Electronic Publishing

Kamron Ackerman

First Edition, 2012

ISBN 978-81-323-2745-5

© All rights reserved.

Published by:
Orange Apple
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Introduction

Chapter 1 - E-Book

Chapter 2 - Metadata Publishing & Online Magazine

Chapter 3 - Comparison of e-Book Formats

Chapter 4 - Digital Edition

Chapter 5 - Online Newspaper

Chapter 6 - Open Access (Publishing)

Chapter 7 - Biblical Software

Chapter 8 - Blog

Chapter 9 - File Sharing

Chapter 10 - Digital Library

Chapter 11 - Digital Preservation

WT

Introduction

Electronic publishing or **ePublishing** includes the digital publication of e-books and electronic articles, and the development of digital libraries and catalogues. Electronic publishing has become common in scientific publishing where it has been argued that peer-reviewed scientific journals are in the process of being replaced by electronic publishing. Although distribution via the Internet (also known as **online publishing** or **web publishing** when in the form of a website) is nowadays strongly associated with electronic publishing, there are many non network electronic publications such as Encyclopedias on CD and DVD, as well as technical and reference publications relied on by mobile users and others without reliable and high speed access to a network.

After an article is submitted to a journal for consideration, there can be a delay ranging from several months to more than two years before it is published in a journal, rendering journals a less than ideal format for disseminating current research. In some fields such as astronomy and some parts of physics, the role of the journal in disseminating the latest research has largely been replaced by preprint repositories such as arXiv.org. However, scholarly journals still play an important role in quality control and establishing scientific credit. In many instances, the electronic materials uploaded to preprint repositories are still intended for eventual publication in a peer-reviewed journal.

There is statistical evidence that electronic publishing provides wider dissemination. A number of journals have, while retaining their peer review process, established electronic versions or even moved entirely to electronic publication.

Electronic publishing is increasingly popular in works of fiction as well as with scientific articles. Electronic publishers are able to provide quick gratification for late-night readers, books that customers might not be able to find in standard book retailers (erotica is especially popular in eBook format), and books by new authors that would be unlikely to be profitable for traditional publishers.

While the term "electronic publishing" is primarily used today to refer to the current offerings of online and web-based publishers, the term has a history of being used to

describe the development of new forms of production, distribution, and user interaction in regard to computer-based production of text and other interactive media.

Examples

Electronic versions of traditional media:

- CD-ROM
- E-book
- Electronic journal
- Online newspaper
- Online magazine

New media:

- File sharing
- Podcast
- Collaborative software
- Blog

Business models

- Online advertising
- Open access (publishing)
- Pay-Per-View
- Print on demand
- Subscriptions
- Self-publishing
- Non-Subsidy Publishing

Technology vendors

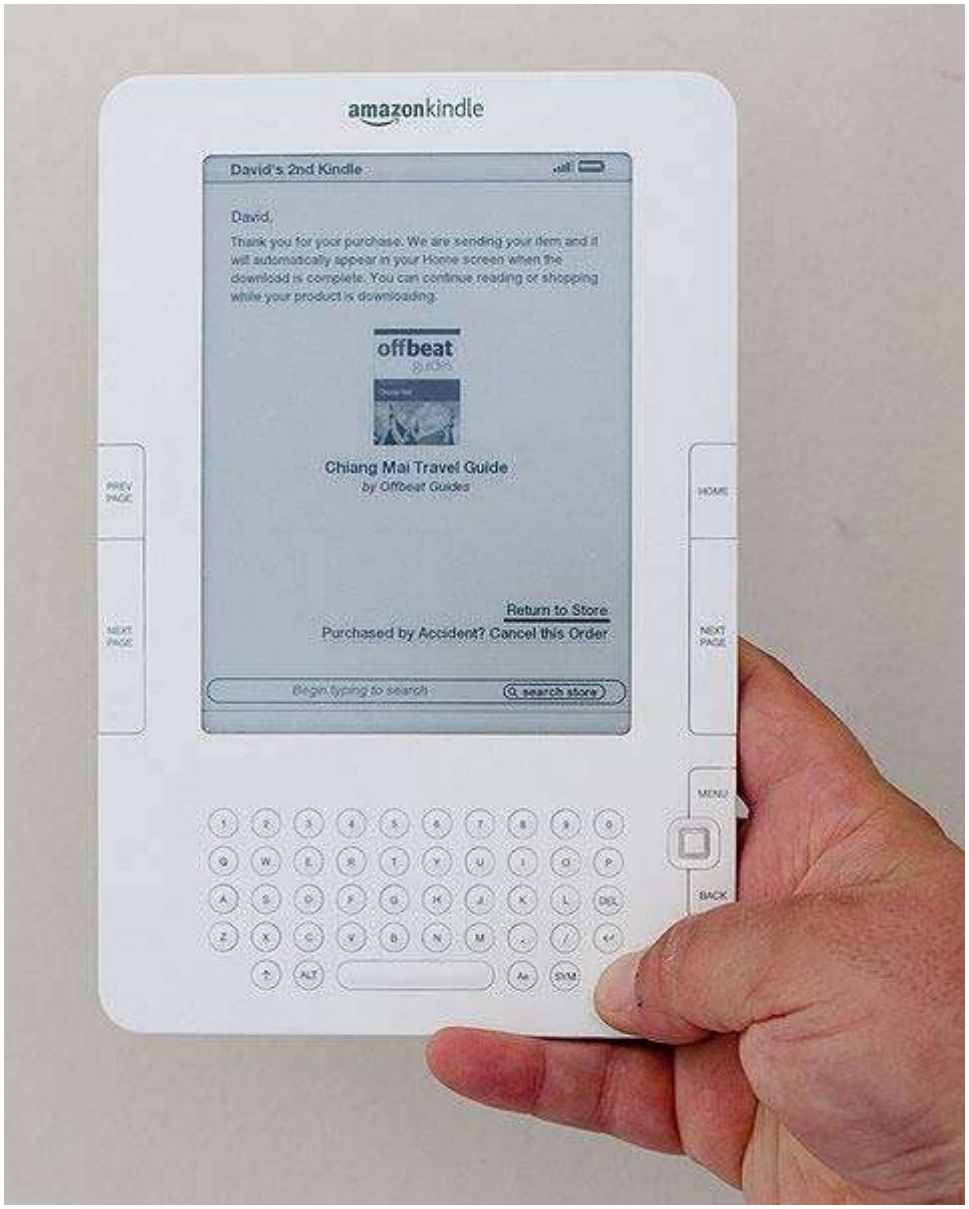
- Eastgate Systems
- eMeta Corporation
- InformIT
- Ingenta
- Mygazines
- Safari Books Online
- Zmags

WT

Chapter 1

E-Book

A user viewing an electronic page on a prototype OLPC



Amazon Kindle 2

An **electronic book** (also **e-book**, **ebook**, **digital book**) is a text and image-based publication in digital form produced on, published by, and readable on computers or other digital devices. Sometimes the equivalent of a conventional printed book, e-books can also be born digital. The *Oxford Dictionary of English* defines the e-book as "an electronic version of a printed book," but e-books can and do exist without any printed equivalent. E-books are usually read on dedicated hardware devices known as *e-Readers* or *e-book devices*. Personal computers and some cell phones can also be used to read e-books.

History

Among the earliest general e-books were those in *Project Gutenberg,* in 1971. One early e-book implementation was the desktop prototype for a proposed notebook computer, the *Dynabook,* in the 1970s at PARC: a general-purpose portable personal computer capable of displaying books for reading.

Early e-books were generally written for specialty areas and a limited audience, meant to be read only by small and devoted interest groups. The scope of the subject matter of these e-books included technical manuals for hardware, manufacturing techniques and other subjects. In the 1990s, the general availability of the Internet made transferring electronic files much easier, including e-books.

Numerous e-book formats, view comparison of e-book formats, emerged and proliferated, some supported by major software companies such as Adobe with its PDF format, and others supported by independent and open-source programmers. Multiple readers followed multiple formats, most of them specializing in only one format, and thereby fragmenting the e-book market even more. Due to exclusiveness and limited readerships of e-books, the fractured market of independents and specialty authors lacked consensus regarding a standard for packaging and selling e-books. In 2010 e-books continued to gain in their own underground markets. Many e-book publishers began distributing books that were in the public domain. At the same time, authors with books that were not accepted by publishers offered their works online so they could be seen by others. Unofficial (and occasionally unauthorized) catalogs of books became available over the web, and sites devoted to e-books began disseminating information about e-books to the public.

U.S. Libraries began providing free e-books to the public in 1998 through their web sites and associated services, although the e-books were primarily scholarly, technical or professional in nature, and could not be downloaded. In 2003, libraries began offering free downloadable popular fiction and non-fiction e-books to the public, launching an e-book lending model that worked much more successfully for public libraries. The number of library e-book distributors and lending models continued to increase over the next few years. In 2010, a Public Library Funding and Technology Access Study found that 66% of public libraries in the U.S. were offering e-books, and a large movement in the library industry began seriously examining the issues related to lending e-books, acknowledging a tipping point of broad e-book usage.

As of 2009, new marketing models for e-books were being developed and dedicated reading hardware was produced. E-books (as opposed to ebook readers) have yet to achieve global distribution. In the United States, as of September 2009, the Amazon Kindle model and Sony's PRS-500 were the dominant e-reading devices. By March 2010, some reported that the Barnes & Noble Nook may be selling more units than the Kindle. On January 27, 2010 Apple Inc. launched a multi-function device called the iPad and announced agreements with five of the six largest publishers that would allow Apple to

distribute e-books. However, many publishers and authors have not endorsed the concept of electronic publishing, citing issues with demand, piracy and proprietary devices.

In July 2010, online bookseller Amazon.com reported sales of ebooks for its proprietary Kindle outnumbered sales of hardcover books for the first time ever during the second quarter of 2010, saying it sold 140 e-books for every 100 hardcover books, including hardcovers for which there was no digital edition. By January 2011, ebook sales at Amazon had surpassed its paperback sales. In the overall U.S. market, paperback book sales are still much larger than either hardcover or e-book; the American Publishing Association estimated e-books represented 8.5% of sales as of mid-2010. In Canada, the option of ebook publishing took a higher profile when the novel, *The Sentamentalists*, won the prestigious national Giller Prize. Owing to the small scale of the novel's independent publisher, the book was initially not widely available in printed form, but the ebook edition had no such problems with it becoming the top-selling title for Kobo devices.

Timeline

1971

- Michael S. Hart launches *Project Gutenberg*.

1985–1992

- Robert Stein starts Voyager Company Expanded Books and books on CD-ROM.

1992

- Charles Stack's Book Stacks Unlimited begins selling new physical books online.

1993

- Zahur Klemath Zapata develops the first software to read digital books. Digital book version 1 and the first digital book is published *On Murder Considered as one of the Fine Arts* (Thomas de Quincey).
- Digital Book, Inc. offers the first 50 digital books in floppy disk with Digital Book Format (DBF).
- Hugo Award for Best Novel nominee texts published on CD-ROM by Brad Templeton.
- Bibliobytes, a project of free digital books online in Internet.

1995

- Amazon starts to sell physical books on the Internet.
- Online poet Alexis Kirke discusses the need for wireless internet electronic paper readers in his article "The Emuse".

1996

- Project Gutenberg reaches 1,000 titles. The target is 1,000,000.

1998

- Kim Blagg obtained the first ISBN issued to an ebook and began marketing multimedia-enhanced ebooks on CDs through retailers including amazon.com, bn.com and borders.com. Shortly thereafter through her company "Books OnScreen" she introduced the ebooks at the Book Expo America in Chicago, IL to an impressed, but unconvinced bookseller audience.
- First ebook Readers: Rocket ebook and SoftBook.
- Cybook / Cybook Gen1 Sold and manufactured at first by Cytale (1998–2003) then by Bookeen.

1999

- Baen Books opens up the Baen Free Library.
- Webscriptions starts selling unencrypted eBooks.

2000

- Microsoft Reader with ClearType technology.
- Stephen King offers his book "Riding the Bullet" in digital file; it can only be read on a computer.

2001

- Todoebook.com, the first website selling ebooks in Spanish.

2002

- Random House and HarperCollins start to sell digital versions of their titles in English.

2004

- Sony Librie with e-ink.

2005

- Amazon buys Mobipocket.
- Bookboon.com is launched, allowing people to download free textbooks and travel guide eBooks.

2006

- Sony Reader with e-ink.
- LibreDigital launched BookBrowse as an online reader for publisher content.
- BooksOnBoard, one of the largest independent ebookstores, opens and sells ebooks and audiobooks in six different formats.

2007.

- Amazon launches Kindle in US.
- Bookeen launched Cybook Gen3 in Europe.

2008

- Adobe and Sony agreed to share their technologies (Reader and DRM).
- Sony sells the Sony Reader PRS-505 in UK and France.
- BooksOnBoard is first to sell ebooks for iPhones.

2009

- Bookeen releases the Cybook Opus in the US and in Europe.
- Sony releases the Reader Pocket Edition and Reader Touch Edition.
- Amazon releases the Kindle 2.
- Amazon releases the Kindle DX in the US.
- Barnes & Noble releases the Nook in the US.
- Bookboon.com achieves over 10 Million downloads in one year — placing the company as the world's largest publisher of free eBooks.

2010

- Amazon releases the Kindle DX International Edition worldwide.
- Bookeen reveals the Cybook Orizon at CES.
- TurboSquid Magazine announces first magazine publication using Apple's iTunes LP format.
- Apple releases the iPad with an e-book app called iBooks. Between its release in April 2010, to October, Apple has sold 7 million iPads.
- Kobo Inc. releases its Kobo eReader to be sold at Indigo/Chapters in Canada and Borders in the United States.
- Amazon.com reported that its e-book sales outnumbered sales of hardcover books for the first time ever during the second quarter of 2010.
- Amazon releases the third generation kindle, available in 3G+Wi-Fi and Wi-Fi versions.
- Kobo Inc. releases an updated Kobo eReader which now includes Wi-Fi.
- Barnes & Noble releases the new NOOKcolor.
- Sony releases its second generation Daily Edition PRS-950.
- PocketBook expands its successful line of e-readers in the ever-growing market.

- Google launches Google eBooks

Formats

There are a variety of e-book formats used to create and publish e-books. A writer or publisher has many options when it comes to choosing a format for production. Every format has its proponents and champions, and debates over which format is best can become intense.

Comparison to printed books

Advantages

There are over 2 million free books available for download as of August 2009. Mobile availability of e-books may be provided for users with a mobile data connection, so that these e-books need not be stored on the device. An e-book can be offered indefinitely, without ever going "out of print". In the space that a comparably sized print book takes up, an e-reader can potentially contain thousands of e-books, limited only by its memory capacity. If space is at a premium, such as in a backpack or at home, it can be an advantage that an e-book collection takes up little room and weight.

E-book websites can include the ability to translate books into many different languages, making the works available to speakers of languages not covered by printed translations. Depending on the device, an e-book may be readable in low light or even total darkness. Many newer readers have the ability to display motion, enlarge or change fonts, use Text-to-speech software to read the text aloud for visually impaired, partially sighted, elderly or dyslectic people, search for key terms, find definitions, or allow highlighting bookmarking and annotation. Devices that utilize E Ink can imitate the look and ease of readability of a printed work while consuming very little power, allowing continuous reading for weeks at time.

While an e-book reader costs much more than one book, the electronic texts are at times cheaper. Moreover, a great share of e-books are available online for free, minus the minimal costs of the electronics required. For example, all fiction from before the year 1900 is in the public domain. Also, libraries lend more current e-book titles for limited times, free samples are available of many publications, and there are other lending models being piloted as well. E-books can be printed for less than the price of traditional new books using new on-demand book printers.

An e-book can be purchased/borrowed, downloaded, and used immediately, whereas when one buys or borrows a book, one must go to a bookshop, a home library, or public library during limited hours, or wait for a delivery. The production of e-books does not consume paper and ink. The necessary computer or e-reader uses less materials. Printed books use 3 times more raw materials and 78 times more water to produce albeit they do not require a machine for use (out of context) Depending on possible digital rights management, e-books can be backed up to recover them in the case of loss or damage and

it may be possible to recover a new copy without cost from the distributor. Compared to printed publishing, it is cheaper and easier for authors to self-publish e-books. Also, the dispersal of a free e-book copy can stimulate the sales of the printed version.

Drawbacks

Ebook formats and file types continue to develop and change through time through advances and developments in technology or the introduction of new proprietary formats. While printed books remain readable for many years, e-books may need to be copied or converted to a new carrier or file type over time. PDF and epub are growing standards, but are not universal.

Not all books are available as e-books. Paper books can be bought and wrapped for a present and a library of books can provide visual appeal, while the digital nature of e-books makes them non-visible or tangible. E-books cannot provide the physical feel of the cover, paper, and binding of the original printed work. An author who publishes a book often puts more into the work than simply the words on the pages. E-books may cause people "to do the grazing and quick reading that screens enable, rather than be by themselves with the author's ideas". They may use the e-books simply for reference purposes rather than reading for pleasure and leisure. Books with large pictures (such as children's books) or diagrams are more inconvenient for viewing and reading.

A book will never turn off and would be unusable only if damaged or after many decades. The shelf life of a printed book exceeds that of an e-book reader, as over time the reader's battery will drain and require recharging. Additionally, "As in the case of microfilm, there is no guarantee that [electronic] copies will last. Bits become degraded over time. Documents may get lost in cyberspace...Hardware and software become extinct at a distressing rate." E-book readers are more susceptible to damage from being dropped or hit than a print book. Due to faults in hardware or software, e-book readers may malfunction and data loss can occur. As with any piece of technology, the reader must be protected from the elements (such as extreme cold, heat, water, etc.), while print books are not susceptible to damage from electromagnetic pulses, surges, impacts, or extreme temperatures.

The cost of an e-book reader far exceeds that of a single book, and e-books often cost the same as their print versions. Due to the high cost of the initial investment in some form of e-reader, e-books are cost prohibitive to much of the world's population. Furthermore, there is no used e-book market, so consumers will neither be able to recoup some of their costs by selling an unwanted title they have finished, nor will they be able to buy used copies at significant discounts, as they can now easily do with printed books. Because of the high-tech appeal of the e-reader, they are a greater target for theft than an individual print book. Along with the theft of the physical device, any e-books it contains also become stolen. E-books purchased from vendors like Amazon or Barnes & Noble.com are stored "in the cloud" on servers and "digital lockers" and have the benefit of being easily retrieved if an e-reading device is lost. Not all e-booksellers are cloud based; if an

e-book is stolen, accidentally lost, or deleted, in the absence of a backup it may have to be repurchased.

The display resolutions of reading devices are currently lower than printed materials. Because of proprietary formats or lack of file support, formatted e-books may be unusable on certain readers. Additionally, the reader's interaction with the reader may cause discomfort, for example glare on the screen or difficulty holding the device. Due to digital rights management, customers typically cannot resell or loan their e-books to other readers. However, some Barnes & Noble e-books are lendable for two weeks via their 'LendMe' technology. Additionally, the potential for piracy of e-books may make publishers and authors reluctant to distribute digitally. E-book readers require various toxic substances to produce, are non-biodegradable, and the disposal of their batteries in particular raises environmental concerns. As technologies rapidly change and old devices become obsolete, there will be larger amounts of toxic wastes that are not easily biodegradable like paper. Paper products are easily sustainable and reusable, unlike many rare earth minerals that are used up in electronic devices.

A rare or fine book can be an art object with a high monetary value. One can invest in first editions and out of print books. Some books will have a very high resale value. Real paper books can be used to decorate a home or office. Some finely bound, limited edition books can be considered very beautiful. Very old books often have great historical importance, and are one of a kind. Archives can easily store old paper books and documents, unlike e-books.

E-books and software can easily track data, times, usage, pages, and details about what one is reading and how often. Similar to this is the growing amount of data available through Google search engines, Facebook, and through data mining. For the first time in history it is now far easier to track and record what specific people might be reading. The notions of privacy, private writing, solitude, and personal reading are changing.

Digital rights management

Anti-circumvention techniques may be used to restrict what the user may do with an e-book. For instance, it may not be possible to transfer ownership of an e-book to another person, though such a transaction is common with physical books. Some devices can phone home to track readers and reading habits, restrict printing, or arbitrarily modify reading material. This includes restricting the copying and distribution of works in the public domain through the use of "click-wrap" licensing, effectively limiting the rights of the public to distribute, sell or use texts in the public domain freely.

Most e-book publishers do not warn their customers about the possible implications of the digital rights management tied to their products. Generally they claim that digital rights management is meant to prevent copying of the e-book. However in many cases it is also possible that digital rights management will result in the complete denial of access by the purchaser to the e-book. With some formats of DRM, the e-book is tied to a specific computer or device. In these cases the DRM will usually let the purchaser move

the book a limited number of times after which he cannot use it on any additional devices. If the purchaser upgrades or replaces their devices eventually they may lose access to their purchase. Some forms of digital rights management depend on the existence of online services to authenticate the purchasers. When the company that provides the service goes out of business or decides to stop providing the service, the purchaser will no longer be able to access the e-book.

As with digital rights management in other media, e-books are more like rental or leasing than purchase. The restricted book comes with a number of restrictions, and eventually access to the purchase can be removed by a number of different parties involved. These include the publisher of the book, the provider of the DRM scheme, and the publisher of the reader software. These are all things that are significantly different from the realm of experiences anyone has had with a physical copy of the book.

Production

Some e-books are produced simultaneously with the production of a printed format, as described in electronic publishing, though in many instances they may not be put on sale until later. Often, e-books are produced from pre-existing hard-copy books, generally by document scanning, sometimes with the use of robotic book scanners, having the technology to quickly scan books without damaging the original print edition. Scanning a book produces a set of image files, which may additionally be converted into text format by an OCR program. Occasionally, as in some e-text projects, a book may be produced by re-entering the text from a keyboard.

As a newer development, sometimes only the electronic version of a book is produced by the publisher. It is even possible to release an e-book chapter by chapter as each chapter is written. This is useful in fields such as information technology where topics can change quickly in the months that it takes to write a typical book (See: Realtime Publishers). It is also possible to convert an electronic book to a printed book by print on demand. However these are exceptions as tradition dictates that a book be launched in the print format and later if the author wishes an electronic version is produced.

As of 2010, there is no industry-wide e-book bestseller list, but various e-book vendors compile bestseller lists, such as those by Amazon Kindle Bestsellers and Fictionwise. There are two yearly awards for excellence in e-books—the EPIC eBook Award (formerly EPPIE) given by EPIC, and the Dream Realm Award for science fiction, fantasy and horror e-books. Both awards have been given since 2000.

e-Readers

e-Readers may be specifically designed for that purpose, or intended for other purposes as well. The term is restricted to hardware devices and used to describe a category type.

Specialized devices have the advantage of doing one thing well. Specifically, they tend to have the right screen size, battery lifespan, lighting and weight. A disadvantage of such

devices is that they are often expensive when compared to multi-purpose devices such as laptops and PDAs.

In 2010, competition sent the price for the most popular electronic reading devices below USD 200.

Research released in March 2011 indicated that e-books and e-book readers are actually more popular with the older generation than the younger generation in the UK. The survey carried out by Silver Poll found that around 6% of over 55s owned an e-book reader compared with just 5% of 18-24 year olds.

The survey also revealed that the Amazon Kindle is the most popular e-book reader in the UK (47%) followed by the Apple iPad (31%) and the Sony Reader (14%).

It has been reported that there is a differing level of dissatisfaction amongst owners of different ebook readers due to poor availability of sought after ebook titles. A survey of the number of contemporary and popular titles available from ebook store, revealed that Amazon.com has the largest collection, over twice as large as Barnes and Noble, Sony Reader Store, Apple iBookstore and OverDrive, the public libraries lending system.

WT

Chapter 2

Metadata Publishing & Online Magazine

Metadata Publishing

Metadata publishing is the process of making metadata data elements available to external users, both people and machines using a formal review process and a commitment to change control processes.

Metadata publishing is the foundation upon which advanced distributed computing functions are being built. But like building foundations, care must be taken in metadata publishing systems to ensure the structural integrity of the systems built on top of them.

Definition of metadata publishing

Published metadata has the following characteristics:

1. Metadata structures available to the general public on a public web site or by a download
2. There is a documented review and approval process for adding or updating data elements to the system
3. New releases are made available without disturbing prior versions
4. A publishing organization that makes a commitment to change control process

Benefits of metadata publishing

When classifying benefits of metadata publishing two groups are usually considered. External parties are usually consumers of information that are not part of the publishing organization. Internal parties are usually the various business units or departments within an organization.

Benefits to external parties

1. Allows external systems (both people and agents) to have a clear understanding of the semantics of data elements in a system
2. Allows third parties to build semantic maps between data models and import and export data between systems
3. Promotes service oriented architectures and allow horizontal sharing of information between traditional information silos
4. Allows systems to participate in accurately indexed and federated search processes

Benefits to internal parties

1. allows parties from diverse business units to agree on shared data definitions and separate department or function specific definitions
2. makes Extract, transform, load (ETL) operations more precise for data warehousing
3. allows user interface designers to access a common pool of screen and report header labels
4. promotion of model-driven architecture

Objections to metadata publishing

- Organizations that publish their metadata could make it easier for unauthorized people to find sensitive data if they breach an organization's firewall
- Vendors that publish their metadata risk customers creating tools that could allow their customers to export their data from computer systems therefor making it easier to migrate off of a vendor's system

Core process in metadata publishing

The following are some of the core processes in metadata publishing

1. Gathering of metadata requirements
2. Selection of metadata registry and metadata publishing tools
3. Training of metadata concepts to project participants
4. Stakeholder group formation
5. Metadata harvesting
6. Glossary consolidation
7. Initial upper ontology construction (abstract data elements)
8. Draft data element loading
9. Data element review process
10. Publishing approved metadata elements in a variety of output formats
11. Creation and maintenance of versions and depreciation of unused or redundant data elements

File format metadata publishing

Organizations that create applications that store data in file systems can also publish metadata definitions. One common way to perform this is to store application data in a compressed XML file format. The XML files can be uncompressed and validated against an external XML Schema. An example of this is done by the Open Source FreeMind tool.

Metadata publishing formats

1. HTML - used for browsing a web site and indexing by text-based search engines
2. Web Ontology Language (OWL) - used by metadata search engines such as Swoogle
3. XML Metadata Interchange (XMI) - OMG standard for exchanging metadata
4. Common Warehouse Metamodel (CMW) - OMG standard for data warehouse metadata
5. Topic maps - an ISO standard for the representation and interchange of knowledge, with an emphasis on the findability of information.
6. KM3 or Kernel Meta Meta Model as used in the Metamodel Zoos. The AtlanticZoo is an open source library of more than 100 metamodels under EPL License. KM3 is a simple Domain Specific Language for specifying metamodels. A number of transformations are available to translate from KM3 to other notations like XMI.

Online Magazine

An **online magazine** shares some features with a blog and also with online newspapers, but can usually be distinguished by its approach to editorial control. Magazines typically have editors or editorial boards who review submissions and perform a quality control function to ensure that all material meets the expectations of the publishers (those investing time or money in its production) and the readership.

Online magazines that are part of the World Wide Web, that is, all or part of a website, are sometimes called ***webzines***. An **ezine** (also spelled **e-zine** and usually is a more specialized term appropriately applied to small magazines and newsletters distributed by any electronic method, for example, by electronic mail (e-mail/email). Some social groups may use the terms **cyberzine** and **hyperzine** when referring to electronically distributed resources. Similarly, some online magazines may refer to themselves as "electronic magazines" to reflect their readership demographics or to capture alternative terms and spellings in online searches.

Many large print-publishers now provide digital reproduction of their print magazine titles through various online services for a fee. These service providers also refer to their collections of these digital format products as online magazines, and sometimes as digital magazines.

Online magazines representing matters of interest to specialists in or societies for academic subjects, science, trade or industry are typically referred to as online journals.

Business model

Many general interest online magazines provide free access to all aspects of their online content although some publishers have opted to require a subscription fee to access premium online article and/or multi-media content. Online magazines may generate revenue based on targeted search ads to web-site visitors, banner ads (online display advertising), affiliations to retail web sites, classified advertisements, product-purchase capabilities, advertiser directory links, or alternative informational/commercial purpose.

The original online magazines, ezines and disk magazines, due to their low cost and initial non-mainstream targets, may be seen as a disruptive technology to traditional publishing houses. The high cost of print publication and large web readership has encouraged these publishers to embrace the World Wide Web as a marketing and content delivery system and another medium for delivering their advertisers' messages.

Growth

In the late 1990s ezine publishers began adapting to the interactive qualities of the Internet instead of duplicating magazines on the web. Publishers of traditional print titles and entrepreneurs with an eye to a potential readership in the millions started publishing online titles. Salon.com founded in July 1995 by David Talbot was launched with considerable media exposure and today reports 5.8 million monthly unique visitors.

In the 2000s, some webzines began appearing in a printed format to complement their online versions. These included *Movie Insider*, *Slate*, *Synthesis* and *Lucire* magazines.

Conferences

Between 1998 and 2005, in San Francisco and New York, a series of webzine-focused conferences brought together independent personal online publishers to share their experiences. Started by Srini Kumar, the "Webzine" conferences were continued primarily by filmmaker Ryan Junell and Eddie Codel. Junell has worked to track the history of the early webzine movement through these festivals; his research is linked below. After a hiatus, Codel and Junell organized the return of the Webzine conference to the Bay Area in 2005. Webzine 2005 took place over two days at the Swedish-American Hall in San Francisco. It consisted of three main areas: speakers and panel discussions, workshops and a self-organizing area called the Master's Lounge modeled after BAR Camp. Webzine 2005 was emceed by veteran Webzine emcee Justin Hall, Annalee Newitz and Charlie Anders.

Today there are many conferences that address online magazine publishing from a variety of perspectives.

Chapter 3

Comparison of e-Book Formats

The following is a **comparison of e-book formats** used to create and publish e-books.

A writer or publisher has many options when it comes to choosing a format for publication. While the average end-user might arguably simply want to read books, every format has its proponents. The myriad e-book formats are sometimes collectively referred to as the "Tower of eBabel".

The storage size for texts without images depends on the file format, but is always relatively small compared with a richly illustrated text.

Format descriptions

Formats available include, but are by no means limited to:

Plain text files

Format: text
Published as: .txt

E-books in plain text exist. The size in bytes is simply the number of characters, including spaces, and with a new line counting for 1 or 2. For example, the Bible, an 800,000-word book, is about 4 MB. The ASCII standard allows ASCII-only text files (unlike most other file types) to be interchanged and readable on Unix, Macintosh, Microsoft Windows, DOS, and other systems. These differ in their preferred line ending convention and their interpretation of values outside the ASCII range (their character encoding).

Hypertext Markup Language

Format: Hypertext
Published as: .htm; .html

HTML is the markup language used for most web pages. E-books using HTML can be read using a Web browser. The specifications for the format are available without charge from the W3C.

HTML adds specially marked meta-elements to otherwise plain text encoded using character sets like ASCII or UTF-8. As such, suitably formatted files can be, and sometimes are, generated *by hand* using a *plain text editor* or *programmer's editor*. Many *HTML generator* applications exist to ease this process and often require less intricate knowledge of the format details involved.

HTML on its own is not a particularly efficient format to store information, requiring more storage space for a given work than many other formats. However, several e-Book formats including the Amazon Kindle, Open eBook, Compressed HM, Mobipocket and EPUB use one HTML file for each book chapter and then Zip compress the files, along with images, metadata and style sheets into one file.

HTML files encompass a wide range of standards and displaying HTML files correctly can be complicated. Additionally many of the features supported, such as forms, are not relevant to e-books.

Amazon Kindle

Format: Kindle
Published as: .azw

With the launch of the Kindle eBook reader, Amazon.com created the proprietary format, AZW. It is based on the Mobipocket standard, with a slightly different serial number scheme (it uses an asterisk instead of a dollar sign) and its own DRM formatting. Because the eBooks bought on the Kindle are delivered over its wireless system called Whispernet, the user does not see the AZW files during the download process. The Kindle format is now available on a variety of platforms.

Open Electronic Package

Format: Open eBook
Published as: .opf

OPF is an XML-based e-book format created by E-Book Systems.

TomeRaider

Format: TomeRaider

Published as: .tr2; .tr3

The TomeRaider e-book format is a proprietary format. There are versions of TomeRaider for Windows, Windows Mobile (aka Pocket PC), Palm, Symbian, iPhone and more. Capabilities of the TomeRaider3 e-book reader vary considerably per platform: the Windows and Windows Mobile editions support full HTML and CSS. The Palm edition supports limited HTML (e.g., no tables, no fonts), and CSS support is missing. For Symbian there is only the older TomeRaider2 format, which does not render images or offer category search facilities. Despite these differences any TomeRaider e-book can be browsed on all supported platforms. The Tomeraider website claims to have over 4000 e-books available, including free versions of the Internet Movie Database.

Arghos Diffusion

Format: Arghos Reader

Published as: .aeh

The AEH format is an XML-based proprietary format developed by the French firm Arghos Diffusion. AEH files use a proprietary DRM and encryption method and are readable only in the *Arghos Player*. It supports various input formats for text, audio or video, such as PDF, WMA, MP3, WMV, and allows multiple interactive functions such as bookmarking, advanced plain-text searching, dynamic text highlighting, etc.

Flip Books

Format: Interaxive media

Published as:

A "Flip Book" is a type of E-Book distinguished by virtual pages that actually "flip", much like turning pages of paper in a real book or magazine. The first dynamic Flip Book Reader was developed in 2003/2004 by Interaxive Media for Nishe Media (Canada) and was therefore called "Nishe Pages". The first version was produced in part by Cybaris (Canada) and was first publicly showcased in August 2004. Soon thereafter, many copycat "flip books" started appearing thanks to technological advances in Macromedia Flash, mostly hard coded using Flash components.

The original software remains unique in that it is powered by a complete server-based CMS system that allows the books to be created, published, and viewed remotely from a web server without requiring any custom software to be installed. Nishe Media went defunct in 2004, leaving the unfinished software to Interaxive Media who continued its development in Hong Kong. Though not widely used outside of Asia, it is now at version 3.0 and can be a server-based E-Book platform. It remains privately held by the original developer, Ryan Sutherland, owner and founder of Interaxive Media.

ANSI/NISO Z39.86 (DAISY)

Format: DAISY

Published as:

The Digital Accessible Information SYstem (DAISY) is an XML-based open standard maintained by the DAISY Consortium for people with print disabilities. DAISY has wide international support with features for multimedia, navigation and synchronization. A subset of the DAISY format has been adopted by law in the United States as the National Instructional Material Accessibility Standard (NIMAS), and K-12 textbooks and instructional materials are now required to be provided to students with disabilities.

DAISY is already aligned with the EPUB open standard, and is expected to fully converge with its forthcoming EPUB3 revision.

FictionBook (Fb2)

Format: FictionBook
Published as: .fb2

FictionBook is a popular XML-based e-book format, supported by free readers such as FBReader, Haali Reader and STDU Viewer.

Text Encoding Initiative

Format: TEI Lite
Published as: .xml

TEI Lite is the most popular of the TEI-based (and thus XML-based or SGML-based) electronic text formats.

Plucker

Format: Plucker
Published as:

Plucker is a free e-book reader application with its own associated file format and software to automatically generate plucker files from HTML files, web sites or RSS feeds. The format is a compressed HTML archive, somewhat like Microsoft's CHM.

Compressed HM

Format: Microsoft Compressed HTML Help
Published as: .chm

CHM format is a proprietary format based on HTML. Multiple pages and embedded graphics are distributed along with proprietary metadata as a single compressed file. In contrast, in HTML, a site consists of multiple HTML files and associated image files in standardized formats.

Portable Document Format

Format: Adobe Portable Document Format
Published as: .pdf

A file format created by Adobe Systems, initially to provide a standard form for storing and editing printed publishable documents. The format derives from PostScript, but without language features like loops, and with added support for features like compression and passwords. Because PDF documents can easily be viewed and printed by users on a variety of computer platforms, they are very common on the World Wide Web. The specification of the format is available without charge from Adobe.

PDF files typically contain brochures, product manuals, magazine articles — up to entire books, as they can embed fonts, images, and other documents. A PDF file contains one or more zoomable page images.

Since the format is designed to reproduce page images, the text traditionally could not be re-flowed to fit the screen width or size. As a result PDF files designed for printing on standard paper sizes are less easily viewed on screens with limited size or resolution, such as those found on mobile phones and PDAs. Adobe has addressed this drawback by adding a re-flow facility to its Acrobat Reader software, but for it to work the document must be marked for re-flowing at creation — meaning that existing PDF documents won't benefit unless they are tagged and resaved. The Windows Mobile (aka Pocket PC) version of Adobe Acrobat will automatically attempt to tag a PDF for reflow during the synchronization process using an installed plugin to Active Sync. However, this tagging process will not work on most locked or password protected PDF documents. It also doesn't work at present (2009–10) on the Windows Mobile Device Center (the successor to Active Sync) as found in Windows Vista and Windows 7. Thus, automatic tagging support during synchronization is limited to Windows XP/2000.

Multiple products support creating and tagging PDF files, such as Adobe Acrobat, PDFCreator, OpenOffice.org, iText, and FOP, and several programming libraries. Adobe Reader (formerly called *Acrobat Reader*) is Adobe's product used to view PDF files; third party viewers such as xpdf are also available. Mac OS X has built-in PDF support, both for creation as part of the printing system and for display using the built-in Preview application.

Later versions of the specification add support for forms, comments, hypertext links, and even interactive elements such as buttons for forms entry and for triggering sound and video. Such features may not be supported by older or third-party viewers and some are not transferable to print.

PDF files are supported on the following e-book readers: Mobipocket, iRex iLiad, iRex DR1000, Sony Reader, Bookeen Cybook, Foxit eSlick, Amazon Kindle (1, 2, International & DX), Barnes & Noble Nook, the iPad, PocketBook Reader, Bebook Neo and the Kobo eReader. Also, pdf files can be read on the iPod Touch using the free Stanza app.

PostScript

Format: PostScript
Published as: ps

PostScript is a page description language used in the electronic and desktop publishing areas for defining the contents and layout of a printed page, which can be used by a rendering program to assemble and create the actual output bitmap. Many office printers directly support interpreting PostScript and printing the result. As a result, the format also sees wide use in the Unix world.

DjVu

Format: DjVu
Published as: .djvu

DjVu is a format specialized for storing scanned documents. It includes advanced compressors optimized for low-color images, such as text documents. Individual files may contain one or more pages. DjVu files cannot be re-flowed.

The contained page images are divided in separate layers (such as multi-color, low-resolution, background layer using lossy compression, and few-colors, high-resolution, tightly-compressed foreground layer), each compressed in the best available method. The format is designed to decompress very quickly, even faster than vector-based formats.

The advantage of DjVu is that it is possible to take a high-resolution scan (300-400 DPI), good enough for both on-screen reading and printing, and store it very efficiently. Several dozens of 300 DPI black-and-white scans can be stored in less than a megabyte.

Microsoft LIT

Format: Microsoft Reader
Published as: .lit

DRM-protected LIT files are only readable in the proprietary Microsoft Reader program, as the .LIT format, otherwise similar to Microsoft's CHM format, includes Digital Rights Management features. Other third party readers, such as Lexcycle Stanza, can read unprotected LIT files. There are also tools such as Convert Lit, which can convert .lit files to HTML files or OEBPS files.

The Microsoft Reader uses patented ClearType display technology. In Reader navigation works with a keyboard, mouse, stylus, or through electronic bookmarks. The Catalog Library records reader books in a personalized "home page", and books are displayed with ClearType to improve readability. A user can add annotations and notes to any page, create large-print e-books with a single command, or create free-form drawings on the reader pages. A built-in dictionary allows the user to look up words.

eReader

Formerly Palm Digital Media/Peanut Press

Format: Palm Media

Published as: .pdb

eReader is a freeware program for viewing Palm Digital Media electronic books. Versions are available for iPhone, PalmOS, WebOS, Android, Symbian, BlackBerry, Windows Mobile Pocket PC/Smartphone, desktop Windows, and Macintosh. The reader shows text one page at a time, as paper books do. eReader supports embedded hyperlinks and images. Additionally, the Stanza application for the iPhone and iPod Touch can read both encrypted and unencrypted eReader files.

The company's web site - ereader.com maintains a wide selection of eReader-formatted e-books, available for purchase and download, with a handful of public domain titles available for free. Those books that aren't free are encrypted, with the key being the purchaser's full name and credit card number. This information is not preserved in the e-book. A one-way hash is used, so there is no risk of the user's information being extracted.

The program supports features like bookmarks and footnotes, enabling the user to mark any page with a bookmark, and any part of the text with a footnote-like commentary. Footnotes can later be exported as a Memo document.

The company also offers two Windows/MacOS programs for producing e-books: the Dropbook, which is free, and the eBook Studio, which is not. Dropbook is a file-oriented PML-to-PDB converter; eBook Studio incorporates a WYSIWYG editor. Both programs are compatible with simple text files.

There is also support for an integrated reference dictionary (with many options up to and including a 476,000-word Merriam-Webster Dictionary, including pronunciation keys) so that any word in the text can be highlighted and looked up on the dictionary instantly. Commercial fonts can also be individually purchased and downloaded at the company's web site, ereader.com.

On July 20, 2009, Barnes & Noble announced that the eReader format will be the method they will use to deliver e-books. Updated versions of the Palm Digital programs for Apple iPhone/Touch, Blackberry, Mac OS X, and Windows platforms were made available on the Barnes & Noble eBooks website.

On October 20, 2009, Barnes & Noble announced that their Nook Reader will support the eReader format. eReader format is also supported by the discontinued eSlick, an e-reading device from Foxit Software. It is not currently supported on Barnes & Noble's NookColor.

Desktop Author

Format: DNL Reader
Published as: .dnl; .exe

Desktop Author is an electronic publishing suite that allows creation of digital web books with virtual turning pages. Digital web books of any publication type can be written in this format, including brochures, e-books, digital photo albums, e-cards, digital diaries, online resumes, quizzes, exams, tests, forms and surveys. DesktopAuthor packages the e-book into a ".dnl" or ".exe" book. Each can be a single, plain stand-alone executable file which does not require any other programs to view it. DNL files can be viewed inside a web browser or stand-alone via the *DNL Reader*.

DNL format is an e-Book format, one which replicates the real life alternative, namely page turning Books. The DNL e-Book is developed by DNAML Pty Limited an Australian company established in 1999. A DNL e-Book can be produced using DeskTop Author or DeskTop Communicator.

Newton eBook

Format: Newton eBook
Published as: .pkg

Commonly known as an Apple Newton book; a single Newton package file can contain multiple books (for example, the three books of a trilogy might be packaged together). All systems running the Newton operating system (the most common include the Newton MessagePads, eMates, Siemens Secretary Stations, Motorola Marcos, Digital Ocean Seahorses and Tarpons) have built-in support for viewing Newton books. The Newton package format was released to the public by Newton, Inc. prior to that company's absorption into Apple Computer. The format is thus arguably open and various people have written readers for it (writing a Newton book converter has even been assigned as a university-level class project).

Newton books have no support for DRM or encryption. They do support internal links, potentially multiple tables of contents and indexes, embedded gray scale images, and even some scripting capability (for example, it's possible to make a book in which the reader can influence the outcome). Newton books utilize Unicode and are thus available in numerous languages. An individual Newton book may actually contain multiple views representing the same content in different ways (such as for different screen resolutions).

Founder Electronics

Format: Apabi Reader
Published as: .xeb; .ceb

APABI is a format deviced by Founder Electronics. It is a popular format for Chinese e-books. It can be read using the Apabi Reader software, and produced using Apabi

Publisher. Both .xeb and .ceb files are encoded binary files. The Iliad e-book device includes an Apabi 'viewer'.

Mobipocket

Format: Mobipocket
Published as: .prc; .mobi

The Mobipocket e-book format based on the Open eBook standard using XHTML and can include JavaScript and frames. It also supports native SQL queries to be used with embedded databases. There is a corresponding e-book reader.

The Mobipocket Reader has a home page library. Readers can add blank pages in any part of a book and add free-hand drawings. Annotations — highlights, bookmarks, corrections, notes, and drawings — can be applied, organized, and recalled from a single location. Images are converted to GIF format and have a maximum size of 64K, sufficient for mobile phones with small screens, but rather restrictive for newer gadgets. Mobipocket Reader has electronic bookmarks, and a built-in dictionary.

The reader has a full screen mode for reading and support for many PDAs, Communicators, and Smartphones. Mobipocket products support most Windows, Symbian, BlackBerry and Palm operating systems. Using WINE, the reader works under Linux or Mac OS X. Third-party applications like Okular and FBReader can also be used under Linux or Mac OS X, but they work only with unencrypted files.

The Amazon Kindle's AZW format is basically just the Mobipocket format with a slightly different serial number scheme (it uses an asterisk instead of a Dollar sign), and .prc publications can be read directly on the Kindle.

Mobipocket has developed an .epub to .mobi converter called KindleGen (supports IDPF 1.0 and IDPF 2.0 epub format, according to the company).

Notably, Eastern European letters with diacritical marks are not supported.

EPUB

Format: IDPF/EPUB
Published as: .epub

The EPUB logo.

The .epub or OEBPS format is an open standard for e-books created by the International Digital Publishing Forum (IDPF). It combines three IDPF open standards:

- Open Publication Structure (OPS) 2.0, which describes the content markup (either XHTML or Daisy DTBook)
- Open Packaging Format (OPF) 2.0, which describes the structure of an .epub in XML
- OEBPS Container Format (OCF) 1.0, which bundles files together (as a renamed ZIP file)

Currently, the format can be read by the Kobo eReader, Apple's iBooks app running on iOS devices such as the iPhone and iPad, Barnes and Noble Nook, Sony Reader, BeBook, Bookeen Cybook Gen3 (with firmware v. 2 and up),COOL-ER, Adobe Digital Editions, Lexcycle Stanza, BookGlutton, AZARDI, Aldiko and WordPlayer on Android, Freda on Windows Mobile and Windows Phone 7, and the Mozilla Firefox add-on EPUBReader. Several other reader software programs are currently implementing support for the format, such as dotReader, FBReader, Mobipocket, uBook and Okular. Another software .epub reader, Lucidor, is in beta.

Adobe Digital Editions uses .epub format for its e-books, with DRM protection provided through their proprietary ADEPT mechanism. The recently developed INEPT framework and scripts have been reverse-engineered to circumvent this DRM system.

DSLibris, a Sourceforge.net project, is able to decode e-books in .epub and .xht format for reading on Nintendo DS systems.

Broadband eBooks (BBeB)

Format: Sony media
Published as: .lrf; .lrx

The digital book format used by Sony Corporation. It is a proprietary format, but some reader software for general-purpose computers, particularly under Linux (for example, calibre's internal viewer), has the capability to read it. The LRX file extension represents a DRM encrypted eBook.

SSReader

Format: SSReader
Published as: .pdg

The digital book format used by a popular digital library company 超星数字图书馆 in China. It is a proprietary raster image compression and binding format, with reading time OCR plug-in modules. The company scanned a huge number of Chinese books in the China National Library and this becomes the major stock of their service. The detailed format is not published. There are also some other commercial e-book formats used in Chinese digital libraries.

TealDoc

Format: TealDoc
Published as: .pdb

TealPoint Software's proprietary reader for Palm OS. In addition to its own format, it opens plain text and PalmDoc files. Newer versions of the software include an editor for Palm OS. Embedded images must be converted to TealPoint's proprietary TealPaint format. The format uses HTML like tags for formatting and has been reverse-engineered for 3rd party programs to edit and convert to/from TealDoc format.

IEC 62448

Format: IEC 62448
Published as:

IEC 62448 is an international standard created by International Electrotechnical Commission (IEC)), Technical Committee 100, Technical Area 10 (Multimedia e-publishing and e-book).

The current version of IEC 62448 is an umbrella standard that contains as appendices two concrete formats, XMDF of Sharp and BBeB of Sony. However, BBeB has been discontinued by Sony and the version of XMDF that is in the specification is out of date. The IEC TA10 group is discussing next steps, and has invited the IDPF organization which has standardized EPUB to be a liaison. It is possible that the current version of EPUB and/or the forthcoming EPUB3 revision may be added to IEC 62448. Meanwhile a number of Japanese companies have proposed that IEC standardize a proposed new Japanese-centric file format that is expected to unify DotBook of Voyager Japan and XMDF of Sharp. This new format has not been publicly disclosed as of November, 2010

but it is supposed to cover basic representations for the Japanese language. Technically speaking, this revision is supposed to provide a Japanese minimum set, a Japanese extension set, and a stylesheet language. These issues were discussed in the TC100 meeting held in October 2010 but no decisions were taken besides offering the liaison status to IDPF.

Comic Book Archive file

Format: compressed images
Published as: .cbr (RAR); .cbz (ZIP); .cb7 (7z); .cbt (TAR); .cba (ACE)

A Comic Book Archive file or ComicBook Reader File consists of a series of image files, typically PNG (lossless compression) or JPEG (lossy compression) files, stored as a single archive file, for the purpose of sequential viewing of images, especially comic books. The idea was made popular by the CDisplay image viewer; since then, many viewers for different platforms have been created. Comic Book Archive files are not a distinct file format; only the file name extension differs from a standard file of the given archive type. Some applications support additional tag information (like artists or story information) in the form of embedded XML files in the archive, or use of the Zip comment function.

Multimedia eBooks

Format: Eveda
Published as: .exe or .html

A multimedia ebook is media and book content that utilizes a combination of different book content formats. The term can be used as a noun (a medium with multiple content formats) or as an adjective describing a medium as having multiple content formats.

The 'multimedia ebook' term is used in contrast to media which only utilize traditional forms of printed or text books. Multimedia ebooks include a combination of text, audio, images, video, and/or interactive content formats. Much like how a traditional book can contain images to help the text tell a story, a multimedia ebook can contain other elements not formerly possible to help tell the story.

With the advent of more widespread tablet-like computers, such as the smartphone, some publishing houses are planning to make multimedia ebooks, such as Penguin.

Comparison tables

Features

Format	Filename extension	DRM support	Image support	Table support	Sound support	Interactivity support	Word wrap support	Open standard	Embedded annotation support	Book-marking
Plain text	.txt	No	No	No	No	No	Yes	Yes	No	No
HTML	.html	No	Yes	Yes	No	No	Yes	Yes	No	No
PostScript	.ps	No	Yes	?	No	No	No	Yes	?	?
Portable Document Format	.pdf	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
DjVu	.djvu	?	Yes	Yes	No	No	No	Yes	Yes	Yes
EPUB (IDPF)	.epub	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FictionBook	.fb2	Yes	Yes	?	No	No	Yes	Yes	Yes	?
Mobipocket	.prc, .mobi	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Kindle	.azw	Yes	Yes	Yes[f 1]	Yes[f 2]	No	Yes	No	Yes	Yes
eReader	.pdb	Yes	Yes	?	No	No	Yes	No	Yes	Yes
TealDoc	.pdb	Yes	Yes	?	No	No	Yes	Yes	?	Yes
Broadband eBook	.lrf, .lrx	Yes	Yes	?	No	No	Yes	No	?	?
WOLF	.wol	Yes	Yes	?	No	No	No	No	?	?
Tome Raider	.tr2, .tr3	Yes	Yes	?	No	No	Yes	No	?	?
ArghosReader	.aeh	Yes	Yes	?	No	No	Yes	No	?	Yes
Microsoft Reader	.lit	Yes	Yes	?	No	No	Yes	No	?	Yes
Multimedia EBook	.exe	Yes	Yes	?	Yes	Yes	No	Yes	Yes	Yes
Repligo	.rgo	?	Yes	Yes	No	No	Yes	No	No	No

1. ^ Supported in all except 1st Generation Kindle. (Support level is as it is in mobipocket)
2. ^ Supported only in kindle for iPhone, iPod, iPad.

Supporting Hardware

Hardware Reader	Plain text	PDF	ePub	HTML	Mobi-Pocket	Fiction-Book (Fb2)	DjVu	Broadband eBook (BBeB) [h 1]	eReader [h 1]	Kindle [h 1]	WOLF [h 1]	Tome Raider [h 1]	Open eBook [h 2]
Amazon Kindle 1	Yes	No	No	No	Yes	No	No	No	No	Yes	No	No	No
Amazon Kindle 2, DX	Yes	Yes	No	Yes	Yes	No	No	No	No	Yes	No	No	No
Amazon Kindle 3	Yes	Yes	No	No	Yes	No	No	No	No	Yes	No	No	No
Android Devices	Yes	Yes	Yes	Yes	Yes [h 3]	Yes	Yes [h 3]	No	Yes [h 3]	Yes	No	Yes [h 3]	Yes [h 3]
Apple iOS Devices	Yes	Yes	Yes	Yes	Yes [h 3]	Yes [h 3]	Yes [h 3]	No	Yes [h 3]	Yes [h 3]	No	Yes [h 3]	Yes [h 3]
Azbooka WISEreader	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No
Barnes & Noble Nook	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No	No
Bookeen Cybook Gen3, Opus	Yes	Yes	Yes [h 4]	Yes	Yes [h 4]	Yes [h 5]	No	No	No	No	No	No	Yes
COOL-ER Classic	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No
Foxit eSlick	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No	No
Hanlin e-Reader V3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No
Hanvon WISEreader	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
iRex iLiad	Yes	Yes	Yes	No	Yes	No	Yes	No	No	No	No	No	No
Iriver Story	Yes	Yes	Yes	No	No	Yes [h 3]	Yes [h 3]	No	No	No	No	No	No
Kobo eReader	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No
Nokia N900	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	Yes
NUUTbook 2	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No

OLPC XO, Sugar	Yes	Yes	Yes	Yes	No	No	Yes	No	No	No	No	No	No
Onyx Boox 60	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
Windows PC	Yes	Yes	Yes	Yes	Yes	*?*	Yes	*?*	Yes	Yes[h 6]	*?*	*?*	Yes
Pocketbook 301 Plus, 302, 360°	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
Sony Reader	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No	No	No
Viewsonic VEB612	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Windows Phone 7	Yes	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No

WT

Chapter 4

Digital Edition

A **digital edition** is an online magazine or online newspaper delivered in electronic form which is formatted identically to the print version. Digital editions are often called digital facsimiles to underline the likeness to the print version. Digital editions have the benefit of reduced cost to the publisher and reader by avoiding the time and expense to print and deliver like a paper edition. This format is considered more environmentally friendly due to the reduction of paper and energy use. These editions also often feature interactive elements such as hyperlinks both within the publication itself and to other internet resources, searching and bookmarking, and can also incorporate multimedia such as video or animation to enhance articles themselves or for advertisements. Some delivery methods also include animation and sound effects replicating page turning to further simulate the experience of their print counterparts. However, the popularity of these facsimile digital editions is limited because they provide neither the best reading experience to the customer, nor a viable revenue stream to the publisher. Additionally some publishers are using other electronic publication methods such as RSS to reach out to readers and inform them when new digital editions are available.

Current technologies are generally either reader-based, requiring download of an application and subsequent download of each edition, or browser-based, requiring no application download (such as Adobe Acrobat) and is often Flash-based. Mygazines and Nxtbook Media are among the main technology providers of web-based digital editions. Some application-based readers allow readers to access editions while not connected to the internet. Dedicated hardware such as the Amazon Kindle and the iPad is also available for reading digital editions of select books, popular national magazines such as Relevant, TIME, Atlantic Monthly, and Forbes and popular national newspapers such as the New York Times, Wall Street Journal, and Washington Post. Other E-book manufacturers that deliver digital editions include Plastic Logic and Sony.

Archives of print newspapers, in some cases going back hundreds of years, are being digitized and made available online. Google is indexing existing digital archives produced by the newspapers themselves or by 3rd parties.

Newspaper and magazine archival is not new with microform film formats solving the problem of efficiently storing and preserving though the format lacked accessibility. Many libraries, especially state libraries in the United States are archiving their collections digitally and converting existing microfilm to digital format. The Library of Congress provides project planning assistance and the National Endowment for the Humanities provides funding through grants from its National Digital Newspaper Program.

Digital magazines, ezines, e-editions and emags are sometimes referred to as digital editions but some of these formats are published only in digital format unlike digital editions which replicate a printed edition as well.

Digital magazines

Digital-replica magazines number in the thousands—consumer and business publications, and house magazines for associations, institutions and corporations – and adoption was still increasing as of 2009.

Adoption by publishers accelerated when the circulation-audit bureaus such as BPA to give publishers the same credit for subscribers receiving digital-replica editions as for subscribers receiving print; this concept is being extended to other media reached by a publication's brand.

A 2008 report funded by digital-replica technology providers and auditing agencies counted 1,786 digital-replica editions having more than 7 million circulation among business-to-business publications, of which 230 editions were audited The same report counted 1,470 digital-replica editions of consumer magazines having 5.5 million digital circulation, of which 240 editions were audited. These authors estimated that by yearend 2009 there would be 8,000 digital magazines, having a combined distribution of more than 30 million people

Surveys have shown that, while not all subscribers prefer a digital edition, some do because of the environmental benefit, also because digital magazines are searchable and may easily be passed along or linked to. One such survey funded by a digital publisher reported on inputs from more than 30,000 subscribers to business, consumer and other digital magazines.

Digital magazine business models

Reduced printing and distribution costs

The ability for publishers to save by moving some or all subscribers from print to digital is widely accepted. Oracle magazine, which has 176,000 of its 516,000 subscribers receiving digital according to its June 2009 BPA circulation statement, is said to be the most widely circulated digital edition of a business-to-business publication. Publishers who do this need to choose whether to make some issues all-digital, move some subscribers to digital edition, add some digital-only subscribers, or send all subscribers the digital edition

Paid subscription revenue

In 2009, a major consumer magazine, PC magazine, went all-digital, charging an annual subscription fee for its digital-replica edition

Many consumer magazines and newspapers are already available in eReader formats sold through booksellers. The Barnes and Noble ecommerce site had 1,289 digital magazines available for purchase as of late October 2009.

Sponsorship and advertising revenue

Digital editions often carry special “front cover” advertising, or advertising on the email message alerting the subscriber to the digital edition. Publishers also produce special digital-only inserts and rich-media ads or advertorials.

Designed-for-digital issues

Another approach is to replace entire printed issues with digital ones, or to use digital editions for extra issues that would otherwise have to be printed.

Where to find digital magazines

There are a number of portal sites available that offer a range of digital editions. Most portal sites offer replica editions (digital versions of a print magazine) rather than stand alone digital titles, including Zinio, Emagazines, and Digital Magazine Deals.

Chapter 5

Online Newspaper

An **online newspaper**, also known as a **web newspaper**, is a newspaper that exists on the World Wide Web or Internet, either separately or as an online version of a printed periodical.

Going online created more opportunities for newspapers, such as competing with broadcast journalism in presenting breaking news in a more timely manner. The credibility and strong brand recognition of well-established newspapers, and the close relationships they have with advertisers, are also seen by many in the newspaper industry as strengthening their chances of survival. The movement away from the printing process can also help decrease costs.

Professional journalists have some advantages over blogs, as editors are normally aware of the potential for legal problems.

Online newspapers are much like hard-copy newspapers and have the same legal boundaries, such as laws regarding libel, privacy and copyright, also apply to online publications in most countries, like in the UK. Also in the UK the Data Protection Act applies to online newspapers and news pages. As well as the PCC rules in the UK. But the distinction was not very clear to the public in the UK as to what was a blog or forum site and what was an online newspaper. In 2007, a ruling was passed to formally regulate UK based online newspapers, news audio, and news video websites covering the responsibilities expected of them and to clear up what is, and what isn't, an online publication.

News reporters are being taught to shoot video and to write in the succinct manner necessary for the Internet news pages. Many are learning how to implement blogs and the ruling by the UK's PCC should help this development of the internet.

Journalism students in schools around the world are being taught about the "convergence" of all media and the need to have knowledge and skills involving print, broadcast and web.

Some newspapers have attempted to integrate the internet into every aspect of their operations, i.e., reporters writing stories for both print and online, and classified advertisements appearing in both media; others operate websites that are more distinct from the printed newspaper. The Newspaper National Network LP is an online advertising sales partnership of the Newspaper Association of America and 25 major newspaper companies.

Introduction

In the developing world online publishers are drawing large amounts of traffic and reaping the rewards of online publishing. The Guardian also leads the way with online news with a revolutionary website that trumps many other UK based newspaper websites. The oldest example of an online newspaper or in this case a weekly summary over the weekend's news is The Weekend City Press Review, set up in 1991 this was a pioneer in the online market. Popular in the city, this subscription based service continues to run today. But they are based on hard copy reports and papers. See 'Hybrid newspapers' section of this page. Truly 'Online Only' newspapers and magazines started much later, with the exception of "News Report", an online newspaper created by Bruce Parrello in 1974 on the PLATO system at the University of Illinois.

Examples of newspaper online

It would be difficult to find a daily newspaper in the UK or United States, in fact in the world, in the 21st century, that does not have or share a website.

Very few newspapers in 2006 will claim to have made money from their websites, which are mostly free to all viewers. Declining profit margins and declining circulation in daily newspapers have forced executives to contemplate new methods of obtaining revenue from websites, without charging for subscription. This has been difficult. Newspapers with specialized audiences such as *The Wall Street Journal* or *The Chronicle of Higher Education*, successfully charge subscription fees. Most newspapers now have an online edition, including, *The Los Angeles Times*, *The Washington Post*, *USA Today*, and *The New York Times*.

The Guardian experimented with new media in 2005, offering a free twelve part weekly podcast series by Ricky Gervais. Another UK daily to go online is *The Daily Telegraph*.

In India, major newspapers went online to provide latest and most updated news from them *Times of India*, *Hindustan Times*, *The Hindu*, *Indian Express* and *The New Indian Express*. Some newspapers even provide E-Paper which is regarded as the digital replica of the newspaper.

In Australia, some newspapers corporations offer an online version to let their readers read the news online, such as The Australian, Sydney Morning Herald.

The Santiago Times operates out of Santiago, Chile and is 100% on line, editions are published in English covering Chilean current events daily Monday through Friday..

Online-only newspapers

The true **online only paper** is a paper that does not have any hard copy connections. An example of this is an independent web only newspaper, introduced in the UK in 2000, called the *Southport Reporter*. It is a weekly regional newspaper that is not produced or run in any format other than 'soft-copy' on the internet by its publishers PCBT Photography. Unlike blog sites and other news websites it is run as a newspaper and is recognized by media groups in the UK, like the NUJ and/or the IFJ. Also they fall under the UK's PCC rules. But even print media is turning to online only publication. As of 2009, the collapse of the traditional business model of print newspapers has led to various attempts to establish local, regional or national online-only newspapers - publications that do original reporting, rather than just commentary or summaries of reporting from other publications. An early major example in the U.S. is the Seattle Post-Intelligencer, which stopped publishing after 149 years in March 2009 and went online only. In Scotland in 2010, Caledonian Mercury was set, as Scotland's first online-only newspaper with the same aims as Southport Reporter, in the UK.

In the US, technology news websites such as CNET, TechCrunch, and ZDNet started as web publications and enjoy comparable readership to the conventional newspapers. Also, with the ever-rising popularity of online media, veteran publications like the US News & World Report are abandoning print and going online-only.

Hybrid newspapers

There are some newspapers which are predominantly an online newspaper, but also provide limited hard copy publishing An example is annarbor.com, which replaced the Ann Arbor News in the summer of 2009. It is primarily an online newspaper, but publishes a hardcopy twice a week.

Soft-copy news sheets

A news sheet is a paper that is on one or two pages only. Soft-copy sheets are like online newspapers, in that they have to be predominantly news, not advert or gossip based. These sheets can be updated periodically or regularly, unlike a newspaper. They must also like a newspaper be regarded as a news outlet by media groups and governments.

Future

The development of electronic newspapers, will very soon be supplementing hard-copy printed papers via electronic paper. In February 2006, the Flemish daily *De Tijd* of

Antwerp announced plans to distribute an electronic-ink version of the paper to selected subscribers. This would have been the first such application of electronic ink to newspaper publishing.

Fair use

In a question and answer session, suggestions that Google and the Internet was eroding the intellectual property rights of newspapers was downplayed.

WT

Chapter 6

Open Access (Publishing)

Open Access logo, originally designed by Public Library of Science

Open access (**OA**) refers to unrestricted online access to articles published in scholarly journals, and increasingly also book chapters or monographs.

Open Access comes in two forms, Gratis versus Libre: Gratis OA is no-cost online access, while Libre OA offers some additional usage rights. Open content is similar to OA, but usually includes the right to *modify* the work, whereas in scholarly publishing it is usual to keep an article's content intact and to associate it with a fixed author. Creative Commons licenses can be used to specify usage rights. The Open Access idea can be extended to the learning objects and resources provided in e-learning.

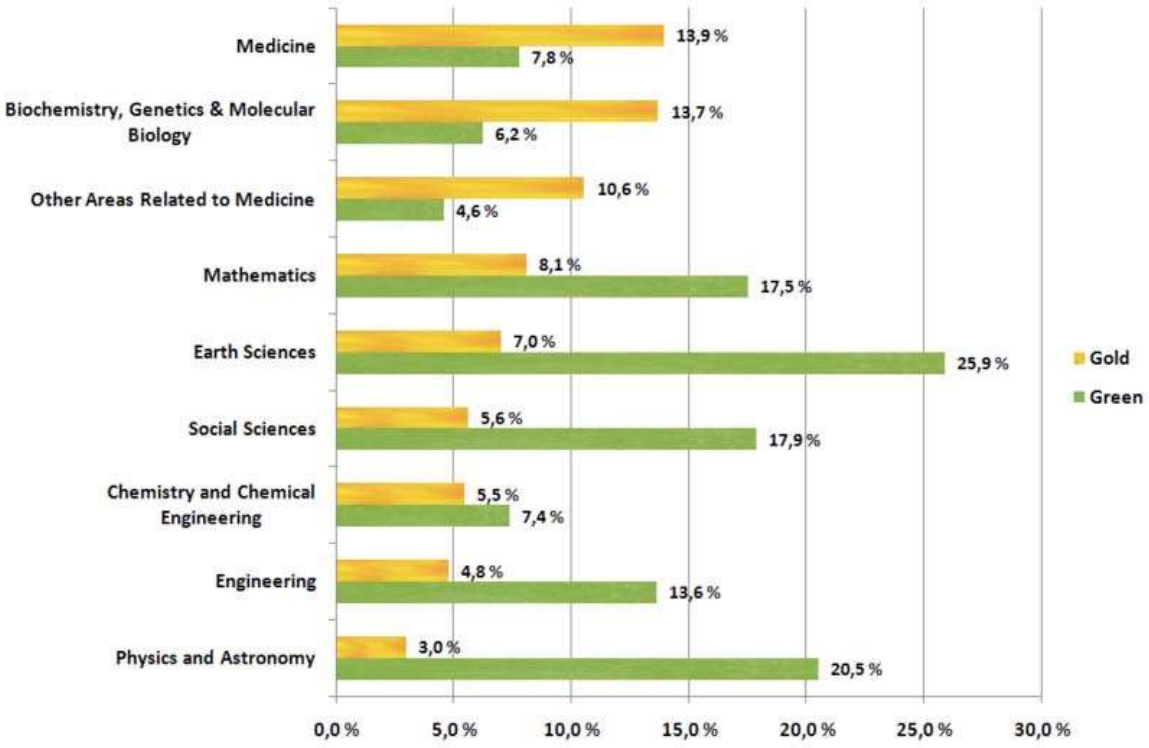
OA can be provided in two ways:

- "Green OA" is provided by authors publishing in any journal and then self-archiving their postprints in their institutional repository or on some other OA website. Green OA journal publishers endorse immediate OA self-archiving by their authors.
- "Gold OA" is provided by authors publishing in an open access journal that provides immediate OA to all of its articles on the publisher's website. (Hybrid open access journals provide Gold OA only for those individual articles for which their authors (or their author's institution or funder) pay an OA publishing fee.)

Public access to the World Wide Web became widespread in the late 1990s and early 2000s. The low-cost distribution technology has fueled the OA movement, and prompted both the Green OA self-archiving of non-OA journal articles and the creation of Gold OA journals. Conventional non-OA journals cover publishing costs through access tolls such as subscriptions, site-licenses or pay-per-view. Some non-OA journals provide OA after an embargo period of 6–12 months or longer. Active debate over the economics and reliability of various ways of providing OA continues among researchers, academics, librarians, university administrators, funding agencies, government officials, commercial publishers, and society publishers.

Adoption statistics

A study published in 2010 showed that of the total output of peer-reviewed articles roughly 20 % could be found Openly Accessible. 8.5 % of the journal literature could be found free at the publishers' sites ("Gold OA"), of which 62 % in full OA journals, 14 % in subscription journals making their electronic versions free after a delay, and 24 % as individually open articles (against payment) in otherwise subscription journals. For an additional 11.9 % of the articles free full text copies were found elsewhere ("Green OA") in either subject-based repositories (43 %), institutional repositories (24%) or on the home pages of the authors or their departments (33%). These copies were further classified into exact copies of the published article (38 %), manuscripts as accepted for publishing (46 %) or manuscripts as submitted (15 %).



Chemistry (13 %) had the lowest overall share of OA of all scientific fields, Earth Sciences (33%) the highest. In medicine, biochemistry and chemistry gold publishing in OA journals was more common than the author posting of manuscripts in repositories. In all other fields author-posted green copies dominated the picture.

Manner of distribution

Like the self-archived Green OA articles, most Gold OA journal articles are distributed via the World Wide Web, due to low distribution costs, increasing reach, speed, and increasing importance for scholarly communication. Open source software is sometimes used for institutional repositories, OA journal websites, and other aspects of OA provision and OA publishing. Gratis OA articles are free online and Libre OA articles have limited copyright and licensing restrictions.

Access to online content requires Internet access, and this distributional consideration presents physical and sometimes financial "barriers" to access. Proponents of OA argue that Internet access barriers are relatively low in many circumstances, that efforts should be made to subsidize universal Internet access, whereas pay-for-access presents a relatively high additional barrier over and above Internet access itself.

OA can be provided by traditional publishers, or under other arrangements. Some OA publishers, such as Public Library of Science (PLoS), publish only OA journals; others publish OA as well as subscription-based journals.

Methods of financing gold OA publishing

Advertising is a major source of funding for mass media that do not charge for content, as well as modern web sites and search engines.

In scholarly publishing, there are many business models for OA journals. Some charge publication fees (paid by authors or by their funding agencies or employers) and some do not. Some of the no-fee journals have institutional subsidies and some do not.

Roughly half the Gold OA journals have author fees to cover the cost of publishing (e.g. PLoS fees vary from $1,300 to $2,850) instead of reader subscription fees. Advertising revenue and/or funding from foundations and institutions are also used to provide funding.

Authors and researchers

The main reason authors make their articles openly accessible is to maximize their research impact. A study in 2001 first reported an OA citation impact advantage, and a growing number of studies have confirmed, with varying degrees of methodological rigor, that an OA article is more likely to be used and cited than one behind subscription barriers. For example, a 2006 study in *PLoS Biology* found that articles published as immediate open access in *PNAS* were three times more likely to be cited than non-open access papers, and were also cited more than *PNAS* articles that were only self-archived. This result has been challenged as possibly due to authors self-selectively making higher quality articles OA, but a recent study comparing self-selected OA with mandated OA found that the citation advantage remained just as big when the OA was mandated.

Scholars are paid by research funders and/or their universities to do research; the published article is the report of the work they have done, rather than an item for commercial gain. The more the article is used, cited, applied and built upon, the better for research as well as for the researcher's career. Similarly, the more *quickly* it is accessible, the better; open access can reduce publication delays, an obstacle which led many research fields to traditions of widespread preprint access.

Some professional organizations have encouraged use of OA: In 2001, the International Mathematical Union communicated to its members that "Open access to the mathematical literature is an important goal" and encouraged them to "[make] available electronically as much of our own work as feasible" to "[enlarge] the reservoir of freely available primary mathematical material, particularly helping scientists working without adequate library access."

Authors who wish to make their work openly accessible have two options. One option is to publish in an OA journal ("Gold OA"). An open access journal may or may not charge a processing fee; open access publishing does not necessarily mean that the author has to pay. Traditionally, many academic journals levied page charges, long before open access became a possibility. When OA journals do charge processing fees, it is the author's

employer or research funder who typically pays the fee, not the individual author, and many journals will waive the fee in cases of financial hardship, or for authors in less-developed countries.

The other option is author self-archiving ("Green OA"). To find out if a publisher or journal has given a green light to author self-archiving, the author can check the Publisher Copyright Policies and Self-Archiving list on the SHERPA RoMEO web site. To find out by journal, the author can check the EPrints Romeo site, which is derived from the SHERPA/RoMEO dataset. The EPrints site itself also provides a FAQ on self-archiving. Extensive details and links can also be found in the Open Access Archivangelism blog and the Eprints Open Access site.

While open access is currently focused on scholarly research articles, any content creators can now decide how to make their content available and, if they wish, they can share their work openly. Creative Commons provides a number of licenses with which authors may easily indicate which uses are allowed.

Users

For the most part, the direct users of research articles are other researchers. Open access helps researchers as readers by opening up access to articles that their libraries do not subscribe to. One of the great beneficiaries of open access may be users in developing countries,where currently some universities find it difficult to pay for subscriptions required to access the most recent journals. Some schemes exist for providing subscription scientific publications to those affiliated to institutions in developing countries at little or no cost. All researchers benefit from OA as no library can afford to subscribe to every scientific journal and most can only afford a small fraction of them – this is known as the serials crisis".

Open access extends the reach of research beyond its immediate academic circle. An OA article can be read by anyone – a professional in the field, a researcher in another field, a journalist, a politician or civil servant, or an interested hobbyist. Indeed, a 2008 study revealed that mental health professionals are roughly twice as likely to read a relevant article if it is freely available.

The Directory of Open Access Journals lists a number of peer-reviewed open access journals for browsing and searching. Open J-Gate is another index of articles published in English language OA journals, peer reviewed and otherwise, which launched in 2006. Open access articles can also often be found with a web search, using any general search engine or those specialized for the scholarly/scientific literature, such as OAIster and Google Scholar. Results may include preprints that have not yet been peer reviewed, or gray literature that will remain unreviewed.

Research funders and universities

Research funding agencies and universities want to ensure that the research they fund and support in various ways has the greatest possible research impact.

Research funders are beginning to expect open access to the research they support. Forty-two of them (including all seven UK Research Councils) have already adopted Green OA self-archiving mandates, and four more (including two in the US) have proposed to adopt mandates.

Canada's Social Sciences and Humanities Research Council, which made a commitment to open access in October 2004, has not yet adopted or proposed a mandate but the Canadian Institutes of Health Research (CIHR) proposed a mandate in 2006 and adopted it in September 2007, the first North American public research funder to do so.

In May 2006, the US Federal Research Public Access Act (FRPAA) was proposed toward improving the NIH Public Access Policy. Besides points about making open access mandatory, to which the NIH complied in 2008, it argues to extend self-archiving to the full spectrum of major US-funded research. In addition, the FRPAA would no longer stipulate that the self-archiving must be central; the deposit can now be in the author's own institutional repository (IR). The new U.S. National Institutes of Health's Public Access Policy took effect in April 2008 and states that "all articles arising from NIH funds must be submitted to PubMed Central upon acceptance for publication". It stipulates self-archiving in PubMed Central rather than in the author's own institutional repository, which some consider a strength and others a weakness.

The Canadian Institutes of Health Research (CIHR) Policy on Access to Research Outputs provides a number of options to researchers, including publication in open access journals, or making their manuscripts available in an online repository such as PubMed Central Canada.

In April 2006, the European Commission recommended: «EC Recommendation A1 : "Research funding agencies... should [e]stablish a European policy mandating published articles arising from EC-funded research to be available after a given time period in open access archives...». This recommendation has since been updated and strengthened by the European Research Advisory Board (EURAB).
The OpenAIRE (Open Access Infrastructure for Research in Europe) project has hence been started. The EC Open Access pilot covers about 20 % of the budget of the Seventh Research Framework Programme.

To somewhat improve on the EC's (and FRPAA's) allowable embargo (of up to six months), EURAB has revised the mandate: all articles must be deposited immediately upon acceptance: the allowable delay applies only to the time when access to the deposit must be made open access rather than to the time when it must be deposited. This is intended to permit individual users to use an eprint request "email eprint" button found on some archives to send a semi-automatic email message to the author requesting an

individual eprint during the embargo period: This is not open access, but in the view of at least some advocates it provides for some needs during any embargo, and might help hasten the demise of embargoes altogether, while facilitating the adoption of self-archiving mandates by funders and universities.

A growing number of universities are providing institutional repositories in which their researchers can deposit their published articles. Eighty-six individual universities and eighteen faculties and departments have already adapted self-archiving mandates (including Harvard, MIT, Stanford, U. College London, U. Edinburgh) and ten further individual multi-university mandates (in Europe and Brazil) have been proposed. Eprints maintains a Registry of OA Repository Material Archiving Policies (ROARMAP). and EnablingOpenScholarship (EPS) provides universities with OA policy-building.

In May 2005, 16 major Dutch universities cooperatively launched DAREnet, the Digital Academic Repositories, making over 47,000 research papers available to anyone with internet access. From 1 January 2007, at the completion of the DARE programme, KNAW Research Information has taken over responsibility for the DAREnet portal. On 2 June 2008, DAREnet has been incorporated into the scholarly portal NARCIS. At the end of 2009 NARCIS provides access to 185.000 open access publications from all Dutch universities, KNAW, NWO and a number of scientific institutes.

Public and advocacy

Open access to scholarly research is argued to be important to the public for a number of reasons. One of the arguments for public access to the scholarly literature is that most of the research is paid for by taxpayers through government grants, who therefore have a right to access the results of what they have funded. This is one of the primary reasons for the creation of advocacy groups such as The Alliance for Taxpayer Access in the US. Examples of people who might wish to read scholarly literature include individuals with medical conditions (or family members of such individuals) and serious hobbyists or 'amateur' scholars who may be interested in specialized scientific literature (e.g. amateur astronomers). Additionally, professionals in many fields may be interested in continuing education in the research literature of their field, and many businesses and academic institutions cannot afford to purchase articles from or subscriptions to much of the research literature that is published under a toll access model.

Even those who do not read scholarly articles benefit indirectly from open access. For example, patients benefit when their doctor and other health care professionals have access to the latest research. As argued by open access advocates, open access speeds research progress, productivity, and knowledge translation. Every researcher in the world can read an article, not just those whose library can afford to subscribe to the particular journal in which it appears. Faster discoveries benefit everyone. High school and junior college students can gain the information literacy skills critical for the knowledge age. Critics of the various open access initiatives point out that there is little evidence that a significant amount of scientific literature is currently unavailable to those who would benefit from it. While no library has subscriptions to every journal that might be of

benefit, virtually all published research can be acquired via interlibrary loan. Note that interlibrary loan may take a day or weeks depending on the loaning library and whether they will scan and email, or mail the article. Open Access online, by contrast is faster, often immediate, making it more suitable than interlibrary loan for high paced research.

Due to the benefits of open access, many governments are considering whether or not to mandate open access to publicly funded research. However, some organizations representing publishers, such as the DC Principles group in the United States, feel that such mandates are an unwarranted governmental intrusion in the publishing marketplace. Lobbying on both sides is fierce, both for pro-OA and contra-OA.

In developing nations, open access archiving and publishing acquires a unique importance. Scientists, health care professionals, and institutions in developing nations often do not have the capital necessary to access scholarly literature, although schemes exist to give them access for little or no cost. Among the most important is HINARI, the Health InterNetwork Access to Research Initiative, sponsored by the World Health Organization. HINARI, however, also has restrictions. For example, individual researchers may not register as users unless their institution has access, and several countries that one might expect to have access do not have access at all (not even "low-cost" access) (e.g. South Africa).

Many open access projects involve international collaboration. For example the SciELO (Scientific Electronic Library Online), is a comprehensive approach to full open access journal publishing, involving a number of Latin American countries. Bioline International, a non-profit organization dedicated to helping publishers in developing countries is a collaboration of people in the UK, Canada, and Brazil; the Bioline International Software is used around the world. Research Papers in Economics (RePEc), is a collaborative effort of over 100 volunteers in 45 countries. The Public Knowledge Project in Canada developed the open source publishing software Open Journal Systems (OJS), which is now in use around the world, for example by the African Journals Online group, and one of the most active development groups is Portuguese.

A 2004 study of open access publishing by Kristin Antelman found that in philosophy, political science, electrical and electronic engineering and mathematics, open access papers had a greater research impact.

Libraries and librarians

Many librarians have been vocal and active advocates of open access. These librarians believe that open access promises to remove both the *price barriers* and the *permission barriers* that undermine library efforts to provide access to the journal literature. Many library associations have either signed major open access declarations, or created their own. For example, the Canadian Library Association endorsed a Resolution on Open Access in June 2005. Librarians also educate faculty, administrators, and others about the benefits of open access. For example, the Association of College and Research Libraries of the American Library Association has developed a Scholarly Communications Toolkit.

The Association of Research Libraries has documented the need for increased access to scholarly information, and was a leading founder of the Scholarly Publishing and Academic Resources Coalition (SPARC).

There is question, however, as to the extent to which Open Access will solve the serials crisis. In a Nature Web Focus forum, The Pros and Cons of Open Access, Kate Worlock discusses whether Open Access is truly the answer to the crisis, or if it is simply an ends to a means in a world with shrinking library budgets. The argument from the publisher is that while the cost of publications have "undisputedly [sic] risen more sharply than the library budgets," the library budget is too small of a portion of the university's (in this example)overall budget at roughly 2%.

At most universities, the library houses the institutional repository, which provides free access to scholarly work of the university's faculty. Some open access advocates believe that institutional repositories will play a very important role in responding to open access mandates from funders. The Canadian Association of Research Libraries has a program to develop institutional repositories at all Canadian university libraries.

An increasing number of libraries provide hosting services for open access journals. A recent survey by the Association of Research Libraries found that 65% of surveyed libraries either are involved in journal publishing, or are planning to become involved in the very near future.

History

The roots of the concept of open access can be found in the distant past, from the very beginnings of publishing, re-emerging with every innovation in publishing technology. The printing press allowed the written word to be printed and distributed, thereby extending literacy to the population at large. Moving from vellum to paper made it possible to print more cheaply. The invention of the postal system provided a means of widespread distribution.

The beginnings of the scholarly journal were a way of expanding low-cost access to scholarly findings. Many individuals anticipated the open access concept long before modern low-cost distribution methods. One early proponent was the physicist Leo Szilard. To help stem the flood of low-quality publications, he jokingly suggested in the 1940s that at the beginning of his career each scientist should be issued with 100 vouchers to pay for his papers. The Common Knowledge project was an attempt to share information for the good of all, the brainchild of Brower Murphy, formerly of The Library Corporation. Brower and Common Knowledge are recognised in the Library Microcomputer Hall of Fame.

The modern Open Access movement (as a social movement) traces its history at least back to the 1960s, but became much more prominent in the 1990s with the advent of the Digital Age. With the spread of the Internet and the ability to copy and distribute electronic data at no cost, the arguments for open access gained new importance.

Probably the earliest book publisher to provide open access was the National Academies Press, publisher for the National Academy of Sciences, Institute of Medicine, and other arms of the National Academies. They have provided free online full-text editions of their books alongside priced, printed editions since 1994, and assert that the online editions promote sales of the print editions. As of June 2006 they had more than 3,600 books up online for browsing, searching, and reading.

An explosion of interest and activity in open access journals has occurred since the 1990s, largely due to the widespread availability of Internet access. It is now possible to publish a scholarly article and *also* make it instantly accessible anywhere in the world where there are computers and Internet connections. The fixed cost of producing the article is separable from the minimal marginal cost of the online distribution.

These new possibilities emerged at a time when the traditional, print-based scholarly journals system was in a crisis. The number of journals and articles produced has been increasing at a steady rate; however the average cost per journal has been rising at a rate far above inflation for decades, and budgets at academic libraries have remained fairly static. The result was decreased access - ironically, just when technology has made almost unlimited access a very real possibility, for the first time. Libraries and librarians have played an important part in the open access movement, initially by alerting faculty and administrators to the serials crisis. The Association of Research Libraries developed the Scholarly Publishing and Academic Resources Coalition (SPARC), in 1997, an alliance of academic and research libraries and other organizations, to address the crisis and develop and promote alternatives, such as open access.

The first online-only, free-access journals (eventually to be called "open access journals") began appearing in the late 1980s. Among them was *Bryn Mawr Classical Review*, *Postmodern Culture* and *Psycoloquy*.

The first free scientific online archive was arXiv.org, started in 1991, initially a preprint service for physicists, initiated by Paul Ginsparg. Self-archiving has become the norm in physics, with some sub-areas of physics, such as high-energy physics, having a 100% self-archiving rate. The prior existence of a "preprint culture" in high-energy physics is one major reason why arXiv has been successful. arXiv now includes papers from related disciplines, such as computer science and mathematics, but computer scientists mostly self-archive on their own websites and have been doing so for even longer than physicists. (Citeseer is a computer science archive that harvests, Google-style, from distributed computer science websites and institutional repositories and contains almost twice as many papers as arxiv.) arXiv now includes postprints as well as preprints. The two major physics publishers (American Physical Society and Institute of Physics Publishing) have reported that arXiv has had no effect on journal subscriptions in physics; even though the articles are freely available, usually before publication, physicists value their journals and continue to support them.

The inventors of the Internet and the Web -- computer scientists—had been self-archiving on their own FTP sites and then their websites since even earlier than the physicists, as

was revealed when Citeseer began harvesting their papers in the late 1990s. The 1994 "Subversive Proposal" was to extend self-archiving to all other disciplines; from it arose CogPrints (1997) and eventually the OAI-compliant generic GNU Eprints.org software in 2000.

In 1997, the U.S. National Library of Medicine (NLM) made Medline, the most comprehensive index to medical literature on the planet, freely available in the form of PubMed. Usage of this database increased a hundredfold when it became free, strongly suggesting that prior limits on usage were impacted by lack of access. While indexes are not the main focus of the open access movement, free Medline is important in that it opened up a whole new form of use of scientific literature - by the public, not just professionals.

In 1998, the American Scientist Open Access Forum was launched (and first called the "September98 Forum"). The *Journal of Medical Internet Research* (*JMIR*), one of the first Open Access journals in medicine, was created in 1998, publishing its first issue in 1999.

In 1999, Harold Varmus of the NIH proposed a journal called E-biomed, intended as an open access electronic publishing platform combining a preprint server with peer-reviewed articles. E-biomed later saw light in a revised form as PubMed Central, a postprint archive.

It was also in 1999 that the Open Archives Initiative and its OAI-PMH protocol for metadata harvesting was launched in order to make online archives interoperable.

In 2000, BioMed Central, a for-profit open access publisher, was launched by the then Current Science Group (the founder of the *Current Opinion* series, and now known as the Science Navigation Group). In some ways, BioMed Central resembles Harold Varmus' original E-biomed proposal more closely than does PubMed Central. BioMed Central now publishes over 170 journals.

In 2001, 34,000 scholars around the world signed "An Open Letter to Scientific Publishers", calling for "the establishment of an online public library that would provide the full contents of the published record of research and scholarly discourse in medicine and the life sciences in a freely accessible, fully searchable, interlinked form". Scientists signing the letter also pledged not to publish in or peer-review for non-open access journals. This led to the establishment of the Public Library of Science, an advocacy organization. However, most scientists continued to publish and review for non-open access journals. PLoS decided to become an open access publisher aiming to compete at the high quality end of the scientific spectrum with commercial publishers and other open access journals, which were beginning to flourish. Critics have argued that, equipped with a $10 million grant, PLoS competes with smaller OA journals for the best submissions and runs danger to destroy what it originally wanted to foster.

The *first major international* statement on open access was the Budapest Open Access Initiative in February 2002, launched by the Open Society Institute . This provided a definition of open access, and has a growing list of signatories. Two further statements followed: the Bethesda Statement on Open Access Publishing in June 2003 and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities in October 2003.

In 2003, the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities was drafted and the World Summit on the Information Society included open access in its Declaration of Principles and Plan of Action.

In 2006, a Federal Research Public Access Act was introduced in US Congress by senators John Cornyn and Joe Lieberman. The act continues to be brought up every year since then, but has never made it past committee.

The idea of mandating self-archiving was mooted at least as early as 1998. Since 2003 efforts have been focused on open access mandating by the funders of research: governments, research funding agencies, and universities. These efforts have been fought by the publishing industry. However, many countries, funders, universities and other organizations have now either made commitments to open access, or are in the process of reviewing their policies and procedures, with a view to opening up access to results of the research they are responsible for.

One of the many librarians involved in advocating the self-archiving approach to open access is Hélène Bosc; her work can be found in her "15-year retrospective".

Criticism

Opponents of the open access model assert that the pay-for-access model is necessary to ensure that the publisher is adequately compensated for their work. Scholarly journal publishers that support pay-for-access claim that the "gatekeeper" role they play, maintaining a scholarly reputation, arranging for peer review, and editing and indexing articles, require economic resources that are not supplied under an open access model, though acknowledging that open access journals do provide peer review. The cost of paper publication may also make open access to paper copies infeasible. Opponents claim that open access is not necessary to ensure fair access to developing nations; differential pricing, or financial aid from developed countries or institutions can make access to proprietary journals affordable. Conventional journal publishers may also lose customers to open access publishers who compete with them. The Partnership for Research Integrity in Science and Medicine (PRISM), a lobbying organization formed by the Association of American Publishers (AAP), is opposed to the open access movement. PRISM and AAP have lobbied against the increasing trend amongst funding organizations to require open publication, describing it as "government interference" and a threat to peer review.

Textbook publishers generally make an even greater investment in the editing process, and electronic textbooks have yet to become widely accepted. For researchers, publishing

an article describing novel results in a reputable scientific journal usually does more to enhance one's reputation among scientific peers, and advance one's academic career. Journal article authors are generally not directly financially compensated for their work beyond their institutional salaries and the indirect benefits that an enhanced reputation provides in terms of institutional funding, job offers, and peer collaboration. It could be argued, then, that the financial reward from writing a successful textbook is an important motivating factor, without which the quality and quantity of available textbooks would decrease.

There are those, for example PRISM, who think that open access is unnecessary or even harmful. It has been argued that there is no need for those outside major academic institutions to have access to primary publications, at least in some fields.

In the entertainment industry, it is argued that, unlike science, there is no pressing social need for widespread and barrier-free access to the content.

One argument against Open Access is highlighted in a Nature (a for-profit publication) Web Focus forum, The Pros and Cons of Open Access. One argument brought up in the forum is that the supposed tax-payer right to access is blown out of proportion by the advocates of Open Access. Kate Worlock, the author of the forum article argues, "...where research is publicly-funded, taxes are generally not paid so that taxpayers can access research results, but rather so that society can benefit from the results of that research; in the form of new medical treatments, for example. Publishers claim that 90% of potential readers can access 90% of all available content through national or research libraries, and while this may not be as easy as accessing an article online directly it is certainly possible." The argument for tax-payer funded research is only applicable in certain countries as well. For instance in Australia, 80% of research funding comes through taxes, whereas in Japan and Switzerland, only approximately 10% is from the public coffers.

Funding issues

The "article processing charges" for open access shifts the burden of payment from readers to authors, which could conceivably create a new set of concerns. For example, budget processes may need adjustments to provide funding for the "article processing charges" required to publish in almost all open access journals (e.g. those published by BioMed Central). Unless discounts are available to authors from countries with low incomes or external funding is provided to cover the cost, article processing charges could exclude authors from developing countries or less well-funded research fields from publishing in open access journals. However, under the traditional model, the prohibitive cost of non-open access journal subscriptions would preclude conducting any research in the first place. Moreover, many open access publishers offer discounts or publishing fee waivers to authors from developing countries or those suffering financial hardship. Self-archiving of non-OA publications also provides a low cost alternative model.

Outside of science and academia, it is unusual for producers of creative output to be financially compensated on anything other than a pay-for-access model. (Notable exceptions include open source software and public broadcasting.) Successful writers, for example, support themselves by the revenues generated by people purchasing copies of their works; publishing houses are able to finance the publication of new authors based on anticipated revenues from sales of those that are successful. Opponents of open access would argue that without direct financial compensation via pay-for-access, many authors would be unable to afford to write, though some would accept the economic hardship of holding down a day job while continuing to write as a "labor of love". However, this argument has no relevance to academic publishing, because scientific journals do not pay royalties to article authors.

Citation study

A study published in the British Medical Journal disputes the claim that open access articles equal more citations. In the study, researchers from Cornell University randomly made some journal articles freely available while keeping others available by subscription only in order to determine whether increased access to journal articles results in more article downloads and citations. They found, in an interim analysis, that in the first year after the articles were published, open-access articles were downloaded more but were no more likely to be cited than subscription-based articles. However, many responses to the paper argue that the interim analysis was premature.

Comparison with other media

Many traditional media such as certain newspapers, television, and radio broadcasts could be considered "open access". These include commercial broadcasting and free newspapers supported by advertising, public broadcasting, and privately funded political advocacy materials. Minor barriers are also present in other media: broadcast media require receiving equipment, online content requires Internet access, and locally distributed printed media requires transportation to a distribution point.

Many other types of material can also be published in this manner: magazines and newsletters, e-text or other e-books, music, fine arts, or any product of intellectual activity.

Chapter 7

Biblical Software

Biblical software or **Bible software** is a group of computer applications designed to view and study biblical texts and concepts. Biblical software programs are similar to e-book readers in that they include digitally-formatted books, may be used to display a wide variety of inspirational books and bibles, and can be used on portable computers. However, biblical software is geared more toward word and phrase searches, accessing study bible notes and commentaries, referencing various modern translations, cross-referencing similar passages and topics, biblical dictionaries, original language texts and language tools, maps, charts, and other e-books deemed relevant to understanding texts from a philological approach.

Bible software varies in complexity and depth, depending on the needs of users, just as the purposes of the users vary from devotional reading and personal study to lesson and sermon preparation, inspirational publishing and even further research tools and translations. Basic bible software is typically aimed at mobile phones, and is designed simply display the text of a single Bible translation, with word and phrase searches as the only available tool. More advanced packages run on personal computers and boast far more features, display a wider variety of theological resources (see above), and may offer features such as synopses and harmonies of the Gospel narratives, morphological and syntactical searches of original texts, sentence diagramming, user notes, manual and dynamic highlighting, lectionary viewers, etc.

History

Interest in using computers to quickly search the Bible and copy sections of the text quickly into lessons and sermons emerged in the early 1980s. Bible software was much faster than traditional study tools in a book forms. Early bible software was aimed simply at word and phrase searches in different modern translations. Later, as computers improved in handling foreign language fonts, the original Hebrew Old Testament and Koine Greek New Testament texts of the Bible were added. When working with the

original biblical languages, one of the first capabilities was morphology or parsing, providing information on the parts of speech of various words to assist in understanding the intent of the text. At this point many bible software programs emerged which are still in publication today.

Library building

Most Bible software publishers offer a variety of initial packages from basic, to intermediate, to advanced levels, ranging in price from a free, to well over the price of the computer it runs on. Bible software producers commonly offer customers expandibility—that users can build on their initial monetary investment with the purchase of additional resources such as dictionaries, commentaries, translations, and other inspirational books. Initial packages normally include many bundled works, while add-on titles represent a more significant investment. Normally, the advanced packages include all the features of the more basic packages, though a customer may stand to benefit from two or more bundles by purchasing packages from different publishers—especially those which work seamlessly in the same format. For instance a user purchasing a package from LOGOS Bible Software, which runs on a Libronix DLS format, could also buy a Thomas Nelson package, which runs in the same format, and all works would be integrated. Similarly, STEP compatible resources from different publishers also could be combined. Olive Tree Bible Software allows several Bibles and other study tools to be combined into a Library study system.

Desktop Bible software

Windows

- **Bible Analyzer –** A freeware Bible study and analysis application with Advanced Searching, Bible Statistics, Parallel Text Generator, Text-To-Speech, Audio, and more. Immediate download of premium modules available.
- **BiblePRO –** Uses an interface reminiscent of a programming IDE. Light version can be downloaded for free. CD with several modules can be purchased for nominal fee.
- **BibleWorks** Orientated towards Greek and Hebrew studies in Seminaries;
- e-Sword – The most popular free Bible software. Includes STEP Library compatibility with a pane and tab system similar to *Pradis*;
- **iLumina –** Animated bible;
- **Interlinear Scripture Analyzer –** (V2.0 basic) Greek/Hebrew Interlinear bible. Based on NA26/27 with Robinson morphology and WLC Hebrew text. This also includes a literal English translation and Strong's Concordance;
- **LOGOS Bible Software –** Producer of the Libronix Digital Library System, or Libronix DLS, which they also license to other publishers. Through the emergence of the predecessor format, the LOGOS Library system, and the growth in the market share which followed, LOGOS is known for advanced tools for Greek and Hebrew studies and allows users to build a virtual library from a variety of publishers which work together seamlessly. Dynamic linking ties works

together to the same biblical text, even as a user moves through the works. The system also provides topical linking; a user accessing a different dictionary will find it already opened to the same topic they chose in another work. LOGOS packages include helps with harmonies, parallel passages, and even offers such things as automatic citation of references, such as APA, MLA, etc., depending on the preference.
    - **eBible library** – A set of Thomas Nelson titles operating within the Libronix DLS, and/or supplementing users' other titles in the LOGOS/Libronix Digital Library System. While Thomas Nelson once launched the New King James Version translation without collaboration, they have demonstrated the ability to market their biblical software titles in the popular Libronix DLS format, while avoiding the perception of a format war.
    - **Pradis** – This was Zondervan's Bible Study Software File Format, until they partnered with LOGOS Bible Software.
    - **SESB** – (Stuttgart Elektronische Studienbibel) A resource library from Deutsche Bibelgesellschaft for the Libronix DLS. This is the only library that offers the critical apparatuses of BHS and NA27 / UBS4.
- Online Bible – Bible Software Package around since 1987. Huge library of material of both English and other languages;
- **PC Study Bible** – Complimented by reviewers for ease to use. Handles Greek and Hebrew studies well;
- **QuickVerse** – This was the first organization to utilize the STEP file format. It remains the only software to exclusively use that file format;
- **SwordSearcher** – An inexpensive but simple bible study program. Includes mostly public domain modules. Not suitable for original languages study.
- **The Word** – free comprehensive Bible study software with features including Bibles in many languages, non-Bible study resources (including maps), rich cross-referencing system, easy user-created and edited modules (with image support, comprehensive hyperlinking, etc.), a "clipboard monitor" to display verse references, full library search capabilities, extensive user preferences, importing modules from other formats, user defined window layouts, and comprehensive copy and paste capabilities;
- **WORDsearch** – Offers a large library of English resources for free and for additional purchase. Very intuitive interface. Greek, Hebrew and other languages are searchable.
    - **Bible Explorer** – Free, light version of WORDsearch.

Mac OSX

- Accordance – It is well known for its extensive library collection, extensive tools, ease of use, affordability and reliability. Accordance dynamically links ties the Biblical texts to other user works in a library, a user accessing various dictionaries or commentaries are able to search in another text, parallel passages, harmonies, automatic citation for bibliographies, complex to simple search techniques, maps, timelines, graphic resources, are all available. Charts, graphs, diagrams, and

powerful analytical tools are all available in Accordances powerful search tools. Generally considered to be the high-end Bible Study software for Mac
- iLumina – Animated bible;
- LOGOS Bible Software – Mac version of Logos, first released December 2008;
- WORDsearch – Offers a large library of English resources. Can search Greek, Hebrew and other languages;

Linux/Unix

- **Bible Analyzer** – A freeware Bible study and analysis application with Advanced Searching, Bible Statistics, Parallel Text Generator, Text-To-Speech, Audio, and more. Immediate download of premium modules available;
- The SWORD Project

Programming API

- The SWORD Project A comprehensive open source multi-lingual Bible project. The project is a programming platform and a collection of resources. Several front end programs are available for Windows, Linux, Mac OSX, various PDAs, etc.
    - Alkitab Bible Study – Cross-platform front end that is based on Netbeans platform;
    - Bible Desktop – Cross platform front end that is built in Java;
    - BibleTime – Front end that is built on the Qt framework for the Windows, Linux, and BSD operating systems;
    - BPBible – Front end for Windows;
    - FireBible – Front end that installs as a Firefox extension;
    - MacSword – Front end for the Mac OS X;
    - PocketSword — Front end for the iPhone;
    - SwordBible – Simple Bible reading software for Windows;
    - SwordReader — Front end for Windows Mobile;
    - The SWORD Project for Windows – Front end for Windows;
    - Xiphos – Front end based on GTK for Windows, Linux, and BSD platforms;

Mobile platform

iPad tablet

- BibleReader – Olive Tree Bible Software

iPhone and iPod Touch

- Accordance – OakTree Software
- BibleReader – Olive Tree Bible Software
- Logos for the iPhone – Logos Bible Software
- PocketSword – Front end for the iPhone

- Good News Bible — Full GNB/TEV text with integrated dictionary, timeline & biblical characters.

Google Android

- and-bible – JSWORD frontend for Android
- BibleReader – Olive Tree Bible Software

Blackberry

- BibleReader – Olive Tree Bible Software

PalmOS

- MyBible: Laridian
- BibleReader – Olive Tree Bible Software (no longer supported)
- Palm Bible Plus – open source fork of *Bible Reader for Palm*

Windows Mobile (formerly Windows CE)

- Bible Pronto – A free bible app for windows phone 7
- BibleReader – Olive Tree Bible Software (no longer supported)
- Pocket E-Sword – in December 2009, it was announced that official support had ceased
- eBible for Pocket PC – this no longer supported by Thomas Nelson Publishers
- BibleSurfer – free download
- SwordReader – Front end for Windows Mobile. Support is available, but development has ceased

Java

- Alkitab Bible Study – Cross-platform front end that is based on Netbeans platform and uses The Sword Project API
- Bible Desktop – Cross platform front end that is built in Java, and uses The Sword Project API
- Go Bible – for JavaME mobile phones

Online tools

- Blue Letter Bible
- Christian Classics Ethereal Library — Defined as a Christian library (like e.g. Libronix). Bible study interface with various bible versions and commentaries available.
- e-Sword Live
- Bible Forge — A free, web based Bible study program supporting KJV only.

Chapter 8

Blog

A **blog** (a blend of the term ***web log***) is a type of website or part of a website. Blogs are usually maintained by an individual with regular entries of commentary, descriptions of events, or other material such as graphics or video. Entries are commonly displayed in reverse-chronological order. *Blog* can also be used as a verb, meaning *to maintain or add content to a blog*.

Most blogs are interactive, allowing visitors to leave comments and even message each other via widgets on the blogs and it is this interactivity that distinguishes them from other static websites.

Many blogs provide commentary or news on a particular subject; others function as more personal online diaries. A typical blog combines text, images, and links to other blogs, Web pages, and other media related to its topic. The ability of readers to leave comments in an interactive format is an important part of many blogs. Most blogs are primarily textual, although some focus on art (art blog), photographs (photoblog), videos (video blogging), music (MP3 blog), and audio (podcasting). Microblogging is another type of blogging, featuring very short posts.

As of 16 February 2011, there were over 156 million public blogs in existence.

History

The term "weblog" was coined by Jorn Barger on 17 December 1997. The short form, "blog," was coined by Peter Merholz, who jokingly broke the word *weblog* into the phrase *we blog* in the sidebar of his blog Peterme.com in April or May 1999. Shortly thereafter, Evan Williams at Pyra Labs used "blog" as both a noun and verb ("to blog," meaning "to edit one's weblog or to post to one's weblog") and devised the term "blogger" in connection with Pyra Labs' Blogger product, leading to the popularization of the terms.

Origins

Before blogging became popular, digital communities took many forms, including Usenet, commercial online services such as GEnie, BiX and the early CompuServe, e-mail lists and Bulletin Board Systems (BBS). In the 1990s, Internet forum software, created running conversations with "threads." Threads are topical connections between messages on a virtual "corkboard."

The modern blog evolved from the online diary, where people would keep a running account of their personal lives. Most such writers called themselves diarists, journalists, or journalers. Justin Hall, who began personal blogging in 1994 while a student at Swarthmore College, is generally recognized as one of the earliest bloggers, as is Jerry Pournelle. Dave Winer's Scripting News is also credited with being one of the oldest and longest running weblogs. Another early blog was Wearable Wireless Webcam, an online shared diary of a person's personal life combining text, video, and pictures transmitted live from a wearable computer and EyeTap device to a web site in 1994. This practice of semi-automated blogging with live video together with text was referred to as sousveillance, and such journals were also used as evidence in legal matters.

Early blogs were simply manually updated components of common Web sites. However, the evolution of tools to facilitate the production and maintenance of Web articles posted in reverse chronological order made the publishing process feasible to a much larger, less technical, population. Ultimately, this resulted in the distinct class of online publishing that produces blogs we recognize today. For instance, the use of some sort of browser-based software is now a typical aspect of "blogging". Blogs can be hosted by dedicated blog hosting services, or they can be run using blog software, or on regular web hosting services.

Some early bloggers, such as The Misanthropic Bitch, who began in 1997, actually referred to their online presence as a zine, before the term blog entered common usage.

Political impact

Since 2002, blogs have gained increasing notice and coverage for their role in breaking, shaping, and spinning news stories. The Iraq war saw bloggers taking measured and passionate points of view that go beyond the traditional left-right divide of the political spectrum.

Talking Points Memo

by joshua micah marshall

(December 6th, 2002 -- 3:20 PM EST // link)

Hard-hitting coverage? We report, you decide.

I've always thought that for all the jokes about age and longevity in office, the one line that really captures how long Strom Thurmond has been around is this: he ran for president against Harry Truman.

Do you really have to say any more than that?

Support TPM Today

TPM

ALL YOU NEED ON YOUR DESERT ISLAND

Of course, Thurmond ran as the presidential candidate on the "States-Rights Democrat" or "Dixiecrat" ticket -- a candidacy that was based *exclusively and explicitly upon the preservation of legalized segregation and opposition to voting rights and civil rights for blacks.*

There's a sort of agreement in Washington these days -- with Thurmond's retirement and hundredth birthday -- to sort of forget about all that unpleasantness.

But look at what Trent Lott said about that candidacy yesterday...

> I want to say this about my state: When Strom Thurmond ran for president we voted for him. We're proud of it. And if the rest of the country had of followed our lead we wouldn't of had all these problems over all these years, either.

Oh, what could have been!!! Just another example of the hubris now reigning among Capitol Hill Republicans.

-- Josh Marshall

On 6 December 2002, Josh Marshall's talkingpointsmemo.com blog called attention to U.S. Senator Lott's comments regarding Senator Thurmond. Senator Lott was eventually to resign his Senate leadership position over the matter.

An early milestone in the rise in importance of blogs came in 2002, when many bloggers focused on comments by U.S. Senate Majority Leader Trent Lott. Senator Lott, at a party honoring U.S. Senator Strom Thurmond, praised Senator Thurmond by suggesting that the United States would have been better off had Thurmond been elected president. Lott's critics saw these comments as a tacit approval of racial segregation, a policy advocated by Thurmond's 1948 presidential campaign. This view was reinforced by documents and recorded interviews dug up by bloggers. Though Lott's comments were made at a public event attended by the media, no major media organizations reported on his controversial

comments until after blogs broke the story. Blogging helped to create a political crisis that forced Lott to step down as majority leader.

Similarly, blogs were among the driving forces behind the "Rathergate" scandal. To wit: (television journalist) Dan Rather presented documents (on the CBS show *60 Minutes*) that conflicted with accepted accounts of President Bush's military service record. Bloggers declared the documents to be forgeries and presented evidence and arguments in support of that view. Consequently, CBS apologized for what it said were inadequate reporting techniques. Many bloggers view this scandal as the advent of blogs' acceptance by the mass media, both as a news source and opinion and as means of applying political pressure.

The impact of these stories gave greater credibility to blogs as a medium of news dissemination. Though often seen as partisan gossips, bloggers sometimes lead the way in bringing key information to public light, with mainstream media having to follow their lead. More often, however, news blogs tend to react to material already published by the mainstream media. Meanwhile, an increasing number of experts blogged, making blogs a source of in-depth analysis.

Mainstream popularity

By 2004, the role of blogs became increasingly mainstream, as political consultants, news services, and candidates began using them as tools for outreach and opinion forming. Blogging was established by politicians and political candidates to express opinions on war and other issues and cemented blogs' role as a news source. Even politicians not actively campaigning, such as the UK's Labour Party's MP Tom Watson, began to blog to bond with constituents.

In January 2005, *Fortune* magazine listed eight bloggers that business people "could not ignore": Peter Rojas, Xeni Jardin, Ben Trott, Mena Trott, Jonathan Schwartz, Jason Goldman, Robert Scoble, and Jason Calacanis.

Israel's was among the first national governments to set up an official blog. Under David Saranga, the Israeli Ministry of Foreign Affairs became active in adopting Web 2.0 initiatives, including an official video blog and a political blog. The Foreign Ministry also held a microblogging press conference via Twitter about its war with Hamas, with Saranga answering questions from the public in common text-messaging abbreviations during a live worldwide press conference. The questions and answers were later posted on IsraelPolitik, the country's official political blog.

The impact of blogging upon the mainstream media has also been acknowledged by governments. In 2009, the presence of the American journalism industry had declined to the point that several newspaper corporations were filing for bankruptcy, resulting in less direct competition between newspapers within the same circulation area. Discussion emerged as to whether the newspaper industry would benefit from a stimulus package by the federal government. President Barack Obama acknowledged the emerging influence

of blogging upon society by saying "if the direction of the news is all blogosphere, all opinions, with no serious fact-checking, no serious attempts to put stories in context, that what you will end up getting is people shouting at each other across the void but not a lot of mutual understanding".

Types

There are many different types of blogs, differing not only in the type of content, but also in the way that content is delivered or written.

Personal blogs

The personal blog, an ongoing diary or commentary by an individual, is the traditional, most common blog. Personal bloggers usually take pride in their blog posts, even if their blog is never read. Blogs often become more than a way to just communicate; they become a way to reflect on life, or works of art. Blogging can have a sentimental quality. Few personal blogs rise to fame and the mainstream, but some personal blogs quickly garner an extensive following. One type of personal blog, referred to as a microblog, is extremely detailed and seeks to capture a moment in time. Some sites, such as Twitter, allow bloggers to share thoughts and feelings instantaneously with friends and family, and are much faster than emailing or writing.

Corporate and organizational blogs

A blog can be private, as in most cases, or it can be for business purposes. Blogs used internally to enhance the communication and culture in a corporation or externally for marketing, branding or public relations purposes are called corporate blogs. Similar blogs for clubs and societies are called club blogs, group blogs, or by similar names; typical use is to inform members and other interested parties of club and member activities.

By genre

Some blogs focus on a particular subject, such as political blogs, travel blogs (also known as *travelogs*), house blogs, fashion blogs, project blogs, education blogs, niche blogs, classical music blogs, quizzing blogs and legal blogs (often referred to as a blawgs) or dreamlogs. Two common types of genre blogs are art blogs and music blogs. A blog featuring discussions especially about home and family is not uncommonly called a mom blog. While not a legitimate type of blog, one used for the sole purpose of spamming is known as a Splog.

By media type

A blog comprising videos is called a vlog, one comprising links is called a linklog, a site containing a portfolio of sketches is called a sketchblog or one comprising photos is called a photoblog. Blogs with shorter posts and mixed media types are called tumblelogs. Blogs that are written on typewriters and then scanned are called typecast or typecast blogs; see typecasting (blogging).

A rare type of blog hosted on the Gopher Protocol is known as a Phlog.

By device

Blogs can also be defined by which type of device is used to compose it. A blog written by a mobile device like a mobile phone or PDA could be called a moblog.

One early blog was Wearable Wireless Webcam, an online shared diary of a person's personal life combining text, video, and pictures transmitted live from a wearable computer and EyeTap device to a web site. This practice of semi-automated blogging with live video together with text was referred to as sousveillance. Such journals have been used as evidence in legal matters.

Community and cataloging

The Blogosphere

The collective community of all blogs is known as the *blogosphere*. Since all blogs are on the internet by definition, they may be seen as interconnected and socially networked, through blogrolls, comments, linkbacks (refbacks, trackbacks or pingbacks) and backlinks. Discussions "in the blogosphere" are occasionally used by the media as a gauge of public opinion on various issues. Because new, untapped communities of bloggers can emerge in the space of a few years, Internet marketers pay close attention to "trends in the blogosphere".

BlogDay

Blogday.org was created with the belief that bloggers should have one day dedicated to getting to know other bloggers from other countries and areas of interest. The designated date is August 31, because when written 3108, it resembles the word "Blog". On that day, bloggers recommend five new blogs to their visitors, so that readers discover new, previously unknown blogs.

Blog search engines

Several blog search engines are used to search blog contents, such as Bloglines, BlogScope, and Technorati. Technorati, which is among the most popular blog search engines, provides current information on both popular searches and tags used to categorize blog postings. The research community is working on going beyond simple keyword search, by inventing new ways to navigate through huge amounts of information present in the blogosphere, as demonstrated by projects like BlogScope.

Blogging communities and directories

Several online communities exist that connect people to blogs and bloggers to other bloggers, including BlogCatalog and MyBlogLog. Interest-specific blogging platforms are also available. For instance, Blogster has a sizable community of political bloggers among its members. Global Voices aggregates international bloggers, "with emphasis on voices that are not ordinarily heard in international mainstream media."

Blogging and advertising

It is common for blogs to feature advertisements either to financially benefit the blogger or to promote the blogger's favorite causes. The popularity of blogs has also given rise to "fake blogs" in which a company will create a fictional blog as a marketing tool to promote a product.

Popularity

Researchers have analyzed the dynamics of how blogs become popular. There are essentially two measures of this: popularity through citations, as well as popularity

through affiliation (i.e. blogroll). The basic conclusion from studies of the structure of blogs is that while it takes time for a blog to become popular through blogrolls, permalinks can boost popularity more quickly, and are perhaps more indicative of popularity and authority than blogrolls, since they denote that people are actually reading the blog's content and deem it valuable or noteworthy in specific cases.

The blogdex project was launched by researchers in the MIT Media Lab to crawl the Web and gather data from thousands of blogs in order to investigate their social properties. It gathered this information for over 4 years, and autonomously tracked the most contagious information spreading in the blog community, ranking it by recency and popularity. It can therefore be considered the first instantiation of a memetracker. The project is no longer active, but a similar function is now served by tailrank.com.

Blogs are given rankings by Technorati based on the number of incoming links and Alexa Internet based on the Web hits of Alexa Toolbar users. In August 2006, Technorati found that the most linked-to blog on the internet was that of Chinese actress Xu Jinglei. Chinese media Xinhua reported that this blog received more than 50 million page views, claiming it to be the most popular blog in the world. Technorati rated Boing Boing to be the most-read group-written blog.

Blurring with the mass media

Many bloggers, particularly those engaged in participatory journalism, differentiate themselves from the mainstream media, while others are members of that media working through a different channel. Some institutions see blogging as a means of "getting around the filter" and pushing messages directly to the public. Some critics worry that bloggers respect neither copyright nor the role of the mass media in presenting society with credible news. Bloggers and other contributors to user-generated content are behind *Time* magazine naming their 2006 person of the year as "you".

Many mainstream journalists, meanwhile, write their own blogs — well over 300, according to CyberJournalist.net's J-blog list. The first known use of a blog on a news site was in August 1998, when Jonathan Dube of The Charlotte Observer published one chronicling Hurricane Bonnie.

Some bloggers have moved over to other media. The following bloggers (and others) have appeared on radio and television: Duncan Black (known widely by his pseudonym, Atrios), Glenn Reynolds (Instapundit), Markos Moulitsas Zúniga (Daily Kos), Alex Steffen (Worldchanging), Ana Marie Cox (Wonkette), Nate Silver (FiveThirtyEight.com), and Ezra Klein (Ezra Klein blog in *The American Prospect,* now in the *Washington Post*). In counterpoint, Hugh Hewitt exemplifies a mass-media personality who has moved in the other direction, adding to his reach in "old media" by being an influential blogger. Equally many established authors, for example Mitzi Szereto have started using Blogs to not only update fans on their current works but also to expand into new areas of writing.

Blogs have also had an influence on minority languages, bringing together scattered speakers and learners; this is particularly so with blogs in Gaelic languages. Minority language publishing (which may lack economic feasibility) can find its audience through inexpensive blogging.

There are many examples of bloggers who have published books based on their blogs, e.g., Salam Pax, Ellen Simonetti, Jessica Cutler, ScrappleFace. Blog-based books have been given the name blook. A prize for the best blog-based book was initiated in 2005, the Lulu Blooker Prize. However, success has been elusive offline, with many of these books not selling as well as their blogs. Only blogger Tucker Max made the New York Times Bestseller List. The book based on Julie Powell's blog "The Julie/Julia Project" was made into the film *Julie & Julia*, apparently the first to do so.

Consumer-generated advertising in blogs

Consumer-generated advertising is a relatively new and controversial development and it has created a new model of marketing communication from businesses to consumers. Among the various forms of advertising on blog, the most controversial are the sponsored posts. These are blog entries or posts and may be in the form of feedbacks, reviews, opinion, videos, etc. and usually contain a link back to the desired site using a keyword/s.

Blogs have led to some disintermediation and a breakdown of the traditional advertising model where companies can skip over the advertising agencies (previously the only interface with the customer) and contact the customers directly themselves. On the other hand, new companies specialised in blog advertising have been established, to take advantage of this new development as well.

However, there are many people who look negatively on this new development. Some believe that any form of commercial activity on blogs will destroy the blogosphere's credibility.

Legal and social consequences

Blogging can result in a range of legal liabilities and other unforeseen consequences.

Defamation or liability

Several cases have been brought before the national courts against bloggers concerning issues of defamation or liability. U.S. payouts related to blogging totaled $17.4 million by 2009; in some cases these have been covered by umbrella insurance. The courts have returned with mixed verdicts. Internet Service Providers (ISPs), in general, are immune from liability for information that originates with third parties (U.S. Communications Decency Act and the EU Directive 2000/31/EC).

In *Doe v. Cahill*, the Delaware Supreme Court held that stringent standards had to be met to unmask the anonymous posts of bloggers and also took the unusual step of dismissing

the libel case itself (as unfounded under American libel law) rather than referring it back to the trial court for reconsideration. In a bizarre twist, the Cahills were able to obtain the identity of John Doe, who turned out to be the person they suspected: the town's mayor, Councilman Cahill's political rival. The Cahills amended their original complaint, and the mayor settled the case rather than going to trial.

In January 2007, two prominent Malaysian political bloggers, Jeff Ooi and Ahiruddin Attan, were sued by pro-government newspaper, The New Straits Times Press (Malaysia) Berhad, Kalimullah bin Masheerul Hassan, Hishamuddin bin Aun and Brenden John a/l John Pereira over an alleged defamation. The plaintiff was supported by the Malaysian government. Following the suit, the Malaysian government proposed to "register" all bloggers in Malaysia in order to better control parties against their interest. This is the first such legal case against bloggers in the country.

In the United States, blogger Aaron Wall was sued by Traffic Power for defamation and publication of trade secrets in 2005. According to Wired Magazine, Traffic Power had been "banned from Google for allegedly rigging search engine results." Wall and other "white hat" search engine optimization consultants had exposed Traffic Power in what they claim was an effort to protect the public. The case was watched by many bloggers because it addressed the murky legal question of who is liable for comments posted on blogs. The case was dismissed for lack of personal jurisdiction, and Traffic Power failed to appeal within the allowed time.

In 2009, a controversial and landmark decision by The Hon. Mr Justice Eady refused to grant an order to protect the anonymity of Richard Horton.

In 2009, NDTV issued a legal notice to Indian blogger Chetan Kunte for "abusive free speech" regarding a blog post criticizing their coverage of the Mumbai attacks. The blogger unconditionally withdrew his post, replacing it with legal undertaking and an admission that his post had been "defamatory and untrue" which resulted in several Indian bloggers criticizing NDTV for trying to silence critics.

Employment

Employees who blog about elements of their place of employment can begin to affect the brand recognition of their employer. In general, attempts by employee bloggers to protect themselves by maintaining anonymity have proved ineffective.

Delta Air Lines fired flight attendant Ellen Simonetti because she posted photographs of herself in uniform on an airplane and because of comments posted on her blog "Queen of Sky: Diary of a Flight Attendant" which the employer deemed inappropriate. This case highlighted the issue of personal blogging and freedom of expression versus employer rights and responsibilities, and so it received wide media attention. Simonetti took legal action against the airline for "wrongful termination, defamation of character and lost future wages". The suit was postponed while Delta was in bankruptcy proceedings (court docket).

In early 2006, Erik Ringmar, a tenured senior lecturer at the London School of Economics, was ordered by the convenor of his department to "take down and destroy" his blog in which he discussed the quality of education at the school.

Mark Cuban, owner of the Dallas Mavericks, was fined during the 2006 NBA playoffs for criticizing NBA officials on the court and in his blog.

Mark Jen was terminated in 2005 after 10 days of employment as an Assistant Product Manager at Google for discussing corporate secrets on his personal blog, then called 99zeros and hosted on the Google-owned Blogger service. He blogged about unreleased products and company finances a week before the company's earnings announcement. He was fired two days after he complied with his employer's request to remove the sensitive material from his blog.

In India, blogger Gaurav Sabnis resigned from IBM after his posts exposing the false claims of a management school, IIPM, led to management of IIPM threatening to burn their IBM laptops as a sign of protest against him.

Jessica Cutler, aka "The Washingtonienne", blogged about her sex life while employed as a congressional assistant. After the blog was discovered and she was fired, she wrote a novel based on her experiences and blog: *The Washingtonienne: A Novel*. Cutler is presently being sued by one of her former lovers in a case that could establish the extent to which bloggers are obligated to protect the privacy of their real life associates.

Catherine Sanderson, a.k.a. Petite Anglaise, lost her job in Paris at a British accountancy firm because of blogging. Although given in the blog in a fairly anonymous manner, some of the descriptions of the firm and some of its people were less than flattering. Sanderson later won a compensation claim case against the British firm, however.

On the other hand, Penelope Trunk wrote an upbeat article in the *Boston Globe* back in 2006, entitled "Blogs 'essential' to a good career". She was one of the first journalists to point out that a large portion of bloggers are professionals and that a well-written blog can help attract employers.

Political dangers

Blogging can sometimes have unforeseen consequences in politically sensitive areas. Blogs are much harder to control than broadcast or even print media. As a result, totalitarian and authoritarian regimes often seek to suppress blogs and/or to punish those who maintain them.

In Singapore, two ethnic Chinese were imprisoned under the country’s anti-sedition law for posting anti-Muslim remarks in their blogs.

Egyptian blogger Kareem Amer was charged with insulting the Egyptian president Hosni Mubarak and an Islamic institution through his blog. It is the first time in the history of

Egypt that a blogger was prosecuted. After a brief trial session that took place in Alexandria, the blogger was found guilty and sentenced to prison terms of three years for insulting Islam and inciting sedition, and one year for insulting Mubarak.

Egyptian blogger Abdel Monem Mahmoud was arrested in April 2007 for anti-government writings in his blog. Monem is a member of the banned Muslim Brotherhood.

After expressing opinions in his personal blog about the state of the Sudanese armed forces, Jan Pronk, United Nations Special Representative for the Sudan, was given three days notice to leave Sudan. The Sudanese army had demanded his deportation.

In Myanmar, Nay Phone Latt, a blogger, was sentenced to 20 years in jail for posting a cartoon critical of head of state Than Shwe.

Personal safety

One consequence of blogging is the possibility of attacks or threats against the blogger, sometimes without apparent reason. Kathy Sierra, author of the innocuous blog "Creating Passionate Users", was the target of such vicious threats and misogynistic insults that she canceled her keynote speech at a technology conference in San Diego, fearing for her safety. While a blogger's anonymity is often tenuous, Internet trolls who would attack a blogger with threats or insults can be emboldened by anonymity. Sierra and supporters initiated an online discussion aimed at countering abusive online behavior and developed a blogger's code of conduct.

Behavior

The **Blogger's Code of Conduct** is a proposal by Tim O'Reilly for **bloggers** to enforce civility on their blogs by being civil themselves and moderating comments on their blog. The code was proposed due to threats made to blogger Kathy Sierra. The idea of the code was first reported by BBC News, who quoted O'Reilly saying, "I do think we need some code of conduct around what is acceptable behaviour, I would hope that it doesn't come through any kind of regulation it would come through self-regulation."

O'Reilly and others came up with a list of seven proposed ideas:

1. Take responsibility not just for your own words, but for the comments you allow on your blog.
2. Label your tolerance level for abusive comments.
3. Consider eliminating anonymous comments.
4. Ignore the trolls.
5. Take the conversation offline, and talk directly, or find an intermediary who can do so.
6. If you know someone who is behaving badly, tell them so.
7. Don't say anything online that you wouldn't say in person.

Chapter 9

File Sharing

File sharing is the practice of distributing or providing access to digitally stored information, such as computer programs, multimedia (audio, images, & video), documents, or electronic books. It may be implemented through a variety of ways. Storage, transmission, and distribution models are common methods of file sharing that incorporate manual sharing using removable media, centralized computer file server installations on computer networks, World Wide Web-based hyperlinked documents, and the use of distributed peer-to-peer networking.

Types of file sharing

Peer-to-peer file sharing

Users can use software that connects in to a peer-to-peer network to search for shared files on the computers of other users (i.e. peers) connected to the network. Files of interest can then be downloaded directly from other users on the network. Typically, large files are broken down into smaller chunks, which may be obtained from multiple peers and then reassembled by the downloader. This is done while the peer is simultaneously uploading the chunks it already has to other peers.

File hosting services

File hosting services are a simple alternative to peer-to-peer software. These are sometimes used together with Internet collaboration tools such as email, forums, blogs, or any other medium in which links to direct downloads from file hosting services can be embedded. These sites typically host files so that others can download them.

History

Files were first exchanged on removable media. Computers were able to access remote files using filesystem mounting, bulletin board systems (1978), Usenet (1979), and FTP servers (1985). Internet Relay Chat (1988) and Hotline (1997) enabled users to communicate remotely through chat and to exchange files. The mp3 encoding, which was standardized in 1991 and which substantially reduced the size of audio files, grew to widespread use in the late 1990s. In 1998, MP3.com and Audiogalaxy were established, the Digital Millennium Copyright Act was unanimously passed, and the first mp3 player devices were launched. MP3.com offered music by unsigned artists, and grew to serve 4 million audio downloads daily.

Usenet was created in 1979. It is a network that was initially based on the UUCP protocol for dial-up connections and has, since being transported over the Internet, used a specialized client-server protocol, the Network News Transfer Protocol (NNTP). Its main purpose was the exchange of text based messages, but through attachments allowed users to encode files and distribute them to participating subscribers of Usenet newsgroups. Usenet remains one of the largest carriers of file sharing and Internet traffic. Legal challenges to P2P systems have spurred a resurgence of Usenet. Usenet itself has also been the target of legal challenges pertaining to its use in file sharing.

Between 1979 and the mid 1990s, file sharing was done through bulletin board systems and Usenet. The term shareware and its distribution model became more popular in part due to the BBS networks and systems. Putting shareware on BBS was a way for some developers to distribute their software and generate income. Games such as Doom became popular as a result of this distribution model. Bulletin boards eventually became obsolete as the Internet grew in popularity.

In June 1999, Napster was released as a centralized unstructured peer-to-peer system, requiring a central server for indexing and peer discovery. It is generally credited as being the first peer-to-peer file sharing system. In the case of Napster, an online service provider could not use the "transitory network transmission" safe harbor in the DMCA if they had control of the network with a server. Many P2P products will, by their very nature, flunk this requirement, just as Napster did. Napster provided a service where they indexed and stored file information that users of Napster made available on their computers for others to download, and the files were transferred directly between the host and client users after authorization by Napster. Shortly after the A&M Records, Inc. v. Napster, Inc. loss in court Napster blocked all copyright content from being downloaded.

Gnutella, eDonkey2000, and Freenet were released in 2000, as MP3.com and Napster were facing litigation. Gnutella, released in March, was the first decentralized file sharing network. In the gnutella network, all connecting software was considered equal, and therefore the network had no central point of failure. In July, Freenet was released and became the first anonymity network. In September the eDonkey2000 client and server software was released.

In 2001, Kazaa and Poisoned for the Mac was released. Its FastTrack network was distributed, though unlike gnutella, it assigned more traffic to 'supernodes' to increase routing efficiency. The network was proprietary and encrypted, and the Kazaa team made substantial efforts to keep other clients such as Morpheus off of the FastTrack network.

In July 2001, Napster was sued by several recording companies. As a result, Napster lost in court against these companies and was shut down. This drove users to other P2P applications and file sharing continued its exponential growth. The Audiogalaxy Satellite client grew in popularity, and the LimeWire client and BitTorrent protocol were released. Until its decline in 2004, Kazaa was the most popular file sharing program despite bundled malware and legal battles in the Netherlands, Australia, and the United States. In 2002, a Tokyo district court ruling shut down File Rogue and an RIAA lawsuit effectively shut down Audiogalaxy.

Demonstrators protesting The Pirate Bay raid, 2006.

From 2002 through 2003, a number of BitTorrent services were established, including Suprnova.org, isoHunt, TorrentSpy, and The Pirate Bay. In 2002, the RIAA was filing lawsuits against Kazaa users. As a result of such lawsuits, many universities added file sharing regulations in their school administrative codes (though some students managed to circumvent them during after school hours). With the shut down of eDonkey in 2005, eMule became the dominant client of the eDonkey network. In 2006, police raids took down the Razorback2 eDonkey server and temporarily took down The Pirate Bay. Pro-piracy demonstrations took place in Sweden in response to the Pirate Bay raid. In 2009,

the Pirate Bay trial ended in a guilty verdict for the primary founders of the tracker. The decision was appealed, leading to a second guilty verdict in November 2010

Networks such as BitTorrent via uTorrent and Azureus and the trackers & indexing sites, gnutella via Limewire and the eDonkey network via eMule managed to survive this turbulent time. Limewire was forced to shut down following a court order in Arista Records LLC v. Lime Group LLC in October 2010, but the gnutella network remains active through open source clients like Frostwire and gtk-gnutella. Furthermore, multi-protocol file sharing software such as MLDonkey and Shareaza adapted in order to support all the major file sharing protocols, so users no longer had to install and configure multiple file sharing programs.

File sharing in academia

File sharing occurs in academic and research circles, where researchers wish to access subscription journals and books, but do not wish to pay a licence fee. File-sharing websites allow researchers to request articles, which are then found by those who do have access to them, and then the articles are posted to the website for all to access, a practice that appears to be unknown to many editors of these journals. The file sharing is extended even further by researchers who share library access codes (usernames and passwords) so that other researchers can access the library databases directly themselves.

WT

Chapter 10

Digital Library

A **digital library** is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers. The digital content may be stored locally, or accessed remotely via computer networks. A digital library is a type of information retrieval system.

The *DELOS Digital Library Reference Model* defines a digital library as:

An organization, which might be virtual, that comprehensively collects, manages and preserves for the long term rich digital content, and offers to its user communities specialized functionality on that content, of measurable quality and according to codified policies.

The first use of the term *digital library* in print may have been in a 1988 report to the Corporation for National Research Initiatives The term *digital libraries* was first popularized by the NSF/DARPA/NASA Digital Libraries Initiative in 1994. These draw heavily on As We May Think by Vannevar Bush in 1945, which set out a vision not in terms of technology, but user experience. The term *virtual library* was initially used interchangeably with *digital library,* but is now primarily used for libraries that are virtual in other senses (such as libraries which aggregate distributed content).

A distinction is often made between content that was created in a digital format, known as born-digital, and information that has been converted from a physical medium, e.g., paper, by digitizing. The term hybrid library is sometimes used for libraries that have both physical collections and digital collections. For example, American Memory is a digital library within the Library of Congress. Some important digital libraries also serve as long term archives, for example, the Eprint arXiv, and the Internet Archive.

Academic repositories

Many academic libraries are actively involved in building institutional repositories of the institution's books, papers, theses, and other works which can be digitized or were 'born digital'. Many of these repositories are made available to the general public with few restrictions, in accordance with the goals of open access, in contrast to the publication of research in commercial journals, where the publishers often limit access rights. Institutional, truly free, and corporate repositories are sometimes referred to as digital libraries.

Digital archives

Physical archives differ from physical libraries in several ways. Traditionally, archives were defined as:

1. Containing primary sources of information (typically letters and papers directly produced by an individual or organization) rather than the secondary sources found in a library (books, periodicals, etc);
2. Having their contents organized in groups rather than individual items.
3. Having unique contents.

The technology used to create digital libraries has been even more revolutionary for archives since it breaks down the second and third of these general rules. In other words, "digital archives" or "online archives" will still generally contain primary sources, but they are likely to be described individually rather than (or in addition to) in groups or collections, and because they are digital their contents are easily reproducible and may indeed have been reproduced from elsewhere. The Oxford Text Archive is generally considered to be the oldest digital archive of academic physical primary source materials.

The future

Large scale digitization projects are underway at Google, the Million Book Project, and Internet Archive. With continued improvements in book handling and presentation technologies such as optical character recognition and ebooks, and development of alternative depositories and business models, digital libraries are rapidly growing in popularity as demonstrated by Google, Yahoo!, and MSN's efforts. Just as libraries have ventured into audio and video collections, so have digital libraries such as the Internet Archive.

According to Larry Lannom, Director of Information Management Technology at the nonprofit Corporation for National Research Initiatives, “all the problems associated with digital libraries are wrapped up in archiving.” He goes on to state, “If in 100 years people can still read your article, we’ll have solved the problem.” Daniel Akst, author of The Webster Chronicle, proposes that “the future of libraries—and of information—is digital.” Peter Lyman and Hal Varian, information scientists at the University of California, Berkeley, estimate that “the world’s total yearly production of print, film,

optical, and magnetic content would require roughly 1.5 billion gigabytes of storage." Therefore, they believe that "soon it will be technologically possible for an average person to access virtually all recorded information."

Searching

Most digital libraries provide a search interface which allows resources to be found. These resources are typically deep web (or invisible web) resources since they frequently cannot be located by search engine crawlers. Some digital libraries create special pages or sitemaps to allow search engines to find all their resources. Digital libraries frequently use the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to expose their metadata to other digital libraries, and search engines like Google Scholar, Yahoo! and Scirus can also use OAI-PMH to find these deep web resources.

There are two general strategies for searching a **federation** of digital libraries:

1. distributed searching, and
2. searching previously harvested metadata.

Distributed searching typically involves a client sending multiple search requests in parallel to a number of servers in the federation. The results are gathered, duplicates are eliminated or clustered, and the remaining items are sorted and presented back to the client. Protocols like Z39.50 are frequently used in distributed searching. A benefit to this approach is that the resource-intensive tasks of indexing and storage are left to the respective servers in the federation. A drawback to this approach is that the search mechanism is limited by the different indexing and ranking capabilities of each database, making it difficult to assemble a combined result consisting of the most relevant found items.

Searching over previously harvested metadata involves searching a locally stored index of information that has previously been collected from the libraries in the federation. When a search is performed, the search mechanism does not need to make connections with the digital libraries it is searching - it already has a local representation of the information. This approach requires the creation of an indexing and harvesting mechanism which operates regularly, connecting to all the digital libraries and querying the whole collection in order to discover new and updated resources. OAI-PMH is frequently used by digital libraries for allowing metadata to be harvested. A benefit to this approach is that the search mechanism has full control over indexing and ranking algorithms, possibly allowing more consistent results. A drawback is that harvesting and indexing systems are more resource-intensive and therefore expensive.

Frameworks

The formal reference models include the DELOS Digital Library Reference Model (Agosti, et al., 2006) and the Streams, Structures, Spaces, Scenarios, Societies (5S)

formal framework The Reference Model for an Open Archival Information System (OAIS) provides a framework to address digital preservation.

Construction and organization

Software

There are a number of software packages for use in general digital libraries, for notable ones see Digital library software. Institutional repository software, which focuses primarily on ingest, preservation and access of locally produced documents, particularly locally produced academic outputs, can be found in Institutional repository software.

Digitization

In the past few years, procedures for digitizing books at high speed and comparatively low cost have improved considerably with the result that it is now possible to plan the digitization of millions of books per year for creating digital libraries.

Advantages

The advantages of digital libraries as a means of easily and rapidly accessing books, archives and images of various types are now widely recognized by commercial interests and public bodies alike.

Traditional libraries are limited by storage space; digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain it. As such, the cost of maintaining a digital library is much lower than that of a traditional library.

A traditional library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Digital libraries may reduce or, in some instances, do away with these fees. Both types of library require cataloguing input to allow users to locate and retrieve material. Digital libraries may be more willing to adopt innovations in technology providing users with improvements in electronic and audio book technology as well as presenting new forms of communication such as blogs; conventional libraries may consider that providing online access to their OPAC catalogue is sufficient. An important advantage to digital conversion is increased accessibility to users. They also increase availability to individuals who may not be traditional patrons of a library, due to geographic location or organizational affiliation.

- **No physical boundary**. The user of a digital library need not to go to the library physically; people from all over the world can gain access to the same information, as long as an Internet connection is available.
- **Round the clock availability** A major advantage of digital libraries is that people can gain access 24/7 to the information.

- **Multiple access**. The same resources can be used simultaneously by a number of institutions and patrons. This may not be the case for copyrighted material: a library may have a license for "lending out" only one copy at a time; this is achieved with a system of digital rights management where a resource can become inaccessible after expiration of the lending period or after the lender chooses to make it inaccessible (equivalent to returning the resource).
- **Information retrieval**. The user is able to use any search term (word, phrase, title, name, subject) to search the entire collection. Digital libraries can provide very user-friendly interfaces, giving clickable access to its resources.
- **Preservation and conservation**. Digitization is not a long-term preservation solution for physical collections, but does succeed in providing access copies for materials that would otherwise fall to degradation from repeated use. Digitized collections and born-digital objects pose many preservation and conservation concerns that analog materials do not. Please see the following "Problems" section of this page for examples.
- **Space**. Whereas traditional libraries are limited by storage space, digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain them and media storage technologies are more affordable than ever before.
- **Added value**. Certain characteristics of objects, primarily the quality of images, may be improved. Digitization can enhance legibility and remove visible flaws such as stains and discoloration.
- **Easily accessible**.

Challenges

Digital preservation

Digital preservation aims to ensure that digital media and information systems are still interpretable into the indefinite future. Each necessary component of the must be migrated, preserved or emulated. Typically lower levels of systems (floppy disks for example) are emulated, bit-streams (the actual files stored in the disks) are preserved and operating systems are emulated as a virtual machine. Only where the meaning and content of digital media and information systems are well understood is migration possible, as is the case for office documents.

Copyright and licensing

Some people have criticized that digital libraries are hampered by copyright law, because works cannot be shared over different periods of time in the manner of a traditional library. The republication of material on the Web by libraries may require permission from rights holders, and there is a conflict of interest between them and publishers who may wish to create online versions of their acquired content for commercial purposes.

There is a dilution of responsibility that occurs as a result of the spread-out nature of digital resources. Complex intellectual property matters may become involved since

digital material is not always owned by a library. The content is, in many cases, public domain or self-generated content only. Some digital libraries, such as Project Gutenberg, work to digitize out-of-copyright works and make them freely available to the public. An estimate of the number of distinct books still existent in library catalogues from 2000BC to 1960, has been made.

The Fair Use Provisions (17 USC § 107) under copyright law provide specific guidelines under which circumstances libraries are allowed to copy digital resources. Four factors that constitute fair use are purpose of use, nature of the work, market impact, and amount or substantiality used.

Some digital libraries acquire a license to "lend out" their resources. This may involve the restriction of lending out only one copy at a time for each license, and applying a system of digital rights management for this purpose.

Metadata creation

In traditional libraries, the ability to find works of interest was directly related to how well they were catalogued. While cataloguing electronic works digitized from a library's existing holding may be as simple as copying moving a record for the print to the electronic item, with complex and born-digital works requiring substantially more effort. To handle the growing volume of electronic publications, new tools and technologies have to be designed to allow effective automated semantic classification and searching. While full text search can be used for some searches, there are many common catalog searches which cannot be performed using full text, including:

- finding texts which are translations of other texts
- linking texts published under pseudonyms to the real authors (Samuel Clemens and Mark Twain, for example)
- differentiating non-fiction from parody (The Onion from The New York Times, for example)

Chapter 11

Digital Preservation

Digital preservation is the active management of digital information over time to ensure its accessibility. Preservation of digital information is widely considered to require more constant and ongoing attention than preservation of other media. This constant input of effort, time, and money to handle rapid technological and organizational advance is considered a major stumbling block for preserving digital information. Indeed, while we are still able to read our written heritage from several thousand years ago, the digital information created merely a decade ago is in serious danger of being lost, creating a digital Dark Age.

Digital preservation is the set of processes and activities that ensure continued access to information and all kinds of records, scientific and cultural heritage existing in digital formats. This includes the preservation of materials resulting from digital reformatting, but particularly information that is born-digital and has no analog counterpart. In the language of digital imaging and electronic resources, preservation is no longer just the product of a program but an ongoing process. In this regard the way digital information is stored is important in ensuring its longevity. The long-term storage of digital information is assisted by the inclusion of preservation metadata.

Digital preservation is defined as: long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required for. Long-term is defined as "long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely". "Retrieval" means obtaining needed digital files from the long-term, error-free digital storage, without possibility of corrupting the continued error-free storage of the digital files. "Interpretation" means that the retrieved digital files, files that, for example, are of texts, charts, images or sounds, are decoded and transformed into usable representations. This is often interpreted as "rendering", i.e. making it available for a human to access. However, in many cases it will mean able to be processed by computational means.

Why active preservation is necessary

Society's heritage has been presented on many different materials, including stone, vellum, bamboo, silk, and paper. Now a large quantity of information exists in digital forms, including emails, blogs, social networking websites, national elections websites, web photo albums, and sites which change their content over time. According an article by Brewster Kahle, in 1996 founder of Internet Archive, "Preserving the Internet", Scientific American, the average life of a URL was, in 1997, 44 days .

The unique characteristic of digital forms makes it easy to create content and keep it up-to-date, but at the same time brings many difficulties in the preservation of this content. Margaret Hedstrom points out that "...digital preservation raises challenges of a fundamentally different nature which are added to the problems of preserving traditional format materials."

Physical deterioration

The media on which digital contents are stored are more vulnerable to deterioration and catastrophic loss than some analog media such as paper. While acid paper is prone to deterioration, becoming brittle and yellowing with age, the deterioration may not become apparent for some decades and progresses slowly. It remains possible to retrieve information without loss once deterioration is noticed. Digital data recording media may deteriorate more rapidly and once the deterioration starts, in most cases there may already be data loss. This characteristic of digital forms leaves a very short time frame for preservation decisions and actions.

Digital obsolescence

Another challenge is the issue of long-term access to data. Digital technology is developing quickly and retrieval and playback technologies can become obsolete in a matter of years. When faster, more capable and less expensive storage and processing devices are developed, older versions may be quickly replaced. When a software or decoding technology is abandoned, or a hardware device is no longer in production, records created with such technologies are at great risk of loss, simply because they are no longer accessible. This process is known as digital obsolescence.

This challenge is exacerbated by a lack of established standards, protocols and proven methods for preserving digital information. We used to save copies of data on tapes, but media standards for tapes have changed considerably over the last five to ten years, and there is no guarantee that tapes will be readable in the future. Recovering these materials may require special tools Hedstrom further explained that almost all digital library researches have been focused on "...architectures and systems for information organization and retrieval, presentation and visualization, and administration of intellectual property rights" and that "...digital preservation remains largely experimental and replete with the risks associated with untested methods".

Strategies

In 2006, the Online Computer Library Center developed a four-point strategy for the long-term preservation of digital objects that consisted of:

- Assessing the risks for loss of content posed by technology variables such as commonly used proprietary file formats and software applications.
- Evaluating the digital content objects to determine what type and degree of format conversion or other preservation actions should be applied.
- Determining the appropriate metadata needed for each object type and how it is associated with the objects.
- Providing access to the content.

There are several additional strategies that individuals and organizations may use to actively combat the loss of digital information.

Refreshing

Refreshing is the transfer of data between two types of the same storage medium so there are no bitrate changes or alteration of data. For example, transferring census data from an old preservation CD to a new one. This strategy may need to be combined with migration when the software or hardware required to read the data is no longer available or is unable to understand the format of the data. Refreshing will likely always be necessary due to the deterioration of physical media.

Migration

Migration is the transferring of data to newer system environments (Garrett et al., 1996). This may include conversion of resources from one file format to another (e.g., conversion of Microsoft Word to PDF or OpenDocument), from one operating system to another (e.g., Windows to Linux) or from one programming language to another (e.g., C to Java) so the resource remains fully accessible and functional. Resources that are migrated run the risk of losing some type of functionality since newer formats may be incapable of capturing all the functionality of the original format, or the converter itself may be unable to interpret all the nuances of the original format. The latter is often a concern with proprietary data formats.

The US National Archives Electronic Records Archives and Lockheed Martin are jointly developing a migration system that will preserve any type of document, created on any application or platform, and delivered to the archives on any type of digital media. In the system, files are translated into flexible formats, such as XML; they will therefore be accessible by technologies in the future. Lockheed Martin argues that it would be impossible to develop an emulation system for the National Archives ERA because the volume of records and cost would be prohibitive.

Replication

Creating duplicate copies of data on one or more systems is called *replication*. Data that exists as a single copy in only one location is highly vulnerable to software or hardware failure, intentional or accidental alteration, and environmental catastrophes like fire, flooding, etc. Digital data is more likely to survive if it is replicated in several locations. Replicated data may introduce difficulties in refreshing, migration, versioning, and access control since the data is located in multiple places.

Emulation

Emulation is the replicating of functionality of an obsolete system. Examples include emulating an Atari 2600 on a Windows system or emulating WordPerfect 1.0 on a Macintosh. Emulators may be built for applications, operating systems, or hardware platforms. Emulation has been a popular strategy for retaining the functionality of old video game systems, such as with the MAME project. The feasibility of emulation as a catch-all solution has been debated in the academic community. (Granger, 2000)

Raymond A. Lorie has suggested a Universal Virtual Computer (UVC) could be used to run any software in the future on a yet unknown platform. The UVC strategy uses a combination of emulation and migration. The UVC strategy has not yet been widely adopted by the digital preservation community.

Jeff Rothenberg, a major proponent of Emulation for digital preservation in libraries, working in partnership with Koninklijke Bibliotheek and National Archief of the Netherlands, has recently helped launch Dioscuri, a modular emulator that succeeds in running MS-DOS, WordPerfect 5.1, DOS games, and more.

Metadata attachment

Metadata is data on a digital file that includes information on creation, access rights, restrictions, preservation history, and rights management. Metadata attached to digital files may be affected by file format obsolescence. ASCII is considered to be the most durable format for metadata because it is widespread, backwards compatible when used with Unicode, and utilizes human-readable characters, not numeric codes. It retains information, but not the structure information it is presented in. For higher functionality, SGML or XML should be used. Both markup languages are stored in ASCII format, but contain tags that denote structure and format.

Trustworthy digital objects

Digital objects that can speak to their own authenticity are called *trustworthy digital objects* (TDOs). TDOs were proposed by Henry M. Gladney to enable digital objects to maintain a record of their change history so future users can know with certainty that the contents of the object are authentic. Other preservation strategies like replication and migration are necessary for the long-term preservation of TDOs.

Digital sustainability

Digital sustainability encompasses a range of issues and concerns that contribute to the longevity of digital information. Unlike traditional, temporary strategies and more permanent solutions, digital sustainability implies a more active and continuous process. Digital sustainability concentrates less on the solution and technology and more on building an infrastructure and approach that is flexible with an emphasis on interoperability, continued maintenance and continuous development. Digital sustainability incorporates activities in the present that will facilitate access and availability in the future.

Digital preservation standards

To standardize digital preservation practice and provide a set of recommendations for preservation program implementation, the Reference Model for an Open Archival Information System (OAIS) was developed. The reference model (ISO 14721:2003) includes the following responsibilities that an OAIS archive must abide by:

- Negotiate for and accept appropriate information from information Producers.
- Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.
- Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided.
- Ensure that the information to be preserved is Independently Understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.
- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.
- Make the preserved information available to the Designated Community.

OAIS is concerned with all technical aspects of a digital object's life cycle: ingest into and storage in a preservation infrastructure, data management, accessibility, and distribution. The model also addresses metadata issues and recommends that five types of metadata be attached to a digital object: reference (identification) information, provenance (including preservation history), context, fixity (authenticity indicators), and representation (formatting, file structure, and what "imparts meaning to an object's bitstream". Prior to Gladney's proposal of TDOs was the Research Library Group's (RLG) development of "attributes and responsibilities" that denote the practices of a "Trusted Digital Repository" (TDR) The seven attributes of a TDR are: "compliance with the Reference Model for an Open Archival Information System (OAIS), Administrative responsibility, Organizational viability, Financial sustainability, Technological and

procedural suitability, System security, Procedural accountability." Among RLG's attributes and responsibilities were recommendations calling for the collaborative development of digital repository certifications, models for cooperative networks, and sharing of research and information on digital preservation with regards to intellectual property rights.

Digital sound preservation standards

In January 2004, the Council on Library and Information Resources (CLIR) hosted a roundtable meeting of audio experts discussing best practices, which culminated in a report delivered March 2006. This report investigated procedures for reformatting sound from analog to digital, summarizing discussions and recommendations for best practices for digital preservation. Participants made a series of 15 recommendations for improving the practice of analog audio transfer for archiving:

- Develop core competencies in audio preservation engineering. Participants noted with concern that the number of experts qualified to transfer older recordings is shrinking and emphasized the need to find a way to ensure that the technical knowledge of these experts can be passed on.
- Develop arrangements among smaller institutions that allow for cooperative buying of esoteric materials and supplies.
- Pursue a research agenda for magnetic-tape problems that focuses on a less destructive solution for hydrolysis than baking, relubrication of acetate tapes, and curing of cupping.
- Develop guidelines for the use of automated transfer of analog audio to digital preservation copies.
- Develop a web-based clearinghouse for sharing information on how archives can develop digital preservation transfer programs.
- Carry out further research into nondestructive playback of broken audio discs.
- Develop a flowchart for identifying the composition of various types of audio discs and tapes.
- Develop a reference chart of problematic media issues.
- Collate relevant audio engineering standards from organizations.
- Research safe and effective methods for cleaning analog tapes and discs.

- Develop a list of music experts who could be consulted for advice on transfer of specific types of musical content (e.g., determining the proper key so that correct playback speed can be established).

- Research the life expectancy of various audio formats.

- Establish regional digital audio repositories.

- Cooperate to develop a common vocabulary within the field of audio preservation.

- Investigate the transfer of technology from such fields as chemistry and materials science to various problems in audio preservation.

Updated technical guidelines on the creation and preservation of digital audio have been prepared by the International Association of Sound and Audiovisual Archives (IASA).

Examples of digital preservation initiatives

- **Xena** is a free Java-based open source archiving solution that can be installed on any desktop PC. It converts proprietary document, graphics and audio file formats to open formats, and normalizes other binary files to ASCII with an XML file wrapper.

- **ArchivalWare** built by PTFS, Inc. is a digital library solution created specifically to house, disseminate, preserve and allow discovery of digital assets. The product supports archival versions and dissemination versions of ingested digital objects, creates PDFa files upon ingestion for long term digital preservation and includes XMP metadata support which allows rich metadata to live in and move with the digital object itself.

- **DSpace** is open source software that is available to anyone who has the World Wide Web. DSpace takes data in multiple formats (text, video, audio, or data), distributes it over the web, indexes the data (for easy retrieval), and preserves the data over time.

- The British Library is responsible for several programmes in the area of **digital preservation**. The National Archives of the United Kingdom have also pioneered various initiatives in the field of **digital preservation**.

- **PADI** is a comprehensive archive of information on the topic of digital preservation from the National Library of Australia.

- **SimpleDL** can store multiple formats, including text, images, video, audio, and data. SimpleDL uses Amazon S3 to provide 99.999999999% durability for the files stored in its preservation system.

Large-scale digital preservation initiatives (LSDIs)

Many research libraries and archives have begun or are about to begin Large-Scale digital preservation initiatives (LSDI's). The main players in LSDIs are cultural institutions, commercial companies such as Google and Microsoft, and non-profit groups including the Open Content Alliance (OCA), the Million Book Project (MBP), and HathiTrust. The primary motivation of these groups is to expand access to scholarly resources.

LSDIs: library perspective

Approximately 30 cultural entities, including the 12-member Committee on Institutional Cooperation (CIC), have signed digitization agreements with either Google or Microsoft. Several of these cultural entities are participating in the Open Content Alliance (OCA) and the Million Book Project (MBP). Some libraries are involved in only one initiative and others have diversified their digitization strategies through participation in multiple initiatives. The three main reasons for library participation in LSDIs are: Access, Preservation and Research and Development. It is hoped that digital preservation will ensure that library materials remain accessible for future generations. Libraries have a perpetual responsibility for their materials and a commitment to archive their digital materials. Libraries plan to use digitized copies as backups for works in case they go out of print, deteriorate, or are lost and damaged.

WT