

Paper still persists: providing print-on-demand options in a web world

Brenda Chawner
School of Information
Management
Victoria University of
Wellington

Donald Gordon
School of Mathematics,
Statistics, and Computer
Science
Victoria University of
Wellington

Glen Walker
Christchurch College of
Education Library
Christchurch

Introduction

In 1998, Walt Crawford used the phrase "paper persists" in an article arguing that physical collections (of books) will continue to be important (Crawford 1999). As the information environment becomes increasingly web-based, there is still a need to provide users with printable versions of information. People print digital information for a variety of reasons, including needing a permanent record (of an online transaction, for example), or for easier reading and better understanding. This paper begins with an overview of factors that affect readability and comprehension of screen- and print-based documents, followed by a description of the Library and Information Association of New Zealand Aotearoa's recent experience in moving from a print to an electronic newsletter. Current techniques for providing print-on-demand options for users of web-based information are then compared. The paper concludes with a description of three projects to provide printable versions of web pages, all based on XML.

Readability and comprehension

Reading text from a computer screen and reading text on paper are fundamentally different. Most of us who have grown up in western societies are introduced to books at an early age. We quickly become familiar with the norms of using the codex format, such as reading from left to right, turning pages, looking for a table of contents at the front of the book and an index at the back, identifying individual pages by their page number, etc. Reading from a computer screen (or a handheld device like a PDA or mobile phone), is more like reading from a scroll, with only a single section available for reading at a given time, while the rest of the document is obscured. In an online reading environment, different elements (such as tables or illustrations) may be on a separate page, or may be presented as thumbnails, requiring a click to be presented in a larger, more readable size. This is likely to be due to the need to use the limited screen real estate efficiently.

Dyson (2004: 377) notes that while there has been extensive research into the legibility of print on paper, there has been little formal research into reading on screen. Factors that have been found to affect readability of text include font type and size, line length, and number of columns. For hard copy, readability is improved by using a justified serif font of 9, 10, 11, or 12 point, depending on line length. For screens, a ragged right sans serif font is generally said to be more readable, and it should be a minimum of 12 points.

There is some evidence to suggest that reading comprehension and behavior differs between print and online publications. Rouet and others (2003) studied differences in reading comprehension between print- and screen-based information by asking journalism students to read the same report in a Brazilian science magazine as either hard-copy or online version and then answer questions about the content. They found that readers of the online version had a similar comprehension of the main text, but lower comprehension of the related documents, than students who read the hard-copy version. More interestingly, students who read the online

version reported that it took more effort to read, and that tables and graphics were difficult to understand. Students who read the paper-based version were less likely to skim or read selectively, while online readers tended to read selectively.

Mercieca's study of MBIT (Master of Business in Information Technology) students and electronic textbooks found that students printed material whenever possible, particularly when it was delivered in pdf format (Mercieca 2004). Students read content delivered in Microsoft reader format on screen only because the software lacked print functionality, and one student used a workaround of copying the content and pasting it into a word processing program to get a printed version.

Many people also prefer the convenience of print, because it is portable and can be read in pieces over time. The following case study describes the Library Association of New Zealand Aotearoa (LIANZA) experience in moving from a print to an online newsletter.

***Library Life* online: the LIANZA experience**

Library Life, the newsletter of Library Association of New Zealand Aotearoa (LIANZA), began as a conventional hard-copy publication in 1978 (continuing the *N.Z.L.A Newsletter* 1956-1977). It had eleven issues per year, and was distributed by mail to members of the Association and subscribers. Over time the content became longer and more sophisticated as the it was adapted to the requirements of an ever-more discerning customer base. As membership numbers grew the cost of producing larger and larger print runs became increasingly unaffordable.

In conjunction with an initiative of significantly reduced annual membership fees, in mid-2004 the LIANZA Council decided to publish *Library Life* online within six months—to be timed with the launch of a new website. The expected benefits were:

- the ability to create, edit and display news easily;
- the ability to leave comments and have discussions attached to articles;
- access to searchable archives; and
- the ability to bookmark or print key articles to which people wished to refer without needing to look through the entire print version.

In addition, printing and mailing costs would be eliminated.

From the beginning, Council members expected that the conversion would be controversial, but because of the lower expected revenue due to the reduced fees they felt it was essential to make the change.

A trial online version of *Library Life* was published in September 2004 using the open-source blog-ware 'CuteNews'. CuteNews is a small set of open-source PHP scripts which can be installed on any server that supports PHP. It can be configured to have a database of members and editors with different levels of access. Readers can comment on articles created via the online administration interface if this option is enabled and the server set to allow RW access to the appropriate folder. CuteNews is the best known product of the European website <http://www.cutephp.com>.

This format was so successful that LIANZA immediately began publishing *Library Life* online at the same time as the print version. This dual publishing mode was in place for two months

before the decision was made to move entirely online, and the last hard copy *Library Life* issue was published in November 2004. During the trial period, there was relatively little feedback, with a few notes of approval filtering through. However, when the print version ceased there was no way to ensure that everyone had looked at the new format, and the Association did not survey members for their opinions. Once print publication ceased, there was no shortage of opinions!

Regardless of their views on the appropriateness of digital distribution, the key point seized upon by most correspondents was the lack of a complete physical copy of any sort, whether for archiving, sharing or even reading. This was (in hindsight) the most predictable of comments—the managing editors had expected that people would simply print pages which interested them, but had not made any provision for printing an entire issue. After several months of digital distribution, there was a lively debate on the nz-libs listserv. A recurrent theme was a desire to print each issue in its entirety, illustrated by the following selection of comments:

I am not sure why a .pdf version which is more readily printable is not available now (several regional branches do it for their regional newsletters).

... when it comes to reading anything, I prefer the tactile, visual and olfactory qualities of paper and ink over the glow of the screen.

And yes, these '40 something' eyes do like to read paper, so, if it's been a really long day I will hit the print button too ;-)

I agree. It was a lunchtime read that blissfully got me away from the computer screen. Paper and electronic would be good.

The most commonly stated advantage for digital publication was improved access, while other people made the point that it was incongruous for librarians to be actively encouraging users to embrace e-publications, while resisting the concept for themselves.

There was initially no print option built into the CuteNews software. An article/page specific PRINT option which removed page formatting was immediately introduced. This was followed shortly by the creation of a PRINT ALL option which presented the entire current issue's data file as a continuous page of lightly-formatted html. Though this was quick and relatively easy, it did little to ease the complaints of those who missed the magazine style formatting of the previous print publication.

Several experiments to create more sophisticated pdf versions of each issue were done using the pdf2you software created for the New Zealand Ministry of Education, used on its Tertiary Education Wiki. We created several trial issues which were formatted in a journal style; this was done by adding appropriate DIV tags around stylistically different sections of the text. The trials were very encouraging, but further use of this software was forestalled by changes at LIANZA.

Comments about the change to a digital format had eased after a few month's issues had been published using CuteNews, suggesting people were either getting used to the new format or that they had given up complaining.

The format of *Library Life* was then changed again in May 2005 as content was migrated to a new LIANZA website. The new process generated each issue using a purpose-built set of scripts, but also failed to allow for easy printing of an entire issue. Though less public or concerted, a number of objections were again raised and the organisation quickly began to include pdf versions of the each issue in addition to the online version. These are created by

copying the raw text into a lightly formatted template and converting it to pdf.

Current print-on-demand options

This section identifies and compares four techniques currently used to provide print-on-demand facilities for users of web-based information.

1. Browser 'print' command

This is the 'do nothing' option, and is always available to users. However, few web pages are designed for easy printing, with many web pages containing information that is not relevant for a printed version, such as navigation menus or advertising. Adams (2003) identifies a number of issues associated with printing web pages, including:

- difficulty in printing a portion of a long page;
- pages that use frames can be difficult to print; and
- printing pages that have light text on a dark background.

Other issues that arise when printing pages using the browser's print function include narrow side margins, tables that break across pages, and text that is cut off or simply doesn't print. In addition, different browsers print the same page differently, so that what people get is not necessarily what they expected.

This option is the easiest one for web page creators, as it requires no extra effort on their part. However, as Potts (2003) notes, providing better print versions will be appreciated by users, and it can be worth the extra effort to implement one or more of the following options.

2. 'Print friendly' version

This approach is commonly found on online news and magazine sources, with a printer icon shown somewhere on the page. Users who click this generally see a reformatted page, often in a separate window, sometimes in a different font, and usually with no advertising or extraneous material. From a user's perspective, it is similar to using a browser 'print' function, but with an extra step. From the content creator's perspective, it is more work because it may require keeping two versions of the page, one formatted for the screen and the other for printing. If the web site includes groups of related pages, or content that spans several pages (such as long articles in an online journal or magazine), this option can generate a single printable document, to save the user's time.

3. Print-oriented stylesheet

Websites that have adopted cascading stylesheets (CSS) have another option for providing print-oriented pages by using a print media-specific stylesheet. This involves using one of three techniques to associate the stylesheet with the main document:

- a) using the LINK element with a media attribute, for example

```
<LINK rel="stylesheet" type="text/css" href="print.css" media="print">
```

- b) using the STYLE element, for example

```
<STYLE type="text/css">
```

```
@media print {
```

```
BODY {font-size: 10 pt; line-height: 120%; background: white}
}
@media screen {
BODY {font-size: medium; line-height: 100%; background: silver}
}
</STYLE>
```

c) using the @import statement, for example

```
@import url (print.css) print;
```

All of these will achieve the same result: when the user issues the browser 'print' command, the print-oriented stylesheet will be used to format the page's content. Print-oriented stylesheets often use serif fonts, and usually omit navigation and advertising links. Developing a good print-oriented stylesheet requires CSS designers to learn about typesetting, and consider such things as the relationship between headings and subsequent text, how links and underlying URLs are indicated, etc. The benefit for the user is that this happens automatically, without needing the extra step of requesting a printer-friendly version of the page. The disadvantage for the content creator is the extra effort of creating another stylesheet, but since it can easily be applied to all of the pages in the site, the extra effort is relatively small. One point to note is that this approach only works for users whose browsers support CSS.

4. Downloadable files

Another common approach is to have the content formatted for printing in Adobe's pdf format. This widely used format is designed to be printed, and lets the content creator fully control the appearance of the printed output. It is particularly well-suited to complex content, such as official documents, where features like tables of contents and page numbers may be important. Like the print-friendly version, it requires both the user and the content creator to do more. The user needs to download the file and read it using a pdf Reader; the creator needs to generate the file. This approach works best for relatively static content, and it is important that web publishing workflows contain checks to make sure that the pdf file is always synchronized with the web version of the content. Good practice is to indicate that the linked file is in pdf format, and it is also helpful to show the file size, because some people might not want to download a large file. This approach is the best for generating high-quality printed output, but it also requires more effort on the part of the content creator or webmaster.

Emerging alternatives

MacManus and Porter (2005) have suggested that the Web is becoming a platform from which to deliver services, rather than a publishing medium; they term this change "Web 2.0". One of the key trends they identify is a shift to describing content using XML-based semantic markup, rather than HTML. With increasing use of semantic markup, developers have new options for transforming content into printable output. A number of products that produce pdf files from HTML are available; they include HTMLDOC (<http://www.easysw.com/htmldoc/>) and Acrobat Web Capture. However, these work from the web pages rather than any underlying XML, meaning that the resulting files do not necessarily reflect the document's structure

accurately. Three XML-based projects to generate printable output are:

1. Prince (<http://www.princexml.com/>)

This project provides batch formatting software to convert XML documents to pdf or PostScript. It supports CSS and a range of DTDs, including DocBook and XHTML. It runs under Windows, Mac Os X, and Linux. A trial version is available free of charge. Prices range from \$US 349 for a single user version to \$US 3800 for a server version. Academic discounts are available.

2. Apache FOP (<http://xmlgraphics.apache.org/fop/>)

FOP (Formatting Objects Processor) is in the early stages of development, and it will render documents to a range of output formats, such as pdf, SVG, and print. PDF is intended to be its primary output. FOP uses XSL-FO to transform XML to print formats.

3. WikiPublisher (<http://www.wikipublisher.org/>)

This project lets cross-organizational virtual teams work on shared documents over the Web, and overcomes problems associated with word-processor approaches, such as:

- people working on different versions of the same document;
- reconciling and merging the efforts of several authors is error-prone and time-consuming; and
- incompatibilities between different word processors or different versions of the same word processor.

The project's objectives are to create a software package that:

- lets an author work with others to create documents ranging in size from a one page letter to a 10 page paper, 100 page report, or 1000 page book, using only a web browser;
- lets authors generate either print (PDF) or web versions of the document with one mouse click, applying suitable formatting for each medium; and
- produces printed output of desktop-publishing quality, without authors having to become publishing experts.

The project is based on existing off-the-shelf components where possible, and it uses the PmWiki engine to create documents; XHTML is generated for web output, and a sequence of enhanced tBook XML, XSLT, LaTeX, ImageMagick, and pdflatex for pdf.

Unlike Prince, WikiPublisher generates pdf files dynamically from the web site's content, and thus always reflects the current version. It handles text, lists, tables, graphics, and equations. Extensive use of metadata lets writers and readings use a web form to control the look of the printed output, such as heading and body text font style, duplex or single-sided layout, watermark, page size, and lists of tables or figures.

WikiPublisher has grown out of the pdf2you project sponsored by the New Zealand Ministry of Education, used on its Tertiary Education Wiki. WikiPublisher is began beta testing in August 2005, and will be released as open source software under the GPL. It will be demonstrated at the LITA National Forum session.

Conclusion

People have been using the Web for more than a decade, and it has become a widely used medium for locating and delivering information. However, reading information in electronic forms is much less popular, and it is commonly said that most people will print anything that is more than 1-2 screens in length. Current approaches to providing print-on-demand generate output of varying quality. Web designers should consider how their pages will print in addition to the way they look on the screen, something that has tended to be overlooked in the past. Developers of information-rich sites need to consider how their material will be printed, as well as how it looks on screen; this should ideally be done at an early stage in the site's development to avoid extensive re-work. Emerging XML-based alternatives offer new functionality, and as we move to more of a read-write Web, solutions such as Apache FOP, Prince, and WikiPublisher will become essential.

References

- Adams, Irene. 2003. Overcoming Web page printing problems. *Online* 27 (5):36-41.
- Brown, Gary J. 2001. Beyond print: reading digitally. *Library Hi Tech* 19 (4):390-399.
- Crawford, Walt. 1998. Paper persists: why physical library collections still matter. *Online* 22 (1):42-48.
- Dyson, Mary C. 2004. How physical text layout affects reading from screen. *Behaviour & Information Technology* 23 (6):377-393.
- Greytower Technologies. 2005. CSS vs. media 2004 [cited 28 August 2005]. Available from <http://www.greytower.net/archive/articles/cssmedia.html>.
- Lie, Håkon Wium, and Michael Day. 2005. Printing XML: why CSS is better than XSL, 19 January 2005 [cited 28 August 2005]. Available from <http://www.xml.com/pub/a/2005/01/19/print.html>.
- Macedo-Rouet, Mônica, Jean-François Rouet, Isaac Epstein, and Pierre Fayard. 2003. Effects of online reading on popular science comprehension. *Science Communication* 25 (2):99-128.
- MacManus, Richard, and Joshua Porter. 2005. Web 2.0 for designers. *Digital Web Magazine*, 4 May. Available from http://www.digital-web.com/articles/web_2_for_designers/
- Mercieca, Paul. 2004. E-book acceptance: what will make users read on screen? Paper read at Vala 2004: Breaking Boundarides: Integration & Interoperability, 3-5 February, at Melbourne, Victoria. Available from <http://www.vala.org.au/vala2004/2004pdfs/32Merci.PDF>
- Potts, Kevin. 2005. Webpage printing: typography and usability considerations (1.2), 6 November 2003 [cited 28 August 2005]. Available from <http://www.graphicpush.com/index.php?id=34>.