure of ongoing preservation processes due to inadequate administrative metadata.

Although assigning quantitative probabilities to many of these risks is clearly impossible, estimating the probabilities and severities for some "bad" events can be based on quantitative research. For example, some studies have been published related to storage media longevity, failure rates and data loss severities. Schroeder and Gibson (2007) examined the annual disk replacement rates in a population of 100,000 drives from four vendors. Assuming that replacement rates are a fair proxy for failure rates, they found that the observed failure rates were far higher than those derived from by the "Mean Time To Failure [MTTF]" data published by vendors. A similar study done by Google (Pinheiro, Weber, and Barroso 2007) found broadly consistent results. However, Schroeder and Gibson also noted the possibility "that disk-independent factors, such as operating conditions, usage and environmental factors, affect [disk] replacement rates more than component specific factors": i.e. that factors other than the failure of the storage media itself could be the cause of the system failure and replacement. Jiang et al. (2008) looked at 1.8 million drives in 39,000 systems and also found that factors other than disk failure were significant causes of storage system failure: indeed, one-third to as many as two-thirds of those failures were due to physical interconnect failures (e.g. cabling) . Finally, based on a personal communication from a colleague at the San Diego Supercomputer Center, Rosenthal (2010) noted, that "over 20 years at the SDSC, operator error is said to have been the cause of three-quarters of all data-loss incidents."

Equally as important as probabilities of failure of storage systems is the severity of the impact of such failures. The first problem comes from so-called latent disk errors, more commonly known as "bit rot." This process, which according to Baker et al. (2006) occurs significantly more frequently than full disk failure, introduces corruption into files silently and invisibly such that the damage is only detected when the file is to be read. A CERN study on data corruption (Panzer-Steindel 2007) found in an 8.7 terabyte [TB] dataset that 1 byte in 30 MB was irreversibly corrupted within 6 months with 1 out of 1,500 files affected. The second problem is the degree to which a partially damaged file can be opened and used. Although the ability to open an uncompressed file appears to be fairly resistant to minor disk errors, the same cannot be said for compressed files. The CERN study found that a single bit error would make a compressed file unreadable with 99.8% probability. According to Wright, Miller and Addis (2009), an uncompressed .WAV file with .4% errors will exhibit barely noticeable differences from the undamaged original whereas a similarly damaged MP3 file will not even open. Heydegger (2008) found that a .01% byte error rate in a compressed JPEG2000 file resulted in at least a 50% loss of information; moreover, a single byte error in a compressed JPEG2000 file could produce obvious damage throughout the image.

In one sense, these quantitative results are important and valuable for discussing and assessing the relative riskiness of different storage strategies. However, in another sense, they

also point to the need for further research before more concrete understanding of the trade-offs can be achieved. For example, it may seem that hard disk drives are riskier than vendor MTTF data suggest. But what if a major contributor to failure of disk-based systems is human error or some other problem with the overall physical plant? It seems that compressed files are riskier than uncompressed. But what if using much larger uncompressed files puts an economic constraint on the number of copies saved? Is the increased risk that a file couldn't be opened offset by having more copies or will the copies fail in some correlated manner? Moore's Law on capacity is driving down storage costs. But is exponential growth of capacity without corresponding advances in reliability an unalloyed benefit? If, in 15 years, we could store a 1 petabyte [PB] research collection on just one disk [instead of the thousand needed today], would we really want to put all those eggs in the one basket even with a back-up copy?

Since the risk profile of a digital asset collection will be based on a mixture of quantitative and qualitative estimates, the common denominator must be qualitative [supported by quantitative data wherever possible]. The JISC Risk Management Toolkit (2009) and the DRAMBORA toolkit (McHugh et al. 2007) provide simple and intuitive schema for probabilities and severities. If one excludes DRAMBORA's "minimal/once in a hundred years" probability category and its "zero loss" severity category, the JISC and DRAMBORA schema can easily be synthesized into one.

This paper's value/vulnerability model incorporates five categories of increasing probability [using the JISC terminology]:

- very low: unlikely to occur;
- low: may occur occasionally;
- medium: as likely to occur as not;
- high: likely to occur; and
- very high: almost certain to occur.

There are six categories of increasing severity [using the DRAMBORA terminology]:

- negligible: isolated but fully recoverable loss of the digital objects' authenticity and understandability;
- superficial: widespread but fully recoverable loss;
- medium: total but fully recoverable loss;
- high: isolated loss including unrecoverable loss;
- considerable: widespread loss including unrecoverable loss or loss recoverable only by a third party; and
- cataclysmic: total and unrecoverable loss of the digital objects' authenticity and understandability.

The vulnerability matrix presented in the model introduced below will be based on these two scaled axes. But first: a word on digital preservation costs.

The Cost of Hedging the Risk in Digital Asset Collections

Estimating the cost of different digital preservation risk management strategies is a fairly difficult process. At first glance, one might think that the costs are mainly about robust storage systems whose costs per TB appear to be falling fairly rapidly. However, both KRDS2 (Beagrie, Lavoie, and Woolard 2010) and the three LIFE initiatives (McLeod, Wheatley and Ayris 2006; Ayris et al. 2008; Wheatley and Hole 2009) illustrate the relatively lower cost contribution from storage per se as compared to the other activities in the digital preservation lifecycle. According to the LIFE projects, these activities include acquisition of the content, ingest to the repository, bit-stream preservation, content preservation and access. Sub-activities within these major categories include intellectual property rights management, physical/electronic storage, metadata creation, technology monitoring and user support. In fact, metadata creation and management by itself is a major cost component. Costing out these lifecycle components is an extremely complex undertaking for all but the smallest and most homogeneous collections.

However, the cost of storage is one activity for which some solid empirical data exist. In a 2007 paper from the San Diego Supercomputer Center, Moore et al. analyzed the all-in cost of data storage from an operational perspective. Beyond just the cost of the storage media, this also included supporting servers and infrastructure, hardware maintenance, software licenses, floor space, utilities and labor. At that time, the SDSC had been running a 24/7 data center for about 20 years and managed 2,500 TB of disk storage as well as 25 PB of tape capacity. Their bottom line for the cost of managed storage: \$1,500 per TB per year for hard disk drives and \$500 per TB per year for tape.

In February 2010, the PrestoPrime project released deliverable 2.1.1 (Addis and Wright 2010) one section of which built on the SDSC's research [as well as those of others including Google (Barroso and Holze 2009)] to present a fairly detailed and current analysis of the storage costs for digital audio-video. Their bottom line cost for managed large-scale storage [e.g. over 500TB]: \$1,000 per TB per year for hard disk drives and about a third of that for tapes. The cost for large-scale storage from Amazon's S3 cloud storage service was seen to be consistent with this estimate. The report also found a multiple of approximately 5 times between the cost of raw media and the cost of managed storage. The trend of prices charged for storage by the SDSC and Amazon's S3 supports a 2-3 year assumption for the half-life of storage cost: for large-scale disk storage, SDSC charged \$1500 per TB per year in 2007, \$650 in 2010 (SDSC 2010); Amazon's S3 was \$1800 in 2007, \$1000 in 2010 (Amazon Web Services 2010).

Princeton University's Office of Information Technology (Goldstein 2010) used the basic mathematics of a convergent series to estimate the total costs in perpetuity for storage based on an assumed halving-period for these costs. For example, with a cost-halving period of two years, a Year 1 cost of \$100 approaches \$341 in cumulative costs at infinity; a three year halving period approaches \$484. Thus, an assumption of a 2-3 year halving period implies an approximate 4X multiple of current annual costs for the lifetime cumulative total. [Note: this simple extrapolation necessarily ignores a whole range of variables including, among other things, inflation, frequency and speed of access required for the collection, and disruptive technology development; future cashflows have also not been discounted to present value.]

Storage costs provide an interesting example of how part of the cost analysis could be performed. The larger challenge for a full risk management strategy cost analysis is to account for all the specific cost components related to the specific digital asset collection needing preservation. It must also correctly account for incremental labor costs as existing staff is tasked with new duties [such as metadata creation and management] as well as any sunken costs from pre-project investments in plant and equipment. The ongoing LIFE projects and the KRDS2 study are both valuable contributions to this complex topic.

Value and Vulnerability: The V-V Model™

The discussions above on value and vulnerability form the basis and create the context for two separate assessments. These assessments are used to create a model consisting of a small set of “vulnerability matrices” calibrated to a specified risk tolerance which can display a collection’s initial and desired risk profiles.

First, the relative value of one or more digital asset collections is assessed, articulated and assigned to one of three ranked value levels based on how much the assets contribute to the successful achievement of the funder’s strategic objectives: at the highest value level are assets which are critical for such success; at the middle level are assets which clearly contribute to that success but may not be critical to it; and, at the lowest level are assets whose contribution to that success is unclear.

Second, the vulnerability of these assets is assessed in terms of the probabilities of “bad” events occurring and the severity of their impacts based on the JISC and DRAMBORA rankings described above. A generic “vulnerability matrix” is a 5X6 matrix with the probability scale on the left, the severity scale across the top and the vulnerability rankings in the body of the matrix that result from multiplying the two axes [see Figure 1 below]. The probability rankings on the left side range from (1) very low/unlikely up to (5) very high/almost certain. The severity rankings across the top range from (1) negligible/an isolated but fully recoverable loss of authenticity and understandability up to (32) a total, unrecoverable loss. Note that the severity

		Severity of Loss					
		1	2	4	8	16	32
Probability	5	5	10	20	40	80	160
	4	4	8	16	32	64	128
	3	3	6	12	24	48	96
	2	2	4	8	16	32	64
	1	1	2	4	8	16	32

Figure 1: generic matrix showing anchor points

scales up exponentially since, in the digital preservation space, the more severe impacts are significantly worse than the lesser ones. This produces the intuitive result that high loss severities

with low probabilities are worse [i.e. have higher vulnerability rankings] than low loss severities with high probabilities.

To orient oneself, one can look at two anchor points in the generic matrix. The green cell in the lower left corner has the lowest vulnerability rank corresponding to an unlikely, yet possible, chance of an isolated, fully recoverable loss; this would presumably be an acceptable vulnerability state for even the highest value assets. In contrast, the red cell in the upper right corner has the highest vulnerability rank corresponding to the near certainty of a total unrecoverable loss; this would presumably be unacceptable for any assets [except perhaps the lowest value ones in an aggressive risk tolerance context].

The next step is to use the generic matrix to generate a specific vulnerability matrix for each of the three asset value levels. Each of the three asset value levels requires its own, specific vulnerability matrix because assets of differing value imply different risk tolerances [i.e. generally speaking, one would accept higher vulnerability for lower value assets]. This is done by multiplying the set of vulnerability rankings in the generic matrix by a simple weighting scheme: for the lowest level the multiplier is 1X; for the middle level it is 3X; and for the highest level it is 5X. Therefore, the lowest value level matrix is the same as the generic one; Figure 2 below shows the matrix for the highest asset value level.

		Severity of Loss						
		1	2	4	8	16	32	
Probability	highest value	5	25	50	100	200	400	800
	5	25	50	100	200	400	800	
	4	20	40	80	160	320	640	
	3	15	30	60	120	240	480	
	2	10	20	40	80	160	320	
	1	5	10	20	40	80	160	

Figure 2: highest value matrix showing 5X multiple on rankings

Next, the three matrices are calibrated to reflect the risk tolerance of the content owners as well as possibly other stakeholders. Elaborate and customized calibrations can be performed. However, a fairly simple calibration can be based on the answers to just two questions about one's highest value assets: 1) what is the worst unlikely, yet possible, severity of loss you can accept? 2) what is the worst near certain severity of loss you can accept? Suppose, for example, that for one's highest value assets one could accept the unlikely, yet possible, chance of a total but fully recoverable loss and the near certainty of an isolated but fully recoverable loss. Referring to figure 2 above, the first answer would imply that the cell in the bottom row and the third column [with a vulnerability rank of 20] is acceptable as would be the cell in the top row and the first column [with a rank of 25]. Therefore, all cells in the three matrices ranked up to and including 20 are green [clearly acceptable], those from 21-25 are yellow [debatable] and those over 25 are red [clearly unacceptable]. Figure 3 on the next page shows the model after calibration. Note that going from the highest asset value matrix to the lowest, the size of the red zone shrinks as the green zone grows which reflects the intuitive sense that relatively greater vulnerability is acceptable for relatively lower value assets. [For contrast, Appendix 1 shows the model calibrated for a more conservative and a more aggressive risk tolerance].

After calibration, the red, yellow, and green zones of the model reflect the articulated risk tolerance for the three asset value levels. The next step is to take the probability and severity estimates for the various risk elements associated with a particular collection and populate the appropriate matrix. As a simple illustration, say that Collection A is exposed to ten risk elements

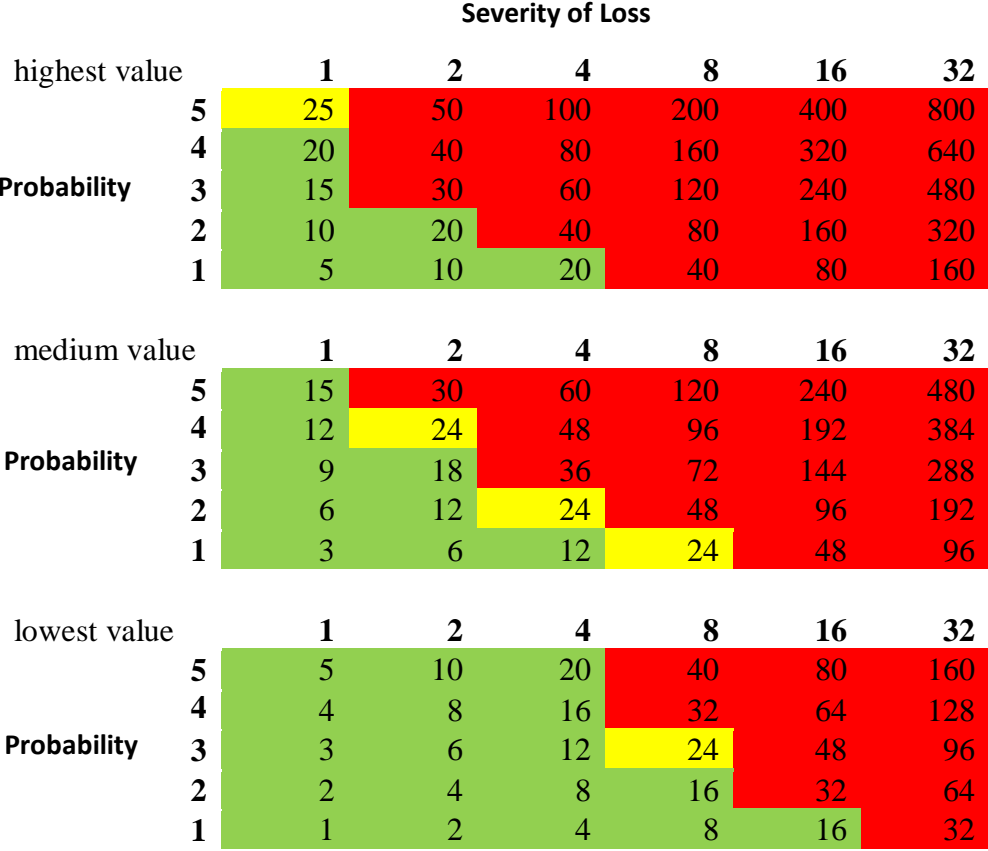


Figure 3: model calibrated for moderate risk tolerance

[in practice, there would likely be more]. For each one, estimate the probability of occurrence and severity of impact on Collection A. Next, plot the results onto the vulnerability matrix appropriate to Collection A’s value level. This will display the initial value/vulnerability profile of Collection A as a 3-D histogram as shown in Figure 4a on the next page [each block in a histogram bar represents one risk element in its probability/severity grid location].

Those risks in the green zone are fine as they are; those in the yellow zone call for discussion; and those in the red zone call for action. Collections of the highest value assets with a substantial number of red-zone risks should be a cause for some alarm. Collections of the lowest value assets can be dealt with at a lower priority; however, red-zone risk concentrations are still outside the specified risk tolerance. For a summary level view of the results of a multi-collection

analysis, vulnerability rankings can be displayed as a function of value in order to set priorities for remedial action.

The goal of the entire process is the transformation of collections with a number of red-zone vulnerabilities into ones with little, or ideally no, risk in the red zone. A collection's risks can be moved out of the red zones by reducing their associated probabilities or severities by implementing risk management strategies. Figure 4b below is an example of an improved risk profile alongside its initial profile which shows the effect of selected risk management strategies on the collection's vulnerability state.

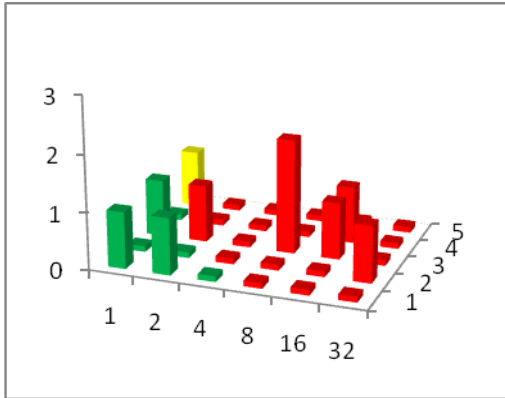


Figure 4a: initial vulnerability state

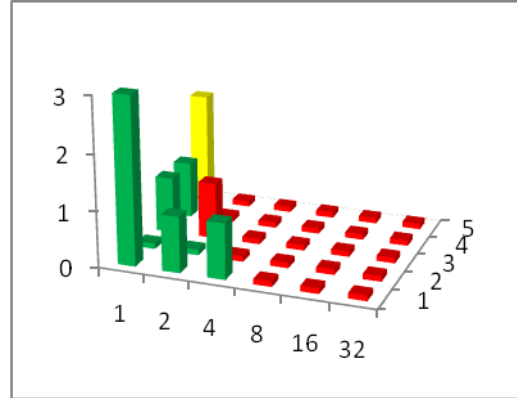


Figure 4b: risk-managed vulnerability state

The model can be used to project which combinations of probability and severity reduction are needed to bring a value/vulnerability profile into alignment with a specified risk tolerance; this is accomplished by implementing risk management strategies. The model can show the impact of a variety of strategies each with its own associated costs. This model can be used to show the initial vulnerability profile of an asset collection and how well or poorly that profile fits within a stated risk tolerance. It can then be used to show the variety of improved risk profiles that result from implementing different risk management strategies. Content stewards, owners and funders can see the cost of achieving a tolerable risk profile; alternatively, they can see the risk profile [tolerable or not] that results from a certain level of investment in digital preservation. In either case, by incorporating the assessments of value and vulnerability as well as the estimated costs of different risk management strategies, the model becomes a structured cost-benefit analysis tool used to support the decisions that must be made to manage the value at risk in digital asset collections.

An Idealized Risk Management Process

The process begins with establishing an organization's strategic context and risk environment. What are the organization's strategic objectives and how does the given digital asset collection support their achievement? What risks threaten the ability of the organization to achieve its objectives and what risks to the given digital asset collection threaten to undermine its contribution to the organization's strategic success?

Assuming that the given digital asset collection is strategically valuable, the vulnerability of these assets should be reduced to an acceptable level. A digital asset risk manager can facilitate the process. First, the overall risk tolerance of the stewards and owners is established and the value/vulnerability model calibrated. Second, a relative valuation assessment and articulation process is undertaken for the given collection. Third, the relevant risks are analyzed and assessed in order to assign probabilities and severities. The initial risk profile of the collection is then established by mapping the collection's risks onto the value/vulnerability matrices. Next, risk management activities are listed along with their cost estimates and their effects in moving risks out of the red or yellow zones. A report is generated showing the cost of each strategy or combination of strategies and the risk reduction benefit as illustrated by the value/vulnerability matrices. If so mandated, the report may include a recommendation. The stewards, owners and funders then decide on a strategy.

The digital asset risk manager may implement or assist with the implementation of the selected risk management strategy. To provide for periodic iterative assessment and communication with and among the stewards, owners and funders, a yearly audit may be performed to verify that the risk management strategy is performing as expected. To complete the risk management cycle, at an interval of 3-5 years, the process would be repeated to take into account the evolution of the use and the value of the preserved assets [whether they have become more valuable or less] as well as the evolution of technology and other aspects of steadily improving digital preservation practice. This would provide the opportunity to periodically roll forward relatively cheap and basic risk management strategies until such time as better solutions become available or more affordable and, in a larger sense, enables the future discovery and implementation of different, potentially more optimal risk management strategies.

Concluding Comments/Further Work

This paper has synthesized recent research relating, in different ways, to the challenges of digital preservation: assessing value, understanding risk and vulnerability, and making cost-effective decisions on strategy. It has brought into the discussion concepts from financial risk management dealing with portfolios of financial assets and applied them in analogous ways to collections of digital assets. It may be possible to push these analogies further to bring increased sophistication to the risk management decision-making models in the digital asset space.

A possible path has been laid out for the asset value problem; however, any ability to add a quantitative aspect would greatly strengthen the process. Along the risk dimension, further work is needed to estimate [even subjectively] probabilities and severities for the various risk elements perhaps leading to a sort of "risk catalog" for different digital asset classes. Developments in research on the full life-cycle cost of digital preservation can be helpful in estimating the costs of different risk management strategies; however, more work is needed to tie the costs of specific components of the digital preservation process to the specific risks being hedged. Ideally, a pilot case study could be undertaken in which a small and relatively simple collection would be run through the risk management decision-making process described above.

The ultimate objective is to provide a tool and a process that stewards and owners of all kinds of digital asset collections - whether large or small, complex or simple - can use to rationally evaluate their options, successfully make their business cases to funders and then implement the risk management and digital preservation strategies that will ensure that these assets will be authentic, accessible, understandable and usable for as long as they are judged to have strategic value.

Appendix 1a: V-V Model™ : Conservative Risk Tolerance

Worst near certain impact acceptable?	50/50 odds isolated/fully recoverable
Worst unlikely impact acceptable?	Widespread/fully recoverable

Green 10
 Yellow 15

		1	2	4	8	16	32
	5	25	50	100	200	400	800
	4	20	40	80	160	320	640
	3	15	30	60	120	240	480
	2	10	20	40	80	160	320
	1	5	10	20	40	80	160
5 highest value		1	2	4	8	16	32
	5	25	50	100	200	400	800
	4	20	40	80	160	320	640
	3	15	30	60	120	240	480
	2	10	20	40	80	160	320
	1	5	10	20	40	80	160
3 medium value		1	2	4	8	16	32
	5	15	30	60	120	240	480
	4	12	24	48	96	192	384
	3	9	18	36	72	144	288
	2	6	12	24	48	96	192
	1	3	6	12	24	48	96
1 lowest value		1	2	4	8	16	32
	5	5	10	20	40	80	160
	4	4	8	16	32	64	128
	3	3	6	12	24	48	96
	2	2	4	8	16	32	64
	1	1	2	4	8	16	32

Appendix 1b: V-V Model™ : Aggressive Risk Tolerance

Worst near certain impact acceptable?	Total/fully recoverable
Worst unlikely impact acceptable?	Total/fully recoverable

Green 20
 Yellow 100

		1	2	4	8	16	32
	5	25	50	100	200	400	800
	4	20	40	80	160	320	640
	3	15	30	60	120	240	480
	2	10	20	40	80	160	320
	1	5	10	20	40	80	160
5 highest value		1	2	4	8	16	32
	5	25	50	100	200	400	800
	4	20	40	80	160	320	640
	3	15	30	60	120	240	480
	2	10	20	40	80	160	320
	1	5	10	20	40	80	160
3 medium value		1	2	4	8	16	32
	5	15	30	60	120	240	480
	4	12	24	48	96	192	384
	3	9	18	36	72	144	288
	2	6	12	24	48	96	192
	1	3	6	12	24	48	96
1 lowest value		1	2	4	8	16	32
	5	5	10	20	40	80	160
	4	4	8	16	32	64	128
	3	3	6	12	24	48	96
	2	2	4	8	16	32	64
	1	1	2	4	8	16	32

References

- Addis, Matthew and Richard Wright. 2010. *PrestoPrime: Keeping Audiovisual Contents Alive: Audiovisual Preservation Strategies, Data Models and Value-chains* (Deliverable 2.1.1). Southampton, UK: University of Southampton IT Innovation Centre.
https://prestoprime.ina.fr/public/deliverables/PP_WP2_D2.1.1_preservationstrategies_R0_v1.00.pdf
- Amazon Web Services. 2010. "Amazon Simple Storage Service (Amazon S3)." Amazon. Accessed August 30, 2010. <http://aws.amazon.com/s3/#pricing>
- Ayris, Paul, Richard Davies, Rory McLeod, Rui Miao, Helen Shenton, and Paul Wheatley. 2008. *The LIFE² Final Project Report*. London: JISC. <http://eprints.ucl.ac.uk/11758/1/11758.pdf>
- Baker, Mary, Mema Roussopoulos, Mehul Shah, Petros Maniatus, Prashanse Bungale, David S. H. Rosenthal, and TJ Giuli. 2006. "A Fresh Look at the Reliability of Long-term Digital Storage." In *EuroSys '06*. <http://www.lockss.org/locksswiki/files/Eurosys2006.pdf>
- Barateiro, Jose, Goncalo Antunes, and Jose Borbinha. 2009. "Addressing Digital Preservation: Proposals for New Perspectives." In *1st International Conference on Innovation in Digital Preservation (InDP '09)*. <http://cs.harding.edu/indp/papers/barateiro7.pdf>
- Barroso, Luiz Andre and Urs Holze. 2009. "The Datacenter as a Computer: An Introduction to the Design of Warehouse-scale Machines." In *Google, Inc. Synthesis Lectures on Computer Architecture no. 6*. San Rafael, CA: Morgan and Claypool.
<http://www.morganclaypool.com/doi/pdf/10.2200/S00193ED1V01Y200905CAC006>
- Beagrie, Neil, Brian Lavoie, and Matthew Woolard. 2010. *Keeping Research Data Safe 2*. London: Higher Education Funding Council for England/JISC.
<http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>
- Blue Ribbon Task Force. 2008. *Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation*. Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access.
http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf
- Currall, James and Peter McKinney. 2006. "Investing in Value: A Perspective on Digital Preservation." *D-Lib Magazine* 12(4). <http://www.dlib.org/dlib/april06/mckinney/04mckinney.html>
- espida. 2007. *espida Handbook: Expressing Project Costs and Benefits in a Systematic Way for Investment in Information and IT*. Glasgow, UK: University of Glasgow/JISC.
https://dspace.gla.ac.uk/bitstream/1905/691/1/espida_handbook_web.pdf
- Goldstein, Serge J. 2010. "DataSpace: A Funding and Operational Model for Long-Term Preservation and Sharing of Research Data." In *Educause Live!*
<http://www.educause.edu/Resources/DataSpaceAFundingandOperationa/212032>

- Heydegger, Volker. 2008. "Analysing the Impact of File Formats on Data Integrity." In *Archiving 2008*. http://slidefinder.net/h/heydegger_archiving2008/5907041/p2
- Jiang, Weihang, Chongfeng Hu, Yuanyuan Zhou, and Arkady Kanevsky. 2008. "Are Disks the Dominant Contributor for Storage Failure? A Comprehensive Study of Storage Subsystem Failure Characteristics." In *6th USENIX Conference on File Storage Technologies (FAST'08)*. http://www.usenix.org/events/fast08/tech/full_papers/jiang/jiang.pdf
- Joint Information Systems Committee (JISC). 2009. *InfoNet InfoKit: Risk Management*. London: JISC. <http://www.jiscinfonet.ac.uk/InfoKits/risk-management/printable-version.pdf>
- Kaplan, Robert S. and David P. Norton. 1992. "The Balanced Scorecard: Measures that Drive Performance." *Harvard Business Review* 70: 58-63. <http://hbr.org/2005/07/the-balanced-scorecard/ar/1>
- Kelly, Kevin. 2008. "Better than Free." *The Technium Blog*. http://www.kk.org/thetechnium/archives/2008/01/better_than_fre.php
- McHugh, Andrew, Raivo Ruusalepp, Seamus Ross, and Hans Hofman. 2007. *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*. Edinburgh, UK: DCC and DPE. <http://www.repositoryaudit.eu/private/?id=5076257c3b4c2699bfd92cf0e010e03e>
- McLeod, Rory, Paul Wheatley, and Paul Ayris. 2006. *Lifecycle Information for E-literature: Full Report from the LIFE Project*. London: University College London and the British Library. <http://eprints.ucl.ac.uk/1854/1/LifeProjMaster.pdf>
- Moore, Richard L., Jim D'Aoust, Robert H. McDonald, and David Minor. 2007. "Disk and Tape Storage Cost Models." In *Archiving 2007*. http://www.cs.ucsb.edu/~chong/290N/dt_cost.pdf
- Panzer-Steindel, Bernd. 2007. "Data Integrity." Draft 1.3. CERN/IT group. <http://indico.cern.ch/getFile.py/access?contribId=3&sessionId=0&resId=1&materialId=paper&onfId=13797>
- Pinheiro, Eduardo, Wolf-Dietrich Weber, and Luiz Andre Barroso. 2007. "Failure Trends in a Large Disk Drive Population." In *5th USENIX Conference on File Storage Technologies (FAST'07)*. http://static.googleusercontent.com/external_content/untrusted_dlcp/labs.google.com/en/us/papers/disk_failures.pdf
- Rosenthal, David. 2010. "Bit Preservation: A Solved Problem?" *The International Journal of Digital Curation* 5(1). <http://www.lockss.org/locksswiki/files/Bit-preservation-ijdc.pdf>
- San Diego Supercomputer Center (SDSC). 2010. "Storage & Backup." The Regents of the University of California. Accessed August 30, 2010. <http://www.sdsc.edu/services/StorageBackup.html>

- Schroeder, Bianca and Garth A. Gibson. 2007. "Disk Failures in the Real World: What Does an MTTF of 1,000,000 Hours Mean to You?" In *5th USENIX Conference on File Storage Technologies (FAST'07)*. http://www.usenix.org/events/fast07/tech/schroeder/schroeder_html/index.html
- Wheatley, Paul and Brian Hole. 2009. "LIFE³: Predicting Long Term Digital Preservation Costs." In *iPRES2009: The Sixth International Conference on Preservation of Digital Objects*. <http://www.life.ac.uk/3/docs/ipres2009v24.pdf>
- Wright, Richard, Ant Miller, and Matthew Addis. 2009. "The Significance of Storage in the "Cost of Risk" of Digital Preservation." *The International Journal of Digital Curation* 4(3). <http://www.ijdc.net/index.php/ijdc/article/viewFile/138/173>