Long-term accessibility of e-books: challenges, obstacles, responsibilities

Maja Krtalić, mcujic@ffos.hr

Damir Hasenay, dhasenay@ffos.hr

Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Osijek

Abstract

The traditional life cycle of books has gone through many changes in the digital environment - enough to start questioning the effects of such changes on the process of creating, publishing, distributing, reading and preserving books. This paper focuses on issues concerning the preservation and archiving of published authors' works in the digital environment for the purpose of their long-term accessibility. The aim of this paper is to give an overview of relevant legal, technical, societal and organisational issues from which challenges, obstacles and responsibilities in ensuring long-term accessibility of e-books arise. Issues related to authorship, editions, changes in content, copyright and digital rights management, selection criteria for preservation, and preservation responsibility will be discussed, specifically focusing on libraries' and publishers' roles in this process. The paper is based on a literature review and content analysis aiming to answer two basic research questions: What specific characteristics of e-books influence their preservation possibilities? Who is responsible for long-term accessibility of e-books?

KEYWORDS: e-books, digital preservation, digital curation, preservation management, heritage collections.

Introduction

The concept of heritage and its preservation has always been a complex one. Making a decision on what and how to preserve depends on the content of heritage, the form of expression of that content, the medium (platform) that carries the content and the context in which heritage content is created, presented and used. The term 'heritage collection' in today's information institutions such as libraries, archives and museums includes the conglomerate of 'traditional' (physical) collections, all sorts of digitised material, and all sorts of born-digital material in all its manifestations where e-books are one of them.

Such hybrid collections are inevitably intertwined with a range of digital services in and for the digital environment. E-books represent one part of this heritage puzzle.

The task of preserving published books as printed heritage for posterity traditionally lied in the domain of libraries. The preservation of cultural heritage has always been an integral part of heritage institutions management and one of its key functions. Even if other parts of the system failed, libraries, archives and museums could take the task of ensuring long-term accessibility to heritage material. With digital documents and especially e-books, the responsibility is divided and therefore not clear.

A large and significant body of professional literature has been written over the years on preserving digital documents of various form and content, from a comprehensive overview of issues (Harvey 2010, Deegan and Tanner 2006, Beagrie and Jones 2008, Cloonan 2014) to specific topics. The Digital Preservation Coalition issues Technology Watch Reports[1] that report on developments in technology, standards and tools which are critical to digital preservation activities. They are focused on the type of the digital content, such as e-books (Kirchhoff and Morrissey 2014), e-journals (Beagrie 2013), moving pictures and sound (Wright 2012), and on the issues that affect long-term accessibility of content, such as standards and models (Lavoie 2014), metadata (Gartner and Lavoie 2013), intellectual property rights (Charlesworth 2012), file formats (Todd 2009), methods and techniques (Leighton 2013), etc. Although digital preservation is a rapidly changing area with many new emerging trends, some key challenges remain to be met. That is why some professional literature is still relevant even if it was published more than a decade ago. For example, Conway raises the question of the concept of preservation in the digital world (1996) and the transformation of preservation principles (2000) that are still relevant questions to be discussed.

Several approaches and standpoints can be identified in the professional literature on digital preservation:

- General ('philosophical') approach discusses movement of preservation focus "from the eternal carrier to the eternal file[2]"

- Terminological approach sorts out the complexity and number of definitions and terms such as content archiving / digital preservation / digital curation / digital stewardship, etc.

- Organisational approach relies on identifying concrete preservation activities in work processes of specific types of institutions.

---

1   Available at: http://www.dpconline.org/advice/technology-watch-reports
2   Phrase used by Schiller 2014. http://blogs.loc.gov/digitalpreservation/2014/11/audio-for-eternity-schuller-and-hafner-look-back-at-25-years-of-change

- Typological approach describes the preservation of different types of data, or types of content (for example textual data, multimedia, scientific data, etc.). One of the most interesting ones is certainly the so-called 'big data'. Big data is usually connected to science and research but lately also to commercial sector.

- Institutional approach implies perspectives on preserving e-books that are coming from information institutions (libraries, archives and museum), publishing community, IT sector or independent preservation institutions. Each branch has their own views not just on how e-books should be preserved, but also on who should actually do it.

- A project approach is oriented towards finding solutions to specific problems.

Preservation issues often do not come in the limelight of professional literature until other more 'urgent' aspects of the topic are solved. That is why this paper aims to contribute to exploring e-books preservation based on the application of the preservation management model to e-book preservation issues. The scope of the term e-book is complex and wide, and there are numerous definitions of the term. In this paper, a wider meaning of the term e-book is implied, similar to the following definition:

> digital file containing a body of text and images suitable for distributing electronically and displaying on-screen in a manner similar to a printed book. E-books can be created by converting a printer's source files to formats optimized for easy downloading and on-screen reading, or they can be drawn from a database or a set of text files that were not created solely for print. (Attwell 2013).

Managing e-books preservation

Specific aspects of preserving e-books will be considered in the general preservation-management model (PMM) (Krtalić & Hasenay 2012) that is comprised of five key components: strategic and theoretical, economic and legal, educational, technical and operational, cultural and social components. The model is based on frameworks in which preservation is conducted (national, institutional, social, and cultural) and on the resources necessary for implementing preservation activities (financial and/or human, including the necessary knowledge and competence).

The guiding principle is that preservation is a complex process that must be strategically planned by following goals and mission set on a national and institutional level. Such plans should be grounded in theoretical knowledge

and achievements in the field of preservation. Furthermore, the preservation process must comply with the economic and legal framework in which the institution functions, but with an emphasis on overcoming limitations that such frameworks often pose. Special attention should be paid to the educational element of managing preservation, given that education on different levels (from the formal education of information or other specialists to the training of staff and users) is believed to contribute to the efficacy of preservation. Since the material and content of heritage items are the focus of preservation activities, an important part of preservation management is handling and safeguarding collections and knowing their conditions and needs. Finally, it is assumed that preservation should result in preserving information (and access to it) that is significant and usable in different areas of cultural and social life. Bearing in mind these basic premises, this preservation-management model comprises five key components, as already mentioned: strategic and theoretical, economic and legal, educational, technical and operational, cultural and social components (Krtalić and Hasenay 2012, 363-364).

Different issues regarding successful preservation management are brought together in these components, such as policies and strategies, financial issues, legal regulations, knowledge and competencies, preservation methods and techniques, user needs, and, lastly, the cultural and social impact of preservation. Among numerous issues in these components, several key issues that reflect the complexity of the topic, specially focusing on cultural and social components will be presented.

Strategic and theoretical components

The strategic and theoretical components are focused on the strategic planning of the preservation process based on relevant theoretical and practical knowledge and research. Challenges in e-book preservation in the strategic and theoretical components of the PMM arise from aligning different theoretical perspectives and approaches that appear in different institutional contexts. One of the interesting challenges is certainly identifying what issues in e-book preservation should be researched. Further on, applying research findings to strategic planning, policies and strategies and creating strategic partnerships represents a key challenge.

Economic and legal component

The economic and legal components are focused on issues that surround and influence the preservation process and arise from legal documents and (un)available financial resources. Key e-book preservation challenges within this component are connected to copyright and digital rights management (DRM) restrictions because in order to preserve an e-book, preservationists (no matter in which institution they work) need to be able to modify and migrate

e-book content without any limitations. Licensing agreements need to be negotiated with that aim in mind (to ensure the stability of leased content). E-books published in proprietary formats, especially those combined with DRM present a special category of preservation risk.

Legal deposit models in some countries can be challenging for e-book preservation. There are various legal deposit models in different countries. Ensuring that national e-book production of a country will be deposited in deposit institutions, usually libraries, will guarantee a professional approach to preservation as well as clear responsibility for long-term accessibility.

Educational component

The educational component of the PMM focuses on knowledge, skills and competences necessary for conducting preservation activities, but also for being able to predict and conceptualize future developments in preservation issues, which might be even more important. Challenges in this model are oriented towards identifying targeted audience that needs to be educated about e-book preservation issues, and designing education materials and training aids. It is evident that, at least for now, education about the necessity of e-book preservation should include the entire preservation community (publishers, aggregators, libraries, but also authors and end users).

Technical and operational components

The technical and operational components of the PMM imply technology that is needed or that influences preservation activities, as well as everyday operational tasks that preservationists carry out. This includes: preserving technology that is used for reading e-books today; preserving formats in which e-books are published so that they are readable and functional, or migrating the content on time; and applying the appropriate set of preservation techniques and methods.

Ensuring that preservation process is successful depends on several factors related to content level of e-book, for example the relations between fixed layout and reflowable text, or the relations between textual content and bibliographical and structural data (such as headings and sections) in, for example, a wrapper format. Content updates (editions) complicate things additionally because they raise the question of selection criteria between many available editions.

Challenges finally need to be met in diverse institutional management frameworks of those institutions that took the preservation responsibility, whether they are library, publisher, consortial preservation institution or somebody else.

Cultural and social components

Cultural and social issues regarding preservation are the most interesting and the most challenging ones, mainly because they need to be recognized and identified early enough to be able to willingly make a choice what and how to preserve. Authorial work in the digital environment raises the question of vast amount of published texts (the term published in this context implies a wider meaning of publically available content created by an author and made available through different means of traditional publishing, self-publishing, blogs, posts etc.). This raises the question of the heritage value of the "published" intellectual content. For example, Jose Saramago's *The Notebook* is a book published based on his blog entries in 2008 and 2009. As one reviewer states "It is not, as its author admits, what is usually considered to be a "real blog". It doesn't contain any links and "I don't have a dialogue with my readers" and "don't interact with the rest of the blogosphere". The Notebook is therefore best thought of as just that, a series of daily jottings which happens to have been first published on the internet, but which might just as well have appeared as a daily newspaper column." [3] There are thousands of similar cases of anonymous authors. Should their authorial works be preserved willingly and in a planned manner, or should they be left to accidental disappearance or survival in the digital world? And who has the right or the obligation to make this decision?

That finally raises the issue of who constitutes the 'preservation community'. A way towards 'preservation ready' e-heritage is the only efficient way here, and that means education on how to create preservation ready e-book for all involved in the process.

Conclusion

The aim of this paper was to give an overview of major e-book preservation issues focusing on theoretical and conceptual level. The basic question remains the question of responsibility. Who is responsible for long-term accessibility of e-books? As it happened many times in the past, preservation comes into the limelight usually after other issues are solved. Printed books, even digitized ones, could wait and survive long-term, often by pure chance, mainly due to the nature of material (if stored under the right conditions) and due to its quantities (many copies of the same item held in many different libraries). E-books, as they are acquired and managed by libraries today, do not have those same chances for accidental long-term survival. No doubt that some e-books will be accessible for the long-term, usually those whose content is explicitly valuable, whose authors or publishers are "big enough" to ensure survival in the digital environment. The problem starts with smaller national markets that are not yet organised enough to ensure preservation of its digital cultural production. Problem continues with

3   Available at: http://conversationalreading.com/new-book-the-notebook-by-jose-saramago/

numerous self-published authors that can easily become invisible. We don't know what exactly we have lost of digital content so far and what of that content had potential heritage value. As John Feather said, all that we inherited from the past came to us because it was preserved. Therefore, in order to provide long-term access to e-books produced today, we need a clear understanding of the responsibilities and strategic partnerships between all parties involved in the process of creating, publishing, distributing and safe-keeping e-books, bearing in mind the complexity of the digital heritage concept.

In 1981, P. Darling wrote about the consequences of giving up on preservation responsibility.

> But suppose we lose our nerve, discouraged by the enormity of the task, diverted by the competing challenges of automation, networking, resource sharing, corporate administration, and belt-tightening fiscal management? Should that happen, the library, and the world, will look a bit different twenty-five years from now. For one thing, we won't have a space problem anymore: retrospective collections will have decayed so seriously through natural deterioration and misuse-the paper crumbled, photographic images vanished, magnetic tape charges jumbled-that we will have discarded most of them, maybe carting them off for landfill or perhaps selling them for recycling as filter material in air pollution masks. Many collections will have been destroyed by fires caused by spontaneous combustion of densely packed paper fragments in large stack areas. And theft will be a major problem as the increasing scarcity of nineteenth- and twentieth-century books and photographs leads to a collectors craze. (Darling 1981, 181).

Fortunately, this has not happened so far, but these words are alarming enough to force us to think about the digital heritage scenario for the next 25 years.

References

Attwell, A. 2013. "E-book." *Encyclopaedia Britannica*. Accessed December 1, 2014. http://www.britannica.com/EBchecked/topic/1235205/e-book

Beagrie, N., and M. Jones. 2008. *Preservation Management of Digital Materials: The Handbook*. Accessed December 1, 2014. www.dpconline.org/advice/preservationhandbook/

Beagrie, N. 2013. *Preservation, trust and continuing access for e-journals.* DPC Technology Watch Report. Accessed December 1, 2014. http://dx.doi.org/10.7207/twr13-04

Charlesworth, A. 2012. *Intellectual property rights and preservation*. DPC Technology Watch Report. Accessed December 1, 2014. http://dx.doi.org/10.7207/twr12-02

Cloonan, M. V. 2014. *Preserving our heritage: Perspectives from antiquity to the digital age*. Chicago: ALA Neal-Schuman.

Conway, P. 1996. *Preservation in the Digital World*. Accessed December 1, 2014. http://www.clir.org/pubs/reports/conway2/index.html

Conway, P. 2000. "Transformation of preservation principles." In *Handbook for digital projects: a management tool for preservation and access*, edited by Maxine K. Sitts, 22-27. Andover, Massachusetts: Northeast Document Conservation Center. Accessed December 1, 2014. https://www.nedcc.org/assets/media/documents/dman.pdf

Darling, P. 1981. "Will anything be left? New responses to the preservation challenge." *Wilson Library Bulletin* 56, 3: 177-181

Deegan, M. and S. Tanner, eds. 2006. *Digital Preservation*. Facet Publishing.

Gartner, R. and B. Lavoie. 2013. *Preservation metadata*. 2nd ed. DPC Technology Watch Report. Accessed December 1, 2014. http://dx.doi.org/10.7207/twr13-03

Harvey, R. 2010. *Digital curation: A How-To-Do-It Manual*. New York, NY: Neal-Schuman Publishers.

Kirchhoff, A. and S. Morrissey. 2014. *Preserving eBooks*. DPC Technology Watch Report. Accessed December 1, 2014. http://dx.doi.org/10.7207/twr14-01

Krtalić, M. and D. Hasenay. 2012. "Exploring a framework for comprehensive and successful preservation management in libraries." *Journal of Documentation* 68, 3: 353-377. Accessed December 1, 2014. http://www.emeraldinsight.com/journals.htm?issn=0022-0418&volume=68&issue=3&articleid=17030676&show=pdf

Lavoie, B. 2014. *The Open Archival Information System (OAIS) Reference Model: Introductory guide*. 2nd ed. DPC Technology Watch Report. Accessed December 1, 2014. http://dx.doi.org/10.7207/twr14-02

Leighton John, J. 2012. *Digital forensics and preservation*. DPC Technology Watch Report. Accessed December 1, 2014. http://dx.doi.org/10.7207/twr12-03

Prom, C. J. 2011. *Preserving email*. DPC Technology Watch Report. Accessed December 1, 2014. http://dx.doi.org/10.7207/twr11-01

Todd, M. 2009. *File formats for Preservation*. DPC Technology Watch Report. Accessed December 1, 2014. http://www.dpconline.org/component/docman/doc_download/375-file-formats-for-preservation

Wright, R. 2012. *Preserving moving pictures and sound*. DPC Technology Watch Report. Accessed December 1, 2014. http://dx.doi.org/10.7207/twr12-01

Sažetak

Dugoročna dostupnost e-knjiga: izazovi, prepreke, odgovornost

U tradicionalnom životnom ciklusu knjige dogodile su se brojne promjene uvjetovane izmještanjem procesa proizvodnje i uporabe knjige u digitalno okruženje. Logično je postaviti pitanje kako su se takve promjene odrazile na cjelokupni proces pisanja, objavljivanja, raspačavanja, čitanja, ali i dugoročnog očuvanja knjiga. Ovaj je rad usmjeren na razmatranje problematike vezane uz zaštitu i arhiviranje objavljenih autorskih radova u digitalnom okruženju u svrhu njihove dugoročne dostupnosti. Cilj je rada dati pregled relevantnih pravnih, tehničkih, društvenih i organizacijskih pitanja iz kojih proizlaze spomenuti izazovi, prepreke i odgovornosti vezane za dugoročnu zaštitu e-knjiga. Raspravit će se pitanja autorstva, izdanja, promjena i autentičnosti sadržaja, autorskih prava, kriterija odabira građe za zaštitu te podjele odgovornosti, posebno se osvrćući na ulogu nakladnika i knjižničara u tom procesu. Rad se oslanja na pregled literature u danom području te na teorijski model upravljanja zaštitom koji se sastoji od pet aspekata, a u ovome radu služi kao okvir za sustavno razmatranje problematike. Dva temeljna pitanja na koja se želi odgovoriti ovim radom su sljedeća: Koje specifične karakteristike e-knjiga utječu na mogućnosti njihove zaštite? Tko je odgovoran za dugoročnu dostupnost e-knjiga?

KLJUČNE RIJEČI: e-knjige, zaštita digitalnih podataka, upravljanje zaštitom, baštinske zbirke.