Keep it together – presentation of born digital and digitized collections to various user segments

Tuula Pääkkönen, Information Systems Specialist

Centre for Preservation and Digitisation, National Library of Finland, Mikkeli, Finland

tuula.paakkonen@helsinki.fi

Abstract. News content has great appeal to the end-users despite the format they have been originally created. However the born digital news media will cause some changes both for the whole lifecycle from the creation of the collections to the presentation layers for the various user segments. End-users desire the content reliably, with features, which give opportunities to participation and enable them to create their own information based on own interests.

This paper aims to go through some options on how the born digital transition can be managed in creation of digital collections from various angles: process, data formats and presentation and how the existing systems can be changed towards that end. The new formats also effect the consumer of these services so some options of serving the user segments are gone through.

The offering of usable user interface for the consumer of the digitized collections is a complicated task, as both the current usage modes need to be fulfilled, but also to prepare for new needs and continuous development as new requirements arise. Some existing processes and practices will need adjustments as the born digital news have different original formats, instead of the traditional paper. New audiences have become more and more accustomed to the digital formats, applications and constant accessibility, which require that also digital collections need to be there where the users are. As a case study the current presentation system of the NLF (http://digi.kansalliskirjasto.fi service) and its current strategies are examined in the case of the electronic deposition. Born digital materials create new possibilities to public-private partnerships as news organisations can also benefit from the new emerging technologies and utilizing those in digital collections, regardless of their creation method.

Keywords: historical newspaper collections, born-digital, presentation systems, content systems

1. Introduction

Media industry is currently facing the change towards digitisation, as the end-users get their news more and more in digital format via internet and mobile applications instead of traditional paper. In a Finnish Newspaper Association follow-up of statistics from 2011-2014 it can be seen that while reading of printed newspaper has gone down, all digital mediums: computer, mobile phone and tablets or e-readers the tide is different, (Sanomalehtien lukeminen, 2014). Therefore the trend that also born-digital news are served to the end-users directly via consumer-friendly medium, e.g. mobiles or websites (Tomas, 2013).

The National Library of Finland (NLF) obtains all of the newspapers, magazines and ephemera from the publishers due to the Act on Collecting an Preserving Cultural materials, which was originally set on the year 1707. The strategic targets of NLF include taking care of permanent preservation and promoting usage of cultural heritage of Finland from researchers to the general public. The earlier digitisation efforts have covered newspapers from years 1771-1910 (Bremer-Laamanen, 2005). Since then the public-private partnerships together with the Finnish news media companies have taken the digitisation forwards with selected newspapers and journals. There are already a few newspapers, who have all the newspapers digitized from their early history to the present day.

In Finland, everyone is able to view the copyright-free digitized newspapers until 1910 freely from http://digi.kansalliskirjasto.fi –web service. Material digitized further can be seen at a few special cultural heritage work stations across Finland, namely at the National Library of Finland, Turku, Jyväskylä, Åbo Akademi, Oulu and University of Eastern Finland University Libraries, (Legal deposit libraries, 2015). Currently the same user interface can be used by general public via internet to view and read the materials until 1910, but also in free deposit libraries the new material can be seen up until today, depending how long material has been digitized.

2. Digital Collections of the National Library of Finland

The digital collections of the National Library of Finland have been created over the years based on the digitisation policy, customer demands, NLF strategy and annual goals, (Digitisation policy, 2010). The newspapers, journals are just one part of the material, as the digitized collections also include ephemera, maps, music, scores and naturally books, to name a few material types. Also certain selected websites are harvested to get the changes of few key websites throughout the day. The quality required, is described as:

> *When deposited, the materials must be complete; free from technical defects; true to what was made available to the general public; and not protected by DRM or other methods.*

> *When deposited, the material must be accompanied, in a digital form, with descriptive metadata as well as with information about copyrights pertaining to it.*

, Keskitalo 2011.)

All of these digital collections have one search portal finna.fi and content services for presentation, which focus to the special qualities of various material types and serving various end-user needs.

2.1 How to define born-digital materials?

In this paper the born-digital materials mean materials created by a computer, which are then either used directly or transformed further to other electronic formats. In a way the definition "created by a word processor" (Jackson, 2003) is also quite suitable as all the phases in creation of newspapers are nowadays digitalized – stories of journalists come electronically, layout and graphics are done digitally. The transferring of the material to the printing houses and partially to the customers is increasingly done electronically - so also the library systems and presentation systems could support this fully digital process. Then born digital materials can be books, e-mails, hard drive contents or even 3D-simulated dance choreographies (Whatley & Warney, 2009), which just shows that the digital world is as versatile as the "real" world.

2.2 Electronical deposition

The Centre for Preservation and Digitisation of NLF had a project from 2011-2014 for setting up a methods, tools, and practices for getting the processes and tools ready for the electronical deposition. This is both for the benefit of the publisher, to enable providing the same print-quality final newspaper towards NLF, conveniently at the same point, when it is provided to the printing house. This enables streamlined workflow as the electronical deposition includes all the necessary author metadata, layout structure and newspaper/journal content. This helps in preventing errors in cataloguing, provides more metadata of the content of the newspaper. All of these digital processes enable getting the material visible for the public in a straightforward way.

The high-level process of the electronical deposition of the NLF can be described as:

1. newspaper delivers print-quality material (content, metadata) to NLF
2. content package is checked for long-term preservation quality (e.g. missing pages, attachments, etc.)
3. content is printed to a microfilm for long-term preservation and storage
4. content is imported to the presentation system

During the partnerships negotiations with media houses it will be agreed on how the electronical deposition is done, and how the material from the newspaper is delivered to the NLF. Quite common is to use e.g. ftp server, where newspaper can deliver the package, basically at the same point it is delivered to the printing house. If there are any errors in the digital version, then those are checked and reported back and fixed before any further processing is done.

Bringing of material to the preservation system can be done after suitable import package has been automatically created from the print-PDF version. At this point the metadata is stored and the text of the pages is indexed and put to the content database and page images to storage disks. This enables showing of the newspaper or journal in the presentation layer to enable both browsing and content-based searchers

of the material. The page text is indexed and so either with full text search or fuzzy search the end-users can find the content they need.

3. One presentation layer for newspapers and journals?

Based on our current experiences, we feel that having a single presentation layer for both historical and "current" newspapers is very beneficial. Our approach has some similarity to the British Library case, where the newspaper collection consists of digitized printed newspapers, e-editions and even harvested web-based newspapers (Day, MacDonald et al., 2014). This brings up a question, that whether instead of talking about the different ways how the digital material is created, maybe the discussion in future should transform more to the thinking of storing of the news content itself – regardless of how the material is originally delivered to the public.

The aspects for acquiring and retaining born-digital materials are cultural, historical, scientific, economic, but then one should not forget the existing of future use cases, (Woods, Lee, Garfinkel, 2011). All of these require commitment from the institution and responsiveness both to the digital partnerships and towards the end-users. The important stakeholders are the end-users, library itself, news media and the future generations of digital heritage via the long-term preservation. In the next chapters, the view point of these stakeholders to the concept of having one presentation systems are gone through within the context of the newspaper material of the NLF.

3.1 End-user benefits

The current users of digitized collections are from very different age groups, regions of Finland and have naturally very different interests. From the recently introduced crowdsourcing features 'Digitalkoot', it has been interesting to see, that few of the most active users have collected article collections from all walks of life: ranging from historical events, early days of automobiles, to dance schools and to the cross words and quizzes. These end-users have used the digitized collections for years and are familiar with the content and the search capabilities. The addition of the new crowdsourcing capabilities has been one reason, which has pulled in more users (NLF statistics, 2014). This has also revealed for the library side the different interests to the content people have. It has also given insight of what kind of content is most appealing to the users, which can be interesting also for current news media organizations, as digital archives of own content is common service, which the newspapers offer in some cases to their digital subscribers.

Therefore, having the same interface also for the newer materials makes the presentation system more approachable. The existing, returning users don't have to learn "yet a new system" and can just get familiar with the new features in the existing one. Currently the digital clipping creation can be done for materials predating 1910, but it and other functionalities are already existing and it might enable opportunities to extend those to new kinds of materials in the future if so desired.

Also for the new users it is useful to have one content system – if they start usage with the copyright-free materials at home, then also at the free deposit libraries, where they can access newer material, the same presentation system is familiar. However, the so called digital natives want to use the library services in a new way, bringing new demands and requiring all content in digital format from anywhere, and utilizing more and more smartphones and tablets (Tomas, 2013). One important thing also mentioned by Bernstein (2006), is the fact that the next generation users do not wait – they want interactivity and quick and easy

access. Therefore enabling users to participate, giving them opportunities to enrich the digital history, is one way to obtain young(er) user segments towards the digital collections.

3.2 Library benefits

Naturally, having a single presentation layer for the content also has many benefits for the library itself – there is need only to plan and specify one system, which also creates less need for the development and maintenance resources. Once the roles and access visibilities for the user groups have been set up, those mechanisms can be directly applied when new newspapers arrive either via electronic deposition or digitisation.

In the case of the digital-first/digital-exclusive books the some libraries have reported benefits rising from cost effectiveness, environmentally friendly publishing and providing instantaneous access to the patrons, (Eaton 2009). The same benefits are quite applicable also for the newspapers, as with digitized processes, the existing limited resources can be used in better ways and it can enable to achieve more with the limited resources and time.

3.3 News media benefits

Via the electronic deposition agreements the news media organization also gets easier way to get own internal access to the digitized newspapers, as local archives of news media organizations can be fragile and partial. The same presentation layer also makes it easier for them to find and locate the old material and maybe even utilize it in creation of new journalistic content. Traditionally the newspaper digitizations have been made from the microfilms, meaning that the newspaper pages can have greyish background. With electronic deposition coming from print-quality digital materials the full colour page images can be utilized via the presentation layer directly. This is also cost-effective for the media house, as the digitisation is slower when it is done in colour.

3.4 Long-term preservation/access

What we have seen also beneficial in the case of the National Library of Finland, is to incorporate the technical implementation of the long-term preservation to the post-processing of material. The National Digital Library has specified to use Metadata Encoding & Transmission Standard (METS) as a collective format to incorporate both access and preservation page images as well as the recognized text from the images[1]. The Analyzed Layout and Text Object (ALTO) XML[2] is used then for containing the words and physical layout positions of the pages. Because it is required that long-term preservation is done for both born digital and traditional newspapers, (Digital preservation, 2012), then having support in the digitisation process saves time and resources in the long run.

---

[1] http://www.loc.gov/standards/mets/
[2] http://www.loc.gov/standards/alto/about.php

4. Conclusion

Naturally, having a single content presentation system is not a silver bullet either. There comes challenges of how to define the requirements of the system – which features it should contain, and who are the users and how to serve them best. Regardless of the system, the whole digitisation workflow and further development of the system anyhow requires effort and investment in resources.

Single system is easier to learn, but customizing it to the needs for different material types can be considerable task. But also creating specific, customized solution requires effort, but from bit different angle. In a way this is a wicked problem – the best solution can only be found by experimenting and open communication with different stakeholders. The library context and content selection has significant impact of what is the best solution, and naturally over the years the situation can evolve as the understanding of the user needs builds up.

At this moment, it seems that in order to serve different user segments better, it could be useful to keep all different digitized contents in one presentation system, at least for the newspapers, journals and ephemera. This would mean enhancing the existing systems with more content of those types, so that the end-users can be served with the same features across material types. Various user segments benefit when they have a generic user interface, which requires less learning and evolves over time with new features. Libraries can benefit by utilizing the existing back-end infrastructure and do less material-specific customizations for the information systems, bringing therefore cost-efficiency to the internal operations of the library. There is still lots of work ahead with any presentation system, so that it is suitable for all different user segments, which makes the times interesting for all of the parties.

References

Bernstein, J. 2006. 'Born Digital, Not Yesterday: Next-generation Web User Seeks Interactivity', *Econtent*, 29, 4, p. 14, Education Research Complete, EBSCO*host*, viewed 22 March 2015.

Bremer-Laamanen, Majlis. 2005. Connecting to the past – newspaper digitisation in the Nordic Countries. World Library and Information Congress: 71th IFLA General Conference and Council, "Libraries - A voyage of discovery" , August 14th - 18th 2005, Oslo, Norway. http://archive.ifla.org/IV/ifla71/papers/019e-Bremer-Laamanen.pdf

Cocciolo, Anthony. 2014. "Challenges to born-digital institutional archiving: the case of a New York art museum", Records Management Journal, Vol. 24 Iss 3 pp. 238 – 250

Digital Preservation Implementation Plan, 2012.
http://www.kdk.fi/images/tiedostot/digital%20preservation%20implementation%20plan.pdf , page checked 1.4.2015.

The digitisation policy of the National Library of Finland, http://urn.fi/URN:NBN:fi-fe201401151119 , 2010.

Day, Michael and MacDonald Ann and Kimura, Akiko and Pennock, Maureen. 2014. Implementing digital preservation strategy: developing content collection profiles at the British library. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries* (JCDL '14). IEEE Press, Piscataway, NJ, USA, 21-24.

Eaton, Lance. 2009. "BOOKS BORN DIGITAL." *Library Journal* 134, no. 9 (May 15, 2009): 26-28. *Academic Search Complete*, EBSCO*host* (accessed March 22, 2015).

Goldman, Ben. 2011. Bridging the Gap: Taking Practical Steps Toward Managing Born-Digital Collections in Manuscript Repositories *RBM Spring 2011 12:11-24*

History of the Finnish Legal Deposit Legislation
http://www.nationallibrary.fi/publishers/deposit/legislation_history.html , page checked 5.2.2015

Honkela, Timo. 2014, *Digital humanities: future directions and opportunities for libraries*, Bibliotheca Baltica, 12th International Symposium, October 9-10, 2014, Stockholm, Sweden.

Jackson, Larry S.. 2003. Preserving state government web publications: first-year experiences. In *Proceedings of the 2003 annual national conference on Digital government research* (dg.o '03). Digital Government Society of North America 1-6

Keough, Brian and Wolfe, Mark. 2012. Moving the Archivist Closer to the Creator: Implementing Integrated Archival Policies for Born Digital Photography at Colleges and Universities, Journal of Archival Organization, 10:1, 69-83, DOI:10.1080/15332748.2012.681266

Keskitalo, Esa-Pekka. 2011. Web Archiving in Finland. Memorandum for the members of CDNL, http://urn.fi/URN:NBN:fi-fe201101311162

Legal deposit libraries, http://www.nationallibrary.fi/publishers/legaldeposit.html#Libraries , page checked 5.2.2015

Sanomalehtien lukeminen eri välineillä. 2014.
http://www.sanomalehdet.fi/sanomalehtitieto/median_kaytto/sanomalehtien_lukeminen_eri_valineilla

Seadle, Michael. 2012. "Archiving in the networked world: authenticity and integrity", Library Hi Tech, Vol.30 Iss 3 pp. 545 – 552, http://dx.doi.org/10.1108/07378831211266654

Suominen, J. and Sivula, A. 2013. Gaming legacy? Four approaches to the relation between cultural heritage and digital technology. ACM J. Comput. Cult. Herit. 6, 3, Article 12 (July 2013), 18 pages. DOI: http://dx.doi.org/10.1145/2499931.2499933

Statistics of NLF, http://www.nationallibrary.fi/infoe/statistics.html , page checked 1.4.2015.

Tomas, K. 2013. Virtual Reality: Why Magazines Should Adopt a Mobile-First Publishing Strategy. *Publishing Research Quarterly*, *29*(4), 301-317. doi:10.1007/s12109-013-9330-7

Whatley, Sarah and Varney, Ross. 2009. Born digital; dance in the digital age, International Journal of Performance Arts and Digital Media, 5:1, 51-63.

Woods, Kam and Lee, Christopher A. and Garfinkel Simson. 2011. Extending digital repository architectures to support disk image preservation and access. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries* (JCDL '11). ACM, New York, NY, USA, 57-66. DOI=10.1145/1998076.1998088 http://doi.acm.org/10.1145/1998076.1998088