The Fragility of Digital Media Content

On Preservation and Loss: Sketching the Pilgrimage of Future Scholars to Recover Our Digital Vellum

Maranke Wieringa

ABSTRACT

Preservation of media content is increasingly problematic with the rise of digital media. While it seems that we can backup everything in clouds, on USB sticks or external hard drives, the heart of the matter is that digital content degrades. Whereas non-digital media such as paper degrade slowly, a single corrupted bit can irrevocably harm a digital file. Digital media are thus on constant life-support. This paper is meant as an thought exercise to chart possible problems that future scholars might encounter when scavenging the digital archives. In particular, a number of hindrances are discussed: the lack of central institutions that take on the task of archiving content, inadequate modes of archiving, obsolescence, instability, and the compromise of the digital artifact's integrity. It seems that the future of preserving our contemporary cultural artifacts is rather bleak, yet some preservation strategies may give a ray of hope.

KEYWORDS

Preservation, Media Content, Digital Media, New Media, Archiving

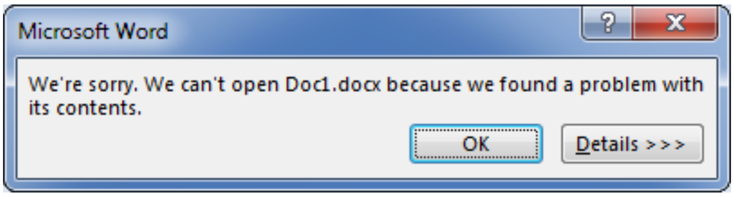
INTRODUCTION

With the various options we have to save our documents on clouds, external hard drives, USB-sticks, and so on, the preservation of media content seems easier than ever. It would seem that laborious journeys, such as that of Poggio Bracciolini, a scholar who discovered the last remaining copy of Lucretius' *De Rerum Natura* in a monastery library in 1417, are a thing of the past. Poggio's journey, which is vividly described in *The Swerve* (Greenblatt 2011), seems to have become unnecessary in our contemporary society. Everything seems backed-up and retrievable with a few key strokes, and a click of a mouse. Simultaneously, however, the same digital media also has a frustrating ephemerality, which contrasts with this trust in the permanence of our data: who amongst us has not experienced the agony of losing work because they forgot to save the document; or perhaps lost vital information because of a corrupted file? It seems, then, that there is more to our 'permanent' file-storage then meets the eye, for behind its 'enduring' appearance lurks a volatile nature. While '[p]rint books and records decline slowly and unevenly - faded ink or a broken off corner of page' -, digital files 'fail completely - a single damaged bit can

Junctions: Graduate Journal of the Humanities
Maranke Wieringa. 2017.
Vol 2(2): 27-37. ISSN: 2468-8282
http://junctionsjournal.org/

render an entire document unreadable' (Rosenzweig 2011, 8–9). This, then, is *The Swerve*'s importance to me: a reminder of the frail nature of media content – especially in the digital age.

It is my hypothesis that, in our current timeframe, digital media content is more prone to being irrevocably lost to us in the future due to its increased fragility, as opposed to that of older (print) media forms. As Roy Rosenzweig (2011, 8-9) so graphically pointed out, there is no slow degradation of digital content. While a scorched book may still be (painstakingly) readable, it is a either/or situation for digital content: either it is readable, or it is not (e.g. Fig. 1). Moreover, one tends to forget digital-born material, when one thinks about digital archiving (Underhill and Underhill 2016, 2). Often, only digitized material is considered in these instances. In this essay I sketch the problems our future Poggio Bracciolinis may encounter in their quest for the recovery of media content. I highlight the lack of a specific public archive, the inadequate means/methods for archiving, and finally the increased pace in which media content needs to be maintained in order to avoid becoming inaccessible. Finally, I will provide some outlooks which provide hopes for the future, as I will discuss preservation strategies which might help prevent a digital dark age.



We're sorry. We can't open Doc1.docx because we found a problem with its contents.

Fig. 1 – Error Message Microsoft Word 2013

WHERE TO LOOK?

While it was fairly straight-forward for Poggio Bracciolini to search for ancient texts in the monastery libraries, there is – as of yet – no obvious archive where future scholars may turn to re(dis)cover digital-born content. Greenblatt's protagonist, Poggio, made a pilgrimage to a monastic library (assumingly that of Fulda), where he recovered Lucretius' *De Rerum Natura* (Greenblatt 2011, 181). The monks preserved (ancient) texts by laboriously copying them (Greenblatt 2011, 47). Similarly, institutions in current times are responsible for the preservation of documents and objects. Yet, as Rosenzweig (2011, 12) notes, '[d]igitiziation has unsettled' this 'system of responsibility'. What is problematic is that 'an alternative system has not yet emerged' (2011, 12). In other words, there is no well organized institution whose responsibility is to archive digital documents and objects.

The few preservation projects, such as the Internet Archive, that have emerged are often in private hands (Rosenzweig 2011, 18). These private archives are unstable, as the archive itself may not endure if there is not enough funding or interest. Such is the case of the (forced) shut down of

Google's 'Library of Alexandria' project (even though this project does not concern itself with digital-born material, but rather with digitized material). Originally, Google wanted to scan (or photograph, to be more precise) and preserve every book ever published. The project was forcibly shut down after copyright infringement claims (Somers 2016). Whereas government archives, for example, are bound by law to preserve specific texts, as the medieval monks preserved books in their monastic tradition (Kuny 1998, 1), private archives do not have the same obligation. Nothing – except perhaps good intentions and/or hopes for profit – can vouch for these private archives' endurance (Howell 2000, 129; Kuny 1998, 3). This, as Google's 'Library of Alexandria' project showed, also makes scholars and other stakeholders weary of trusting the preservation practices of such parties. This is why UNESCO pleads for collaboratively founded archives between cultural institutions and the 'creators of information and of software producers' (UNESCO and National Library of Australia 2003, 6). Rosenzweig (2011, 20) adds that '[a] combination of technical and organizational approaches promises the greatest chance of success', but also warns us that this success lies in a synthesis between the state – which would provide continuance, guaranteed access and funding – and the 'experiential and ad hoc spirit' of the private parties.

What further underlines this collaboration between the cultural institutions and the private/commercial sector is the digital information's need to be maintained from the moment of its creation (Harvey 2012, 12). Unlike print media, you cannot wait and see if something will be valuable in the future and then take preservative action. Due to rapid obsolescence, increased media instability and challenges to the integrity of the object, early action is required, otherwise we risk losing access to the data (Levi 2008).[1]

Apart from the pressing matters regarding the permanence of private archives and the need for rapid action, the issue of selection also comes to the fore. What complicates the process of selection is that the vast amounts of data that are currently saved can never be sifted through (Workshop on research challenges in digital archiving and long-term preservation 2003, 3). There is – paradoxically – too much data right now, while there may be too little adequately preserved for the future (Rosenzweig 2011). Selection is thus 'necessary because there are usually more things – more information, more records, more publications, more data – than we have the means to keep. Every choice to preserve is at the expense of something else' (UNESCO and National Library of Australia 2003, 70).

Who, then, gets to decide what is preserved and what is discarded? While current private preservation initiatives, such as the Internet Archive or the leftovers of Google's 'Library of Alexandria' project in the shape of Google Books, are surely better than nothing, they gather a limited amount of information that may not be the most valuable. The Internet Archive cannot

[1] Obsolescence, instability and challenges to integrity are discussed later in this essay.

archive password- or copyright-protected sites and so forth. This is problematic in a time where 'a great deal of the interesting digital information is not in the public domain' (Howell 2000, 129), but instead belongs to companies or other private parties (Kuny 1998, 9). Thus, if we want to have any hope in enabling future Poggios to rediscover valuable artifacts, it is important to strengthen collaborative efforts between the state and private endeavors.

INADEQUATE MODES OF ARCHIVING

To make matters worse, the current archives are inadequate in their mode of archiving. An 'archival language' for digital-born material does not seem to exist; and cultural institutions are struggling to find proper ways to archive digital-born material because they lack a precedent for this kind of archiving (Ernst 2013, 82). Where the medieval monks worked with an indexical gestural language to refer to particular books (Greenblatt 2011, 43–44), indexing has become trickier in the digital era.[2] The books the monks requested were static objects whose content would not vary (much). The digital objects we are encountering in our own time, however, are anything but static. While books are rigid objects, digital objects are a 'constant dynamic flow of information in cyberspace' (Ernst 2013, 122). In a time in which web pages can continuously be altered, there is a lack of a final object, and a potentially infinite amount of these incomplete web pages linked to the page in question.

This dynamic and interactive nature of media content is something that archives and cultural institutions are struggling with. As UNESCO notes, institutions were given little time to develop adequate strategies, as the '[d]igital evolution has been too rapid' for them to be able to keep up with it (UNESCO and National Library of Australia 2003, 13). The applied methods are often labor-intensive, and are not feasible trajectories in the long-run, especially when facing large-scale collections.[3] More troubling, we still lack 'well-developed methodologies for preserving many of today's complex data types and formats' (Workshop on research challenges in digital archiving and long-term preservation 2003, 7). Thus, even if our future Poggio would have successfully located a digital archive which could potentially hold valuable artifacts, a lot of it may still not be (fully) accessible due to inadequate archiving methodologies**.**

[2] The monks were not allowed to speak during their task, so they used a sign language to ask for particular texts.

[3] Preservation methods that are currently employed are print-outs, preserving the older technology, data migration and emulation (Rosenzweig, 2011, pp. 13-14).

OBSOLESCENCE, INSTABILITY AND COMPROMISED INTEGRITY: CONTEMPORARY BOOKWORMS?

Current digital preservation methodologies fall short because the pre-digital mode of archive differs dramatically from digital archiving (Harvey 2012, 10–13). One of the pre-digital strategies for preservation has been 'benign neglect' (Harvey 1993, 140). This strategy, however, 'no longer works for digital materials' (Harvey 2012, 17). Sure, the pre-digital times had their bookworms which deteriorated the archived material, but it would not render a book illegible as fast as we are currently experiencing. The ferocious bookworms we are facing nowadays are obsolescence, instability and challenges to the digital object's integrity (Harvey 2012, 55).

In short, where Poggio Bracciolini would be relatively sure that ancient texts could survive for years without attention – thus increasing his odds at finding something noteworthy – this seems to no longer hold true. Digital materials require 'different timetables for preservation action' (National Library of Australia 2013). What is worse, these new timetables are more far more pressing than their pre-digital counterparts.

Obsolescence

As the Digital Preservation Coalition notes: 'the speed of changes in technology means that the timeframe during which action needs to be taken is measured in a few years, perhaps only 2-5' (2008, 32). As hardware and software are continuously renewed and updated, older versions and their corresponding file formats become obsolete or illegible. Even in the event that digital storage media (such as floppy disks or CDs) are still in pristine condition, the computer or software may not be able to properly access the information. As Roy Rosenzweig (2011, 9–10) puts it:

> the ones and zeros lack intrinsic meaning without software and hardware, which constantly change because of the technological innovation and competitive market forces. Thus this lingua franca requires translators in every computer application, which in turn, operate only on specific hardware platforms. Compounding the difficulty is that the language is being translated keep changing every few years.

Thus, the digital media system is prone to obsolescence in three ways: hardware, software and code language. If one of these becomes obsolete, the data may become inaccessible. The pace in which one of these three is at risk of obsolescence is staggering. I, for one, have worked with seven sorts of medium storage/sharing technologies and eight different operating systems in fewer than twenty-five years.[4] Yet, even though I have worked with them, 3 ½ inch floppy disks seem to be unrecognizable as a storage medium for the current generation of children. We can only hope

[4] That is a new medium storage/sharing technology and operating system every ± 3.28 years on average.

that our future Poggio may indeed know what he is looking for and, perhaps more importantly, finds a way to retrieve information from the obsolete storage media (e.g. Fig. 2).

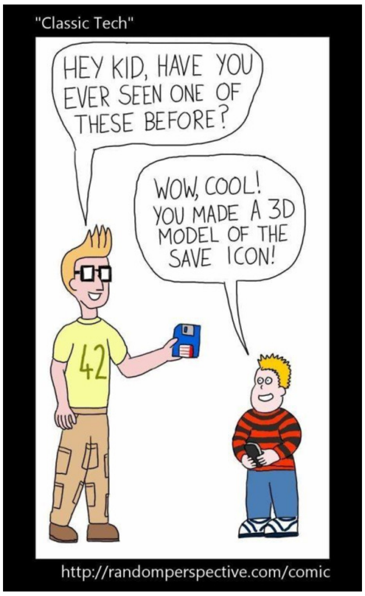

"Classic Tech"
HEY KID, HAVE YOU EVER SEEN ONE OF THESE BEFORE?
WOW, COOL! YOU MADE A 3D MODEL OF THE SAVE ICON!
42
http://randomperspective.com/comic

Fig. 2 – Rapid Obsolescence

Instability

Although we may conceive digital documents and objects as immaterial, they certainly are not without physicality. CDs/DVDs, USB-sticks, and so forth, would be the most obvious examples; but even the famous cloud is not without its materiality. As the Digital Preservation Coalition (2008, 33) notes: 'The media digital materials are stored on is [sic] inherently unstable and without suitable storage conditions and management can deteriorate very quickly even though it [sic] may not appear to be damaged.' For example, the physicality of storing information in the cloud comes in the form of data farms. These server agglomerations need extensive amounts of cooling (and electricity) to function. Whereas a book would not be (much/rapidly) affected when no attention is given, a data farm will overheat when not continuously cooled, which in turn may result in massive data loss (Fig. 3). However, it does not need to be a mass-scale catastrophe. A demagnetized floppy disk, for instance, is just as good an example, as data would similarly be lost.

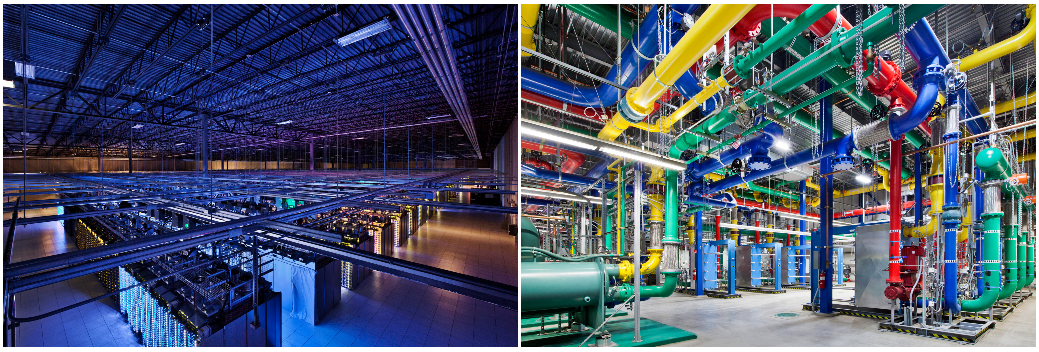


Fig. 3 – Interiors of Google Data Farms Left: Servers Right: Cooling Installation

Challenges to the digital object's integrity

The final contemporary bookworm is the compromised integrity of the digital artifact. 'The ease with which digital objects can be altered' means that we no longer trust in the object's integrity by default. To make matters worse, some preservation techniques (migration, for instance) may even compromise integrity (Harvey 2012, 54). This is why UNESCO (2003, 14) believes it is vital that the documentation about the artifact, that is, the digital material, is vital.

Documentation is not a novel idea. Poggio Bracciolini was supposedly familiar with Cicero and Ovid's references to Lucretius' *De Rerum Natura* (Greenblatt 2011, 51–52). These kinds of references to the work in the writings of other authors helped to identify the text as legitimate. What has changed is the intensity of documentation required. Now a document has to be documented to its details, for future scholars to be more certain they are dealing with the real thing, and not a facsimile (Harvey 2012, 54).

REASONS TO BE OPTIMISTIC(?)

While there are pressing and worrisome complications in the matter of digital archiving, there are also some optimistic voices. The National Archives of the United Kingdom believe that cloud storage may be a feasible option, provided that archives find ways to assure continuation and overcome legal and practical hindrances (Beagrie, Charlesworth, and Miller 2014). There are obvious advantages involved in cloud archiving, such as delocalization, potential cost-saving (for smaller archives), and the flexibility of cloud-storage (Beagrie, Charlesworth, and Miller 2014, 11). It is also important to look at some best practices in the field. For example, the University of California Irvine (UCI) acquired a digital-born collection of 2.5 terabytes in 2014, which became their largest digital-born collection up to that point (Uglean Jackson and McKinley 2016). Uglean Jackson and McKinley (2016, 67–70) found that the digital preservation strategies practiced up to that moment were too laborious for this big collection. Therefore, they adopted different strategies. They note, for instance, that the importance of donor interviews. Interviewing the benefactor is a way to ensure the data's integrity, but helps archivists to contextualize the data. Such interviews can then be complemented with algorithmic metadata retrieval. Similar to UNESCO (2014) and D. R. Harvey (2012), they also emphasize the importance of documentation. Nevertheless, Uglean Jackson and McKinley also encountered setbacks, as they 'decided to sacrifice easy file retrieval in order to continue using a reliable system that was already integrated into our preservation infrastructure' (Uglean Jackson and McKinley 2016, 74). Moreover, the accessibility of the files was not realized in 2016.

The National Digital Stewardship Alliance has also published on ways in which archives can check if their digital material is adequately archived. Among the most helpful tools for archives and archivists is the 'Levels of Digital Preservation' tool (Phillips, Bailey, Goethals, and Owens 2013). This tool (Table 1) formulates several different levels of digital preservation. As one progresses from level one to latter levels, 'one is moving from the basic need to ensure bit preservation towards broader requirements for keeping track of digital content and being able to ensure that it can be made available over longer periods of time' (Phillips, Bailey, Goethals, and Owens 2013, 4). While level 4 is undoubtedly preferred, this tool articulates the bare necessities and potential augmentation, even when such a comprehensive approach is not feasible. Thus, for smaller archives, or archives with lower budgets, such a tool provides an oversight and starting point for what to take into account.

These recent developments provide us with a hopeful outlook, but many obstacles still need to be overcome. The cloud archive proposed by the National Archives still needs to be maintained and secured, not only digitally, but physically as well. As the data is stored on server(s) (farms), it is still at risk of overheating or power failures. Nevertheless, best practices, such as the digital-born collection of UCI, may help pave the way for future preservation efforts. Moreover, organizations, such as the National Digital Stewardship Alliance, help archives formulate the bare essentials of digital preservation, and to chart what actions need to be taken.

CONCLUSION

'Benign neglect' as a preservation method, stable archives, and trusted methods of archiving build the foundation of Poggio Bracciolini's success. However, they change in the digital era, for digital data is on constant 'life-support', as it requires constant maintenance (Workshop on research challenges in digital archiving and long-term preservation 2003, 7). A future Poggio-like figure may, therefore, be confronted with massive hindrances.

	Level 1 (protect your data)	**Level 2 (know your data)**	**Level 3 (monitor your data)**	**Level 4 (repair your data)**
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs
Metadata	- Inventory of content and its storage location - Ensure backup and non-collocation of inventory	- Store administrative metadata - Store transformative metadata and log events	- Store standard technical and descriptive metadata	- Store standard preservation metadata
File formats	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

Table 1 – Version 1 of the Levels of Digital Preservation (Philips, Bailey, Goethals, and Owens 2013, 3)

The first obstacle would be to find where to look for digital material, as there is no clear 'system of responsibility' for the preservation of digital artifacts at the moment (Rosenzweig 2011, 12). Contemporary archives that attempt to preserve digital information are often private or commercial endeavors. Archives of these kind may not endure, as the parties involved often lose interest for various reasons. If archives are meant to last, it is recommended that they collaborate with multiple parties to safeguard its existence. Another reason for collaboration is that, currently, these private or commercial parties get to decide what is worth preserving and what is not, and their selection criteria may not be in future historians' best interests (Harvey 2012, 66–67). For a singular person, what is worth preserving may simply be what he/she finds interesting. For a company it may be what results in the biggest profit or good reputation. A final reason for collaborative archiving is that digital material requires early action, as it needs to be maintained from its creation onward (Harvey 2012, 12).

The second obstacle is the as-of-yet inadequate ways of archiving digital material. As opposed to the static texts that Bracciolini's monks were used to, the digital object is dynamic and interactive. It is connected to a potentially infinite number of other constellations and it needs not be textual. Thus, if our future Poggio would be able to locate an archive – he/she would not be able to trust that all the material has been archived in a suitable way. These preservation methodologies fall short because of our contemporary bookworms: obsolescence, instability and compromised integrity. Obsolescence has to do with the increased technological developments, which renders older technology irrelevant and often illegible to its successors. Instability refers to the digital medium's vulnerability. Whereas books could be left unattended for years, digital information needs active maintenance, which may result in more data loss. Finally, if our future Poggio manages to circumvent all of these problems, he/she may face a final bookworm: compromised integrity, when the authenticity of digital data can no longer be vouched for.

It seems, then, that a future Poggio Bracciolini would have quite a hard digital pilgrimage ahead of him/her to recover '[t]he traces of information (…) from our digital vellum' (Kuny 1998, 10). Nevertheless, recent efforts and innovations give a ray of hope. While there are still many obstacles to overcome, there have been fruitful best practices, interventions, and guidelines that raise the chances of preserving our future *De Rerum Natura*s.

REFERENCES

Beagrie, N., A. Charlesworth, and P. Miller. 2014. "How Cloud Storage can address the needs of public archives in the UK.*"* http://www.nationalarchives.gov.uk/documents/archives/cloud-storage-guidance.pdf

Digital preservation coalition. 2008. "Preservation Management of Digital Materials : The Handbook.' www.dpconline.org/component/docman/doc_download/299-digital-preservation-handbook.

Ernst, W. 2013. *Digital Memory and the Archive*. Minneapolis/London: University of Minnesota Press.

Greenblatt, S. 2011. *The Swerve*. New York/London: W.W. Norton & Company.

Harvey, D.R. 2012. *Preserving Digital Materials*. Berlin/Boston: Walter de Gruyter.

Harvey, R. 1993. *Preservation in Libraries: Principles, Strategies and Practices for Librarians*. London/Melbourne/Munich/New Jersey: Bowker-Saur.

Howell, Alan. 2000. "Perfect One Day—Digital The Next: Challenges in Preserving Digital Information." *Australian Academic & Research Libraries* 31 (4): 121–41.

Kuny, Terry. 1998. "The Digital Dark Ages? Challenges in the Preservation of Electronic Information." *International Preservation News* 17 (May): 8–13.

Levi, Yaniv. 2008. "Digital Preservation : An Ever-Growing Challenge." *Information Today* 25 (8): 22.

National Library of Australia. 2013. "Digital Preservation Policy 4th Edition." http://www.nla.gov.au/policy-and-planning/digital-preservation-policy.

Rosenzweig, R. 2011."'Scarcity or Abundance?" In *The Future of the Past in the Digital Age*, edited by R. Rosenzweig, 3–27. New York: Columbia UP.

Somers, J. 2017. 'Torching the Modern-Day Library of Alexandria'. *The Atlantic*. April 20. https://www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/

Underhill, A.-M., and Underhill, A. 2016. *A Digital Dark Now? Digital Information Loss at Three Archives in Sweden.* Lund University (Master thesis). Retrieved from http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=8876749&fileOId=8876760

UNESCO, and National Library of Australia. 2003. 'Guidelines for the Preservation of Digital Heritage.' Paris. http://unesdoc.unesco.org/images/0013/001300/130071e.pdf.

Workshop on research challenges in digital archiving and long-term preservation. 2003. "It's about Time: Research Challenges in Digital Archiving and Long-Term Preservation." Washington.

ILLUSTRATIONS

Design Science. 2004. 'MathType: TechNote 64: "The disk is full" or "We can't open" Error Message in Microsoft Word', *Design Science*. https://dessci.com/en/support/mathtype/tsn/tsn64.htm

Dickson, B. 2013. 'Classic tech', *Random perspective*. http://randomperspective.com/comic/34/.

Google. 2012. 'Google Datacenters', *Google*. https://www.google.com/about/datacenters/gallery/#/all.