

Conceptualising Optimal Digital Preservation and Effort

Sean Mosely, Jessica Moran,
Peter McKinney, Jay Gattuso
National Library New Zealand
P O Box 12349
Wellington 6001
+64 4 474 300
{sean.mosely,
jessica.moran,peter.mckinney,
jay.gattuso}@dia.govt.nz

ABSTRACT

In this paper we describe the National Library of New Zealand's attempts to conceptualise how we measure the degrees of effort required to achieve acceptable levels of digital preservation. We argue that understanding digital preservation practice in terms of "optimal effort" may help us conceptualise where and how best to achieve the greatest impact in proportion to effort. The paper examines the various roles of digital preservation, including the archival/curatorial, digital object handling, preservation management, and policy roles through case studies of our experience. We argue that through conceptualising our ideal digital preservation and the levels of effort required to achieve those, we will be able to better understand where our effort is being expended and the levels of preservation we are achieving.

Keywords

Digital preservation; digital preservation roles; effort; ingest; format migration.

1. INTRODUCTION

The mission of digital preservation is relatively straightforward – to ensure that digital objects are kept safe and accessible for as long as they are required. In some sense this mission will always be aspirational. In this paper we will describe some of the challenges inherent in the practice of digital preservation, as well as some of the National Library of New Zealand's (NLNZ) attempts to define a level of comfort in our practice. The discipline of digital preservation demands practitioners be able to acquire digital objects, maintain them in a way that ensures their physical and contextual integrity, and to deliver them for consumption when required. Assuming that the time period for requiring these objects is infinite (or at least undefined), then the task of preservation will never be complete – only once the period of requirement has ended will we know whether our mission was successful for that object. Therefore, our goal is to understand where and how our effort should best be focused.

There is another aspect of our work where we will always be aspiring to an idealised, abstract goal; the relationship of the "original experience" of the object versus how it will be consumed in the future. Regardless of the preservation methodologies employed – migration, emulation, normalisation, hardware/software museum-based, etc. – there may always be some qualitative difference between how the object was previously consumed and how it will be consumed both now and in the future. It is coincidental that of these cited approaches, the one that appears to get the closest to that idealised goal (the hardware/software museum-based approach) also appears to be the hardest to guarantee over time. And even when all technical factors have been controlled, the objects will still have been removed from their original temporal context.

Over time, the actions we perform on digital objects will also threaten them, whether from necessary changes in format, environments, the behaviour of emulators, or some other as-yet unknown factors. Our efforts will always be our best attempt to retain what can never be fully retained—the pursuit of a myth of Total Preservation.

This paper uses the various roles or preservation actors, such as the archival or curatorial, digital object handling, preservation management, and policy, to examine the levels of effort and preservation achieved through two case studies: an ingest of a collection of archival organizational records, and a format migration of a set of WordStar files.

2. TOTAL PRESERVATION?

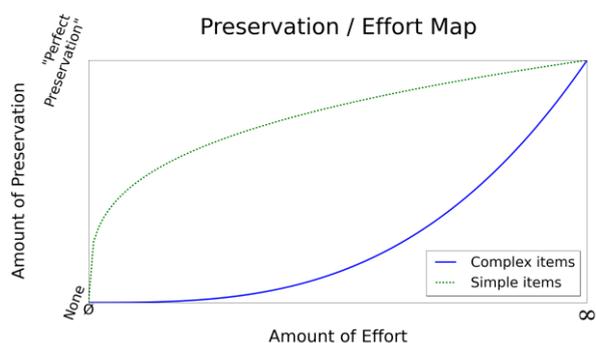


Figure 1. Relationship between preservation and effort over time, for both simple and complex items.

Figure 1 shows a basic illustration of the relationship between the goal of "Total Preservation" and the amount of effort required to achieve this goal. The presumption in digital preservation is that as our efforts increase, be they in the amount of time we take to understand context, identify and validate file formats, ensure we have stable systems to manage and store, and develop and maintain adequate policy to manage our process, the closer we will get to "Total Preservation."

However, it should be noted that in practice the above model does not hold true for all digital objects. For certain types of use cases, it appears that the initial preservation challenges will be very difficult, but that once these initial challenges are ironed out, the effort-to-reward ratio will most likely start to reflect that of simpler objects.

3. OPTIMAL PRESERVATION

3.1 Maximising Outcomes in Proportion to Effort

These two abstractions of effort-versus-reward are intended to reflect two hypothetical use cases – the first being a simple, self-contained resource which was created according to a widely-endorsed standard and can be rendered faithfully in a variety of applications (a tiff image, for example).

The second abstraction is intended to reflect a more complicated use case – such as a multi-file resource that is delivered via a server and involves the dynamic generation of content from a database, augmented by technologies which are being regularly updated or replaced (such as JavaScript-driven web application, retrieved from a LAMP stack and rendered through a modern web browser).

These abstractions, while useful tools for intellectualising the scale of digital preservation workloads, will change depending on the specifics of each use case and the preservation methodologies employed. Later in this paper, when the NLNZ’s preservation actions are plotted in regard to this effort-versus-reward spectrum, the exponential curves will be replaced with straight-line steps, in order to situate those actions in a more quantifiable space.

All memory institutions, regardless of their size or the extent of their resources, are affected by the realities of this effort-versus-reward ratio. As the discipline of digital preservation has become more widespread and more institutions begin to address their backlogs of digital content, more practitioners have started to discuss how to maximise their output for their efforts. Such conversations have given way to initiatives like POWRR (Preserving digital Object With Restricted Resources) [10] and the State and University Library of Denmark’s Minimal Effort Ingest approach.[5] These initiatives acknowledge the difficulty of adhering to the ‘best practice’ ideals of the discipline, and the practitioners seek to establish more achievable baselines for digital preservation.

The goal of such initiatives is laudable. By attempting to lower the barrier of entry to the discipline, these initiatives have the potential to encourage additional institutions to implement their own preservation strategies, and start to actively preserve content before it approaches a point of obsolescence. However, the terminology used in such initiatives may be problematic – POWRR’s approach of “good enough” digital preservation and the Danish State and University Library’s “minimal effort” are couched in language that has the potential to misrepresent the very nature of preservation.

To a degree, this use of diminishing language reflects the broader Information Technology industry as a whole. ‘Laziness is a virtue’ has long been a mantra of developers and system administrators¹, and the notion of ‘good enough’ or ‘just in time’ workflows has driven many large-scale IT businesses.² However, whereas digital technologies in general may benefit greatly from an approach that seeks to limit extraneous effort (for example, developing an application in Python rather than

C++ when it is determined that the development time savings will outweigh the potential performance gains), preservation is often a different matter. The discipline of digital preservation is still very much in its infancy, and if our language suggests to new practitioners that it is prudent to shy away from the emergent challenges, then there is a much greater risk that the alarmist claims of a ‘Digital Dark Age’ will become real.

In light of this, the discussions of “good enough” and “minimal effort” should perhaps be reframed as “optimal effort” – in other words, how do we find the best way to measure and maximise our efficiency for preserving digital objects? We want to leave room in our model as well to stress the importance of contributing to investigations into new preservation technologies, as innovations will allow us to preserve more content and further maximise our outcomes in proportion to effort.

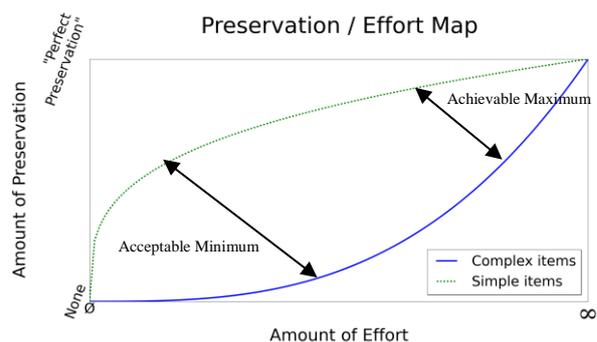


Figure 2. Optimal preservation will tend towards a point between ‘acceptable minimum’ and ‘achievable maximum’.

3.2 Preservation Actors

In order to evaluate our potential for efficiency in preserving digital objects it is helpful to conceptually break up the discipline into its different roles and responsibilities. Even in small institutions where the curator may also be responsible for technical analysis or system administration, the scope for acceptable loss may be different depending on the role. For instance, it may be considered an acceptable risk to undertake sampled quality control of descriptive metadata or access derivatives, but a sampled approach to checksum validation may be unacceptable. For the purposes of this paper, we have chosen to break up the responsibilities of digital preservation into four responsibilities:

- Archival/Curatorial
- Object Handling
- Preservation Management
- Policy

3.3 Archivist /Curator

This role is the traditional agent who advocates for the collection and manages the relationship between the collection, its scope and the contents of individual files and digital objects. This role typically relies heavily on human decision making, using training and experience to understand the intellectual, evidential, heritage, or value of an item. Further, this role should understand the relative impact of any changes, gaps, and other such subjective measures that might be encountered through the processing of any given file.

¹ The three virtues of programming – laziness, impatience and hubris – are widely attributed to Larry Wall, founder of the Perl language. Wall, along with co-authors Tom Christiansen and Jon Orwant, promote these virtues in their highly-successful and influential book *Programming Perl*. [12]

² The example of Amazon.com’s initial book sales business model (where stock would not be kept on-hand, but rather ordered only once the customer had paid for the purchase) is a canonical example that, for the sake of brevity, will stand in here for a more comprehensive summary.

3.4 Object handling

This role is charged with delivering the technical expectations of the Archivist/Curator role, and ensuring that files are engaged with according to recommendations. The role also provides some technical information and advice to the archivist / curatorial role, helping intellectual decisions to be informed where relevant by technical parameters. We imagine this role to be an even mixture of human decision making, and scripted logic and inference.

3.5 Preservation Management

This role is responsible for the underpinning technologies that bind digital collections together. We imagine this role to manage the Digital Preservation System (DPS) in its widest definition. Parts include digital shelving, per item cataloguing and indexing, processing workflows, and managing other generally automated functions. This role also includes management of digital storage, and regular system updates, testing, and implementation

We imagine this role to be primarily systemic in essence, including the combined processing and feature set of all related applications used to manage and process collections.

3.6 Policy

This role moves across all decision making layers, informing the institutionally agreed processes and functions that are applied to all digital collections.

We imagine this role to represent the collective institutional wisdom and knowledge that determines what can and cannot be undertaken, and the process through which any final outcomes are measured and approved.

3.7 Interpretation

The interpretation continuum (Figure 3) represents the type of reasoning that is required at the various levels to ensure that any interaction or intervention with any given file or digital object is being properly handled. The Archival/Curatorial role is predominantly interested in the intellectual content, context, provenance, and chain of custody of objects. These concerns include: what the digital object represents, its expected access by readers, its relationship to other objects in the collection, and how pre-ingest and ingest activities may affect an object's authenticity. The Object Handling role provides information to the Archival/Curatorial role on the technological possibilities and limits. This role also works closely with the Archivist/Curator to ensure digital objects are handled properly and technical solutions are developed. The System role is predominantly concerned with the technical representation of the object – what encodings are being used to bind information to any representation, what processes or operations are permitted and how they are carried out, how the host operating system and file store understands the binary stream and its attendant metadata. The Policy role helps develop the principles and directions to which the other roles will work.



Figure 3. Interpretation continuum.

Figure 3 attempts to take the notion that for any unit of effort, many files may be processed in a light or basic way, or few files may be processed in an intensive or complex way, and understand the way that all the roles rely upon and work together. In this construct, effort is seen as a mixture of resources (people, time, money) and capability (skill, tools, knowledge).

Essentially, while we conceptualise these roles as separate and independent, in effect they must work together, bringing their different expertise's to bear on the decision making processes. As we understand it, our work is continually being informed by both policy and the intellectual and technical parameters necessary to achieve what we think of as optimal digital preservation at any one moment.

4. APPLYING CONTINUUM TO A COLLECTION

4.1 Collection Description

To understand how we are applying this continuum model in more detail, it is helpful to apply it to a sample collection from the Library. This sample collection consists of the business records of an organization and was transferred to the NLNZ in 2013. The records came to the Library in two transfers over the course of two months with the organisation's IT department transferring the records from shared drive storage to external hard drives supplied by the Library. Prior to the transfer, a Curator and Digital Archivist visited the organisation, interviewed the donor about the kind of materials to be transferred and were given a high-level overview of the records. Together, we selected the areas of the shared drive the Curator appraised to have transferred to the Library based on the Library's collection and digital preservation policies.[7] Like other institutions, we believe time and effort should first be expended to develop policies around what and how we will collect and preserve digital content. We rely on these policies to guide our decision making throughout the appraisal and ingest workflow.[1] Based on this visit we suspected the records to consist largely of business records created in a Windows environment, using standard and well-supported file formats including the Microsoft Office Suite, pdf, tif, and jpg files, and most created in the last 10-15 years. At this stage our understanding of the collection was based only on an initial visual appraisal of the records.

4.2 Technical Appraisal

Once the collection was ready to be transferred to the Library, technical and intellectual analysis of the collection began. We determined that the collection consisted of 4239 individual files, and while at the top levels the records were well organized, in many cases the file paths were five or six levels deep; the collection had a total of 355 folders and a total size of only 4 GB. The collection dates ranged from 1997-2012.

We expected that because these were current business records, maintained by the organisation, and using widely adopted file formats, that our digital preservation challenges would be minimal. However, during technical appraisal the digital archivist discovered that the collection included 316 files with mismatched file extensions and 10 files whose format was unidentifiable,³ as well as a number of files with special characters in the filenames.

³ The Library uses DROID for format identification as part of its pre-ingest technical appraisal and within its DPS.

The digital archivist at this pre-ingest phase had a number of decisions to make in terms of how to best prepare the collection for ingest into the digital preservation system (DPS). The collection had to be appraised, and arrangement and description performed. These processes were done by subject experts in those roles with advice from the digital archivist. In this case, because the collection had been transferred intact from the shared drive of the organisation, there were some records that the Curator did not want to collect and preserve. Based on a desire to retain the deep file structure of the collection the digital archivist worked with the arrangement and description librarian to describe and further appraise the collection. At these points the digital archivists provide technical advice to the curator and arrangement and description librarian about the types of file formats in the collection, how they were created, and how best to describe them for future preservation and access. In this work the digital archivist is informed by the Library's digital preservation policy as well as both intellectual and technical knowledge about the records and an understanding of the DPS system and its strengths and limitations.

4.3 Preparing for Ingest

A number of policy and business decisions inform how we ingest material into our DPS. These include: records going into the DPS will have at least a scant collection management record, and files should pass validation checks, including format identification and validation. While neither of these is necessary for ingest, the Library has made the policy decision that by doing this work at ingest we are better prepared to understand our growing collections overtime and better able to make decisions about what sort of preservation actions we may need to perform in the future.[9]

Once appraisal and arrangement and description were complete, we were ready to being the process of ingesting the collection. First, we identified those files which could be easily ingested via the Library's internal submission tool already developed and integrated with the DPS. Using this tool, the digital archivist is able to build Submission Information Packages (SIPs) that automatically deposit the files and metadata to our DPS. In this case we selected those files which would need no preconditioning or provenance notes added to their preservation metadata, and that were all part of the same folder groupings. If these two conditions were met they could be quickly deposited using our ingest tool. This method accounted for 966 IEs or 23% of the collection, and in this case the greatest effort was expended earlier in the development of the ingest tool. Next, the digital archivist filtered out from the remaining records all those files that had been appraised out of the collection during appraisal or arrangement and description. This accounted for another 705 files, or 17% of the collection. However, that left us with about 60% of the collection that could not be quickly or easily ingested. In this case, that was due to the organizational structure of the files, lack of format identification, mismatched file extensions, or some combination of the above. At this point the digital archivists handed the collection over to the preservation analyst to do more of the object handling and determine the best way to ingest the rest of the collection. The 60% of the collection represented about 2500 files, and while this is still a relatively small number of files, we deemed it too large a number to be ingest using our ingest application, because in order to retain both the file structures and apply the preconditioning and provenance notes would mean hand building hundreds of individual SIPs. We deemed this to be too much manual effort. Instead, we developed a script that could identify some of the main issues that would cause the files to fail validation during ingest, automatically address those issue

that could be fixed, apply the accompanying provenance notes, and prepare the files for ingest.[6].

This second round of ingest accounted for another 2429 IEs, or 57% of the collection. In this part of the ingest process most of the time and effort was taken in developing and testing the script and data wrangling to prepare the files for ingest. We now had less than 200 files remaining that could not be ingested. Some of these were more complex multi-file IEs identified during processing and loaded separately by the digital archivists. The remaining files included 4 files that can be loaded following the next PRONOM update, 12 files whose format we have been able to identify, but do not yet have signatures written, and 5 files whose format we could not identify and that still require further research.

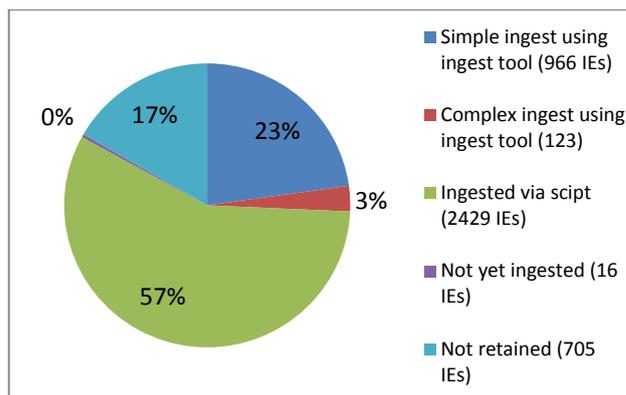


Figure 4. Types of ingest in collection.

Figure 4 illustrates the percentage of the collection by ingest type, and hints at the effort expended by type of ingest. For the first batch of 966 IEs (which were deposited by the simple ingest method), the effort in that case came in the development of our ingest tool and its stable integration with our system. In other words, this ingest method was simple because all the tools were already in place. Next, the 2429 IEs ingested via script required more upfront effort in understanding the objects, developing, and testing the script and the automated ingest method. Once that development and testing effort has been expended we anticipate being able to transfer the knowledge and tools developed in this collection for use in other similarly complex collections. The remaining 3% of the collection required the most effort, through manually preparing the files for ingest, format identification, writing of signatures, and other object handling.

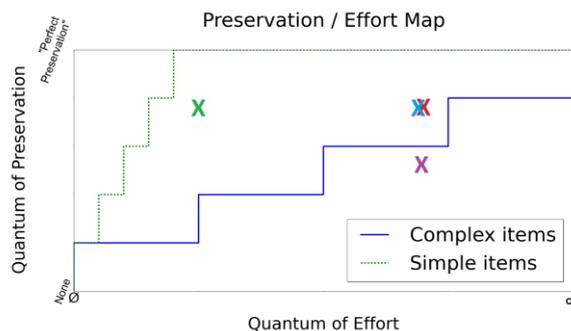


Figure 5. Effort mapped for sample collection ingest.

Figure 5 shows how we see the various parts of the collection mapped against our continuum of effort and preservation. For those parts of the collection that we were able to ingest with relatively little effort using tools already in place, we think we were able to reach an acceptable level of preservation with less effort. For those files where we had to either build more complicated SIPs using our ingest tool, or develop a scripted process to prepare the files for ingest, a much higher degree of effort was required to reach a similar level of preservation. The remaining files (those which have not yet been ingested into our DPS) have required more effort, but a much lower level of preservation has as yet been achieved. We can extend the same thinking across collections (Figure 6). For many of the collections that come into the Library and would not be categorised as complex by us, with less effort we can be confident in a high degree of preservation. For those collections that take longer to appraise and process, either due to size, file format, or condition of the collection, greater effort is needed to reach the same level of confidence in our level of preservation. Indeed in some cases we have already expended a great deal of effort for a much lower degree of degree of preservation.

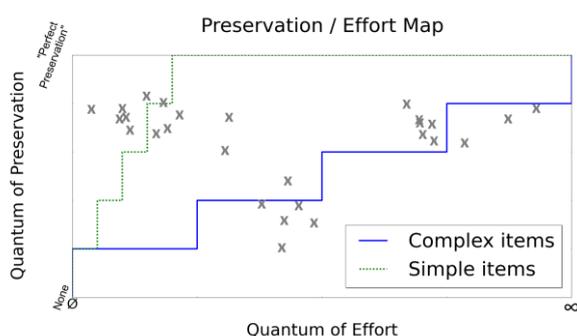


Figure 6. Effort mapped for ingest across many collections.

4.4 Preservation Management

Once the collection is ingested into the system, it then comes under the purview of preservation management. The activities in this area are in place to ensure that the objects remain free from corruption, are available and can be accessed over time (while maintaining any access restrictions).

Some of this work can be system automated. This includes routines such as periodic virus scanning and checking of fixity values. Other processes, however, are not automated. These include risk analysis and preservation planning and actions. Automation is not available for these processes due to the immature nature of the work. For example, we have recently undertaken a migration of WordStar files [4]. This migration was a handcrafted solution, both in terms of generating tools from scratch and taking deliberate time and care with decision-making. It was handcrafted because: a) existing tools for the conversion were untrustworthy; b) we wanted to ensure the process was robust; and c) non-technical staff required time to understand the activities we were undertaking and had to be assured that the resulting files could stand as authentic representations of the originals (and conceptually replace them).

NLNZ is embarking on a programme of migrations. One of the outcomes of this programme (beyond the primary concern of mitigating risks) is that a clear process, including decision-making routines, is agreed upon. With this in place, far less

effort will be required to achieve the same outcomes as the WordStar conversion.⁴

In addition to these management aspects, there is of course the underlying architecture upon which the management takes place. A preservation system is not (at least in our experience) a “plug-and-play” operation that can be left to its own devices. The underlying storage is reconfigured regularly, as is the processing layer. Such reconfiguration is due most often to the availability of new technologies, changes in requirements, new external vendors, and collection growth. There are also updates to the preservation system itself. Requirements need to be written, negotiation over the solution undertaken, testing, and eventual roll-out.

Preservation management can be best described as a container for iterative processes undertaken across the lifetime of the objects being cared for – the “lifetime”, in our case, being perpetuity.

4.5 Access

Access is in fact an added benefit of our current process and one not always discussed in digital preservation literature.[8] One of the benefits of expending this effort upfront to identify, prepare, and validate all the files that go into our DPS is that they are accessible immediately to researchers. Because the files have been identified, validated, and have the correct extensions, files in our DPS system are accessible and can be delivered to our users either through viewers or they are downloadable and accessible to users, even those with little technological confidence. By performing our quality assurance and pre-conditioning actions during ingest, the files can be delivered back to our researchers in a format they and their computers are more likely to understand with little or no intervention from us.

4.6 Policy

Underpinning all of this work is the policy layer. Policy informs and aids decision making at each step of the process. Our policy is aimed at the level of operating principles⁵. That is to say, we describe the goals and aims for policy areas and describe the high level processes that should be undertaken and within what boundaries. However, the policies do not go into highly detailed specifics. For example, the fixity policy will contain principles detailing that more than one fixity information method must be used, but it does not specify which ones should be used.

Each policy created requires a large amount of effort to create. There is consultation, drafting, further consultation, redrafting and, if lucky, sign-off (our experience tells us that this process has to be tightly controlled in order to avoid a constant spinning in the drafting and consultation phase). After a few years, there is also the review of the policies. No little effort is expended across all preservation stakeholders while creating and reviewing these policies.

While policies bring us towards optimum effort through normalizing and codifying practice, they do not completely diminish effort. The policies do not define the exact steps that must be taken by each staff member; they are not business process models. Therefore there is still some layer of effort by the member of staff as they put into practice policy principles.

⁴ Indeed, we are even testing a migration that could be classed as “quick and dirty”. This should help us explore the boundaries of our comfort around the integrity and authenticity of the content and what measures are required to give this comfort.

⁵ For an excellent discussion on preservation policies and the different levels they operate at, see Sierman, 2014.[11]

5. EFFORT

The 'effort' axis on all of the charts above is unencumbered by any sort of scale. Effort can be measured across many different parameters and given many different scales. It could be staff hours, costs involved, or even perceived effort (a la the Borg rating of perceived exertion [3]).

In our conceptual world, effort is predominantly envisaged as staff hours (and perceived staff hours). But it also contains a trace of costs (some of our tool development is outsourced and therefore monetized in a way that our team efforts are not). Hidden in there are also costs of storage and consultancy, which we interpret as part of the effort of preserving.

It seems pertinent to note that even if the scale measured defined costs it is very hard to get a true sense of what, for example, staff resources are being spent on. If an organisation employs two members of staff and assigns them key tasks it is, we would argue from experience, difficult to actually gauge the time spent on, say, format identification, or risk analysis, or the validation of preservation actions. Additionally, are these areas of work more valuable than the support work that is spent wrestling with recalcitrant tools, or tracking down bugs in code, re-architecting server configurations, and testing system upgrades? This is one of the reasons we are reluctant to lay down any definite scale for effort.

This paper is deliberately vague about effort and therefore also about costs. These should be understood in relation to the reader's own organizational context. The determination of exact costs for digital preservation activities, and thus, defining an exact scale on the "quantum of effort" axis would require an in-depth community-agreed process for digital preservation. Even delving into small aspect of the preservation process highlights differences in practice that would change costs. For example, some institutions require and exact single format identification for each object, thus requiring in some cases relatively extensive research to be undertaken. Others accept all possibly format identifications including "format unknown". The Blue Ribbon Task Force have worded this far more eloquently suggesting that arriving at an estimate for preserving an amount of data over time is "a task over-laden with the details of a particular implementation and preservation context" [2].

Likewise, the notion of deferred costs has not been explicitly addressed. This is not to say that there are not ongoing discussions about the upfront costs and delayed costs. Should we be spending n amount of days at the point of ingest on certain issues, or rather bring in the material as is and resolve any issues at a later date? Our experience tells us that that when we leave something for a later date, that later date rarely occurs. Our current (non-written) policy is that we must give best efforts now in order to give future staff (and even our future selves) the best possible chance of continuing to offer access to the material.

The final word to address is "efficiency". Efficiency has very good associations to higher rates of productivity, but also has quite negative connotations of adequacy and trimming of outcomes. For the sake of this paper, we will focus only on the positive and consider one final example.

We have mentioned above the work undertaken on converting WordStar files. A great deal of effort was expended in that process. It is clear though that we could have made that process far more efficient: we deliberately slowed much of the process down in order to guarantee that all stakeholders followed every single step of the process even if they were not directly involved in each step. The final graph below shows how we could have put far less effort in for probably exactly the same outcomes (in terms of the content). But what the graph (figure

7) does not show us is the institutional outcomes that were achieved by taking this slowly, slowly approach.

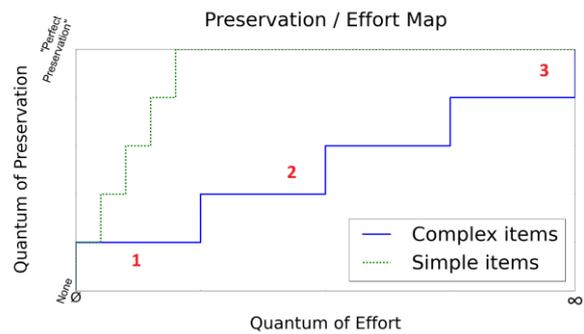


Figure 7. Effort mapped against a single format.

In the graph above we have set the levels of preservation as a very coarse series of descriptors; No preservation; Binary only; Format Identified; Generally Renderable; Obsolesce risk migrated. This is a rough scale for demonstration purposes, and not a measurement we would specifically advocate for.

The quantum of effort is also an approximation, based on the work required to successfully ingest and migrate the original content.

The first marker (1) indicates the state we found the collection in. We could have ingested the original WordStar files as unknown binary items with minimal effort, but as this is against our general working methodology and business rules of only ingesting content that has a format identification we can use the map to indicate where we start.

The second marker (2) indicates the effort required to create a successful format signature, get that included into the format registry, and ingest the content into the preservation system.

The third marker (3) indicates the effort required to migrate the content from the original format into its new contemporary format.

6. CONCLUSION

By conceptualising where our idealised digital preservation is, and what levels of effort are required to achieve it, we can better understand where we are currently expending our effort and what level of preservation we are achieving. Charting this effort and preservation will allow us to begin quantifying what we are doing, what direction we want to move in, and how best to expend our effort to achieve better efficiencies in our digital preservation work. What we do believe (but won't be able to test until some future time) is that the effort we expend will result in the National Library of New Zealand being able to deliver a digital collection back to a user in a way that they can understand its organisation, its context, have trust in its authenticity, and can easily access the objects and their metadata into perpetuity.

7. ACKNOWLEDGMENTS

Thank you to other staff in the Library whose input has shaped our thinking including Leigh Rosin, Gillian Lee, Susanna Joe, Sholto Duncan, Kirsty Smith, Cynthia Wu, and Suman Sajwan.

8. REFERENCES

- [1] AIMS Work Group. *AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship* (January 2012). https://dcs.library.virginia.edu/files/2013/02/AIMS_final_t_ext.pdf.
- [2] Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010). *Sustainable Economics for a Digital Planet: Ensuring Long-term Access to Digital Information*. http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf.
- [3] Borg GA (1982). 'Psychophysical bases of perceived exertion'. *Med Sci Sports Exerc* 14 (5): 377–81.
- [4] Gattuso, J. & McKinney, P. "Converting WordStar to HTML4" *Proceedings of the 11th International Conference on Digital Preservation* (Melbourne, Australia, 6-10 October 2014.)
- [5] Jurik, B., Blekinge, A., and Christiansen, K. Minimal Effort Ingest. In *Proceedings of the 12th International Conference on Digital Preservation* (Chapel Hill, North Carolina, November 2-6, 2015.)
- [6] Moran, J. and Gattuso J. Beyond the binary: pre-ingest preservation of metadata. In *Proceedings of the 12th International Conference on Digital Preservation* (Chapel Hill, North Carolina, November 2-6, 2015.)
- [7] National Library of New Zealand. *National Library of New Zealand Collections Policy* (October 2015.) [http://natlib.govt.nz/about-us/strategy-and-policy/collections-policy/national-library-of-new-zealand-](http://natlib.govt.nz/about-us/strategy-and-policy/collections-policy/national-library-of-new-zealand-collections-policy)
[collections-policy](http://natlib.govt.nz/about-us/strategy-and-policy/collections-policy/national-library-of-new-zealand-collections-policy) and Joint Operations Group, Policy (JOGP), Department of Internal Affairs, 2012, *Digital Preservation Policy Manual*. <http://ndha-wiki.natlib.govt.nz/assets/NDHA/About-Us/Strategic-Partnerships/Digital-Preservation-Policy-Manual.pdf>.
- [8] Peltzman, Shira. Expanding NDSA levels of preservation. In *The Signal Digital Preservation Blog*. (April 12, 2016). <http://blogs.loc.gov/digitalpreservation/2016/04/expanding-ndsas-levels-of-preservation/>
- [9] Rosin, L. 2014. Applying theoretical archival principles to actual born-digital collections. *Archival Journal* 4 (Spring 2014) <http://www.archivejournal.net/issue/4/notes-queries/applying-theoretical-archival-principles-and-policies-to-actual-born-digital-collections/>.
- [10] Schumacher, J. The Digital POWRR project – a final report to the Institute of Museum and Library Services (February 2015) DOI= <http://hdl.handle.net/10843/13678>
- [11] Sierman, Barbara. 'The SCAPE Policy Framework, maturity levels and the need for realistic preservation policies'. *Proceedings of the 11th International Conference on Digital Preservation* (Melbourne, Australia, 6-10 October 2014.)
- [12] Wall, L., Christiansen, T. and Orwant, J., *Programming Perl*. Third edition (Sebastopol, California, O'Reilly, 2000)