Emerald Insight

Collection Building

The hidden cost of digitization – things to consider
Trevar D. Riley-Reid

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

The hidden cost of digitization – things to consider

Trevar D. Riley-Reid

Morris Raphael Cohen Library, The City College of New York, New York, USA

Abstract

Purpose – The paper aims to explore some of the hidden costs of these endeavors through an exploration of some the current literature. Despite the perceived ease of creating digital access, the road to digitization is long and fraught with a number of obstacles.

Design/methodology/approach – Some of the key ideas researched are the steps to consider when planning a digitization project, the long-term sustainability of such projects and issues of copyright and ownership. This article will benefit any institution planning to embark upon a future digitization project.

Findings – More research needs to be conducted as the needs of users change and as the technology shifts and changes.

Originality/value – This paper will benefit any institution planning to embark upon a future digitization project.

Introduction

There is no doubt that we are in the midst of a digital revolution and as such, the exponential advancement of technology over the past few years has meant that the way libraries look and operate now, is not going to be what they will look like in the future. Even the concept of a book, once a concrete term that meant a very particular thing, has now taken on multiple meanings. One can read a "book" on a tablet, an e-reader, a computer screen or as a phone app. Libraries and cultural institutions have to contend with growing born-digital assets and need to quickly learn ways to file, sort and store these assets. In addition, there is also a push to take those analog items that are either paper or cultural artifacts and make them accessible digitally. These days, the advantages of having ready access to information are clear; our patrons also expect this from us.

Great strides are being made in terms of technology and the way information can be disseminated to ever wider audiences. Our world is rapidly being converted from analog to digital, making the transfer of information fast, cheap and easy. But there are drawbacks. From e-books to preservation, the long arm of digitization can turn out to be a chokehold and not the warm embrace we have all been led to believe. It appears as though the leap from analog to digital comes at a heavy price – much of it monetary but some of the costs reflect unforeseen socioeconomic and socio-technical issues that have a way of short-circuiting the vast potential of where digital repositories might take us.

The physical boundaries may have been torn away while other boundaries and challenges have taken their place. So a librarian discovers a treasure trove of CD-ROMs which contain a number of historical photos in a storage closet; she/he wants to make these available online but does not have the resources or the time to figure out how to make this a reality. He might feel the pressure from his institution or his constituents to hop on the digital bandwagon. Unfortunately, hopping on the digitization bandwagon too quickly, without taking into account the various pitfalls could result in wasted time and money. This is just one illustration of how we are affected in our everyday dealings with digital assets.

It is important for an institution to analyze its motives in embarking upon a digitization project and in looking at whether mass *digitization*, or simply *preservation* is called for. It is crucial to understand what digitization can mean when compared to preservation and what the implications are when choosing either action. Will your project entail digitization for better access to materials that might otherwise be difficult for users to get to? Do you want to expand your user base to make books or other items available to anyone with a computer and Internet access? This might be cultural artifacts, rare books or other ephemera that an institution wishes to make readily available to a wider audience. Creating a digital repository would mean taking these newly generated files (or gathering already existing, born-digital files) and organizing them in a user-friendly interface. Digitization might be useful in augmenting an institution's prestige by creating demand and highlighting a resource that is unique to that particular institution. Having better access to this digital asset would also increase awareness and build prestige for that particular institution.

This is a huge, constantly evolving, highly complex issue of which this paper is by no means a comprehensive

---

examination. There are simply too many facets of digitization to address in one small space – much more extensive research is needed. For the purposes of this article, the term *digitization* applies to the process and planning involved in actually creating digital files out of books, photos or data. *Preservation* on the other hand, has slightly different motivating factors (of which digitization is an outcome). An institution might have documents or materials in danger of disintegration by time, overuse or other environmental influences. In this case, an institution's incentive for embarking upon a digitization project would be different and possibly more immediate. The urgency in preserving materials can be the prompt to get an institution to move quickly in embarking upon a digitization project, thereby leaving the nuance and details of how users will access these digitally preserved items for a later time. It should be recognized that many issues discussed in this paper are shared among those preserving data that is both born digital and data that is converted to digital. What will be examined here are the steps in planning a digitization project and some of the questions that institutions need to ask themselves before getting started; issues related to long-term sustainability and the fear of obsolescence and what that might entail; and last, some of the issues surrounding copyright laws and their implications for the future. The biggest question is what access really means in terms of the cost.

Planning

The key to successful digitization is to create a comprehensive plan – as much as possible in the beginning stages. The initial step an organization needs to take before embarking upon a digitization project is to employ a practice used in journalism which is to ask about the who, what, where, why and how. Who are the intended users? Who will benefit the most by accessing these digital files now and into the future? Who will be responsible for maintaining the files? What specifically should we digitize? What are the long-range benefits of digitizing? Where will the digital files be kept and accessed? Why should we digitize? How will we obtain all of the necessary resources to effectively maintain the repository into the future? With time and money so scarce these days, a lot of care and thought needs to go into assessing the most benefit using the fewest resources. An internal digitization project will certainly look much different from a project intended for use by the community outside of the library or institution. Those stakeholders in the digital project need to analyze available resources in the present and then need to conceptualize what might be required to sustain the project over the long term. Planning is hard and in this case it is even harder, as the plan needs to include long-range projections for what might be needed in the future. As we all know, sometimes even the best laid plans can backfire, but spending some time thinking about these points and making plans accordingly is an essential step in the digitization process.

Why do libraries want to even consider the long, sometimes arduous, confusing task of creating digital files and/or access to their holdings? There are two key motives for why libraries might want to embark upon a digital project: one is to offer more access and the second is to preserve deteriorating holdings. The impetus for each outcome will naturally take on different characteristics. An institution might want to make

materials open and available that would otherwise be limited to only a handful of people – all one would need is a computer and Internet access as opposed to having to physically go to a library or institution to conduct research. In the case of preservation, there might be a more immediate need to get things digitized due to environmental degradation or the simple wear and tear of usage over time. The purpose of this article is to elucidate the steps involved and the things to consider regarding digitization from a project management standpoint whether one is approaching the process with the goal of obtaining open access or in the case of preservation. The two goals are not mutually exclusive of one another; preservation can also mean better access. In an effort to showcase the cultural heritage of an institution, digitization "raises the profile of [that] institution as users worldwide utilize its collection remotely," (Lopatin, 2006, p. 274). Institutions with rare books want to maintain them for the future *and* make them available to a larger audience. "The best reading, for the largest number, at the least cost", is the core tenet of the American Library Association (1988). If there is information to be discovered, then it must be made available using the fewest resources. Users also need to be able to fully *engage* with the digital material in a number of ways. As Kevin Bradley explains, access is more than just retrieval; it is being able to "interact with, display, or run the digital item in such a way that users can be assured that what they are viewing satisfies their needs," (Lopatin, 2006, p. 274). Hence, a library will have to carefully review the wants and needs of its users, and will need to measure those needs against available resources, and then the institution will be able to provide a more enriching digital experience.

The plan

There are so many issues that need to be considered when planning a digitization initiative; all too often, those involved in the planning fail to take into account the myriad facets of a digitization project. As a result, some libraries and institutions would rather embark on more traditional, tried and true collection development instead of the riskier gamble of a digitization project (Lee, 2001). Even a small project can involve unforeseen circumstances or unaccounted for steps in the process (Moss and Currall, 2004). There just are not ways to fully realize how far and wide a digital project can take an institution, but the following steps, which are not definitive, may offer an organization a rough foundation from which to work:

- determine goals;
- assess the collection;
- analyze work flows;
- identify financing/resources;
- maintain quality control;
- create/maintain metadata;
- educate/train (users and staff); and
- identify Legal/copyright issues.

Determine goals
What are the goals for embarking upon this project? Stakeholders who have a vested interest in the project need to be a part of the goal-making process from the very beginning. Clear, specific reasons or goals for digitizing need to be established by all of those involved. The administration, along

with the necessary departments, needs to be in agreement about what the project will entail. As Lampert and Vaughn (2009) explain, the decision-makers need to buy into the benefits of digitization. Project ideas need to be prioritized and clear. Institutions must consider a variety of facets of the project. Will these digital files be freely available on the Internet or through the institution's Intranet? If certain divisions within the institution want their material digitized, are they willing to provide resources or do they only want the library to do that? It is hard to know the extent of the full scale of the digitizing process without giving serious thought about what and why to digitize. Some key goals of a preservation policy should be to: maintain physical control of the repository; care for the physical media; identify and retrieve the data into the future; and to keep the digital objects "whole and unimpaired [. . .] [so that] all the parts relate to each other," (Dappert and Enders, 2010, p. 158).

Assess the collection
What needs to be digitized? Not only should user needs be carefully examined, assessment should also include going through the organization with a discerning eye to root out any books or materials that might be hidden in dark corners which might benefit from being digitized. Libraries often accumulate so many "hidden gems" over the years amongst its holdings, that it can be easy to lose track of certain things, especially if these items are not part of the regular collection or have been donated and not catalogued.

Analyze work flows
Which department will be responsible for the digitizing project? Will there need to be a new unit or department developed just to handle this and possible future digitization projects? How will this department handle long-term upkeep of the repository? It is important to figure out how departments will work together and how the work flows will operate between these entities. Because of the tech-heavy nature of the digitization process, the IT department is going to have to be a key player in the planning and implementation of any digitization project. Sometimes, communication with IT takes on a different quality with the demands of a digitization project – there needs to be more collaboration and information technology will have to take on more decision-making roles. Is there a dedicated staff member to digitize and maintain the repository? If there are volunteers, who could help in a pinch, who will manage the volunteers? It really helps for there to be a dedicated department similar to the setup at the University of Las Vegas (Lampert and Vaughan, 2009).

Identify financing/resources
Probably the most difficult question to answer is "where will the money come from?" As budgets get streamlined or eliminated altogether, institutions have a hard enough time keeping the lights on without taking on the added responsibility of digitization projects. As in the case of the University of Las Vegas Nevada (Lampert and Vaughan, 2009), many institutions might want to seriously consider grants as an option to making their digitization dreams a reality. At the very least, the plan for a digital repository needs to "consider how to address the financial sustainability of [a] digital preservation program," (Bishoff, 2010, p. 22). Digital

preservation needs to reflect economic realities as well as technical ones. The question must be asked: is it cost-effective in the long run? As more and more digital content is created, it is becoming more apparent that our ability to manage and sustain this data effectively is being called into question. Many institutions are not keeping up and the economic downturn of 2009 has meant that fewer funds have been set aside for digital preservation; even today, institutions are still trying to play catch up financially.

Maintain quality control
It is important that what is being digitized is of good quality. If the items are pictures, are they clear and focused? If the information is a series, are all items in the proper order? How many clicks does a user need to employ to get through the Web site to access the necessary information? These and so many other questions need to be asked regarding the usability and viability of the digital content. Once items have been digitized, an institution must be able to describe and analyze what has been digitized in detail in order for that material to be deemed authentic (Webb *et al.*, 2013). What is the sense in putting in all the time and money if what is being stored is not useful, organized or authentic?

Create/maintain metadata
Another key component in making digital assets available is metadata. A clear metadata policy is important in defining how to manage digital assets into the future to avoid possible losses. According to Dappert and Enders (2010), there are four categories of metadata which are necessary for long-term preservation: administrative, technical, descriptive and structural. Each category is essential in maintaining a comprehensive metadata infrastructure. Having this metadata infrastructure makes the migration to updated software formats more seamless and will make data losses less probable. The likelihood of keeping your data intact will be increased if the metadata plan is clear and is followed for each upgrade.

Educate/Train (Users and staff)
An important thing to consider is who will be responsible for maintaining and uploading new content to the digital repository? Which department and which staff? Staff turnover can prove to be problematic, as anyone new will need to be trained.

Identify legal/copyright issues
In the next section, this paper will briefly examine some of the issues regarding copyright. Questions arise regarding legality and ownership when content is made too available for users.

The rights and wrongs of copyright

The biggest issue since the very beginning of the wave of digitization has been the question of copyright. So much of what we think we know about copyright laws and ownership have been upended with the digital revolution; the old rules and laws do not match with the demands and needs of users. The laws, which were written when books were simply paper editions, do not fully take into account the expectations of the digital user despite the recent updates and amendments to the law. On the contrary, the updates have created even more questions. Ironically enough, American copyright laws

originated in sixteenth century Britain, when Queen Mary Tudor allowed a group of printers in London the rights to print only publications approved exclusively by the monarchy. These printers in turn established the notion that perpetual exclusive rights belonged to the first to print a particular work (Warwick, 2002, p. 236). Early American proponents of copyright sought to promote learning and prevent monopoly. How could a democracy thrive if its citizens did not have access to information? The library would be the great "democratizer" – it was the place that upheld the notion of free and continued access to information while being an arbiter of copyright laws.

With the change to a more digital society, the laws have not kept pace and the relationship between libraries and copyright holders has changed dramatically to a point where the library and other institutions have been viewed as scofflaws:

> The potential for digital initiatives [. . .] to encroach upon the central interest of copyright owners also is recasting libraries as potential infringers [. . .] rather than as agents of a larger public good, (Buttler and Crews, 2002, p. 257).

Libraries and museums are most concerned with the doctrine of *fair use* and the right of *first sale* which are both a part of the copyright laws and apply to physical objects. Fair use addresses these questions:

Q1. Is the use of the work for commercial or nonprofit educational purposes?

Q2. What effect will the use have on the value or market for the copyrighted work?

Q3. Is the work an image or a recording (the answer of which has different implications)?

Things get tricky when applying the notion of fair use when digital copies or files are made from these objects. There are no rights ascribed to these new digital creations. Digitizing or copying a book in its entirety is infringement, so is copying an image, no matter how small, but what about some old newspapers that the library owns? To make its contents available online, the library would have to get permission from the newspaper owner to make that happen. The law is far from being simple, clear and direct, especially when one analyzes the various ways that digital content is used. It is no wonder that these copyright issues are so contentious in our twenty-first century digital world.

The power of *first sale* allows libraries, museums and archives the right to lend materials – reflecting the idea that any person or organization has the right to do anything to that "fixed expression" object like loan it out or sell it. A big problem with digital resources is that many institutions do not own these assets but license them instead, or if they own them, they do not necessarily have permission to digitally copy and electronically distribute the contents as in the newspaper example listed above. There have been a number of revisions to the copyright laws; one, the revision of the Copyright Act of 1976 and the resulting Congressional Commission on New Technological Uses of Copyrighted Works which established a series of guidelines to determine the minimum amount that a person could copy under a series of exemptions while avoiding copyright infringement.

In 1998, Congress passed the Digital Millennium Copyright Act (DMCA) which allowed for a certain amount of digital reproduction and distribution. Specifically in Section 108, certain allowances have been made for libraries to make copies of works for preservation purposes and unpublished archival materials. A lot of time has elapsed since the DMCA was passed by Congress, yet the end result has been "many negative consequences for copyright-dependent industries, costly and drawn-out litigation [. . .] and convoluted legal standards" (Di Palma, 2014, p. 801). A variety of court decisions have resolved some issues but then as some things got worked out, other problems have arisen to take their place.

The law is still vague – how can it be otherwise, when technology has created such different usage demands and these demands keep shifting and changing? Congress has attempted to keep up with the advances in new technologies in regards to the copyright law, the way in which the law moves overall is through gradual change; yet, with each attempt to elucidate the law, other more pointed questions have arisen of which there are no clear-cut answers.

Even when things related to copyright should be apparent, they are not. Take Google for example. Their idea to digitize all of the books in the world (or at least the ones in the public domain) has been met with a great deal of criticism and even a class action suit from a number of authors and publishers. Even when Google charged back by only providing excerpts of books, the company was still sued in 2005. In 2008, a settlement was reached but has since been overturned. In November 2013, Judge Denny Chin ruled in favor of Google and the case is once again under appeal. According to Samuelson (2014), this might actually work in the favor of libraries and cultural institutions, as it would mean an affirmation of previous fair use rulings and would allow these institutions the opportunity to mass digitize their collections which would provide increased access. The line of objectors who oppose Google and their digitization efforts is long and continues to grow. They include the governments of France and Germany who accuse Google of cultural hegemony by inflicting American values and language on their countries; academics and librarians who fear monopoly and price gouging along with the lack of user privacy; and competitors who assert, the deal for Google "to commercialize all in-copyright out-of-print books" (Samuelson, 2009), is a monopoly. Samuelson (2009) posits a main concern about the Google Book Settlement, that once it is finally approved by the courts, there are overwhelming corporate interests that may become involved – the settlement would give Google "the right to sell the corpus to anyone – Rupert Murdoch or China – if it so chooses". Congress is the one who needs to step in and settle these issues, not the courts.

The debate over rights and copyright infringement has led to a degree of instability in terms of delivery of digital content. Some organizations just do not know how to proceed from here. Institutions can be doing more and if they are not being sidetracked by the lack of funding, or delaying delivery because of the lack of technology, they are hesitating because they just do not know what to do to for fear of copyright infringement. In order not to incur the wrath of the law, they do nothing, thus becoming prisoners of their own inertia while the users get nothing. The updates that were supposed to

clarify the situation have only made things worse to some degree with more questions than answers. Compounding the confusion are the rapid advances in technology that have catapulted all of us into this netherworld of disorder and perplexity. We all need to take a step back and try to find a clear path in the chaos.

Conclusions

As we move forward into the future under the rosy glow of technological advancement and ease of access, it is difficult to resist the lure of jumping on the digitization bandwagon. We should not resist completely; we just need to be more cautious with clear goals in mind. By no means exhaustive, this article was an attempt to explore some things to consider when formulating a digitization project, while analyzing various aspects of a repository's long-term sustainability. Embarking on a digitization project is not easy, nor is it clear cut, but by looking closely at the long range goals and making a plan, the project can progress more smoothly.

References

American Library Association (1988), "Mission and priorities", available at: www.ala.org/aboutala/missionpriorities (accessed 17 October 2014).

Bishoff, L. (2010), "Digital preservation plan: ensuring long term access and authenticity of digital collections", *Information Standards Quarterly*, Vol. 22 No. 2, available at: http://dx.doi.org/10.3789/isqv22n2.2010.03

Buttler, D. and Crews, K.D. (2002), "Copyright protection and technological reform of library services", in Lipinski, T.A. (Ed.), *Libraries, Museums, and Archives: Legal Issues and Ethical Challenges in the New Information Era*, Scarecrow Press, Lanham, MD, pp. 257-273.

Dappert, A. and Enders, M. (2010), "Digital preservation metadata standards", *Information Standards Quarterly*, Vol. 22 No. 2, pp. 156-166.

Di Palma, J. (2014), "The digital millenium copyright act and the clash between authors and innovators: the need for a legislative amendment to the safe harbor provisions", *Loyola of Los Angeles Law Review*, Vol. 47 No. 3, pp. 797-830.

Lampert, C.K. and Vaughan, J. (2009), "Success factors and strategic planning: rebuilding an academic library digitization program", *Information Technology and Libraries*, Vol. 28 No. 3, p. 166.

Lee, S.D. (2001), "Digitization: is it worth it?", *Computers in Libraries*, Vol. 21 No. 5, pp. 28-33.

Lopatin, L. (2006), "Library digitization projects, issues and guidelines: a survey of the literature", *Library Hi Tech*, Vol. 24 No. 2, pp. 273-289.

Moss, M. and Currall, J. (2004), "Digitisation: taking stock1", *Journal of the Society of Archivists*, Vol. 25 No. 2, pp. 123-137.

Samuelson, P. (2009), "The Google settlement", *Nation*, Vol. 289 No. 17, pp. 8-10.

Samuelson, P. (2014), "Mass digitization as fair use", *Communications of The ACM*, Vol. 57 No. 3, pp. 20-22.

Warwick, S. (2002), "Copyright for libraries, museums and archives", in Lipinski, T.A. (Ed.), *Libraries, Museums, and Archives: Legal Issues and Ethical Challenges in the New Information Era*, Scarecrow Press, Lanham, MD, pp. 235-255.

Webb, C., Pearson, D. and Koerbin, P. (2013), "'Oh, you wanted us to preserve that!' Statements of preservation intent for the National Library of Australia's digital collections", *D-Lib Magazine*, Vol. 19 Nos 1/2, p. 2.

Further reading

Abruzzi, R. (2015), "Undertaking a digitization project?", *Online Searcher*, Vol. 39 No. 2, pp. 28-32.

Bradley, K. (2007), "Defining digital sustainability", *Library Trends*, Vol. 56 No. 1, pp. 148-163.

Lawrence, G.W., Kehoe, W.R., Rieger, O.Y., Walters, W.H. and Kenney, A.R. (2000), *Risk Management of Digital Information: A File Format Investigation*, Council on Library and Information Resources, Washington, DC.

About the author

Trevar D. Riley-Reid is an Assistant Professor at the Morris Raphael Cohen Library of The City College of New York located in upper Manhattan. As Access Services Librarian, she is chief of the Circulation Division and manages a staff of five full-time and 12 part-time workers. In addition to being a former high school English teacher and concentrating her research on evolving student needs in higher education, her research also focuses on digitization and the use of technology in libraries. A fierce advocate for the advancement of libraries, she firmly believes that despite the proliferation of information, librarians will continue to be a necessary part of aiding users in finding the accurate, pertinent, reliable information that they will need into the future. Trevar D. Riley-Reid can be contacted at: trileyreid@ccny.cuny.edu

This article has been cited by:

1. Barbara CostelloAcademic Libraries in Partnership with the Government Publishing Office: 87-110. [CrossRef]