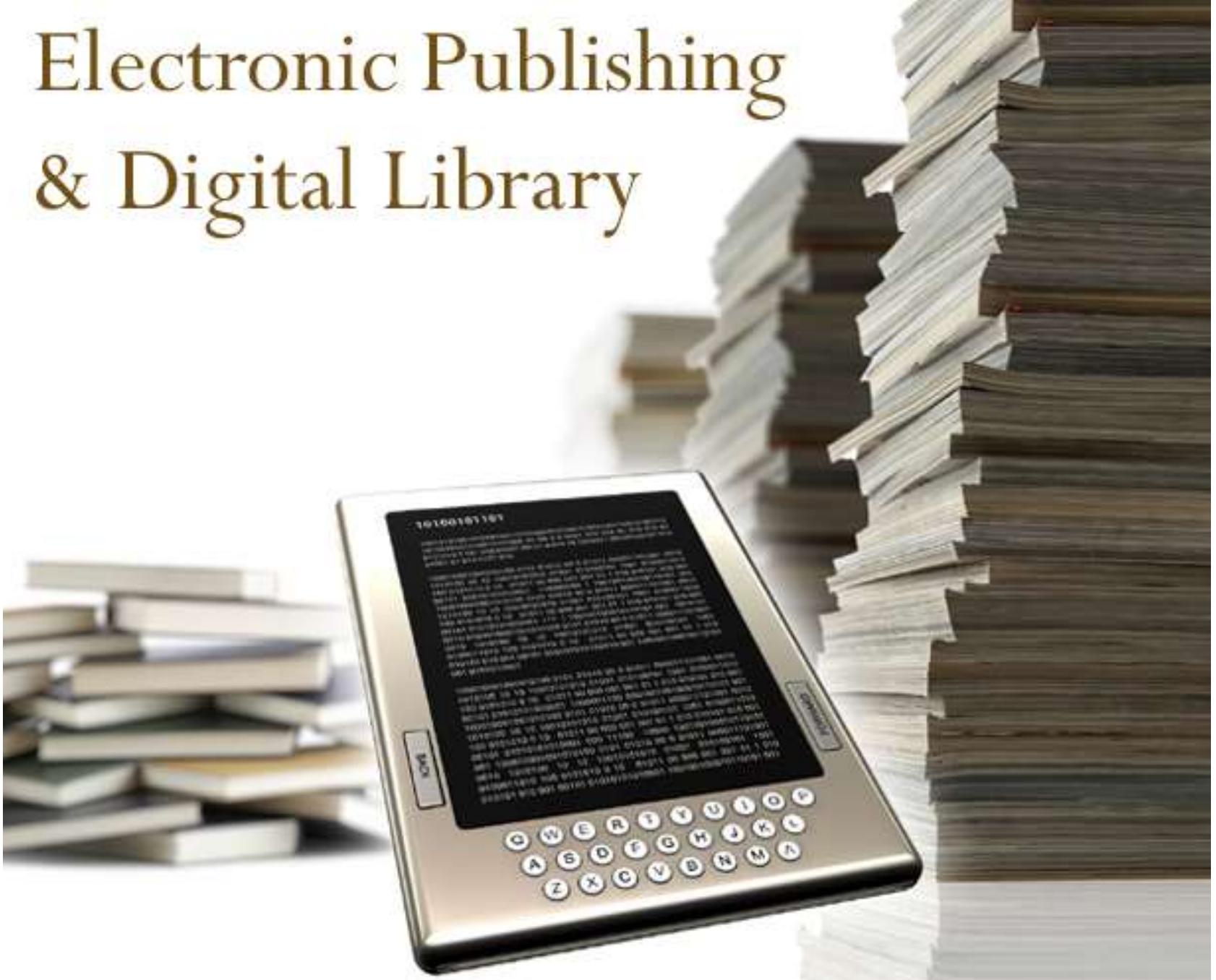
Electronic Publishing & Digital Library



Gordon Jean

Kamron Ackerman

First Edition, 2012

ISBN 978-81-323-1474-5

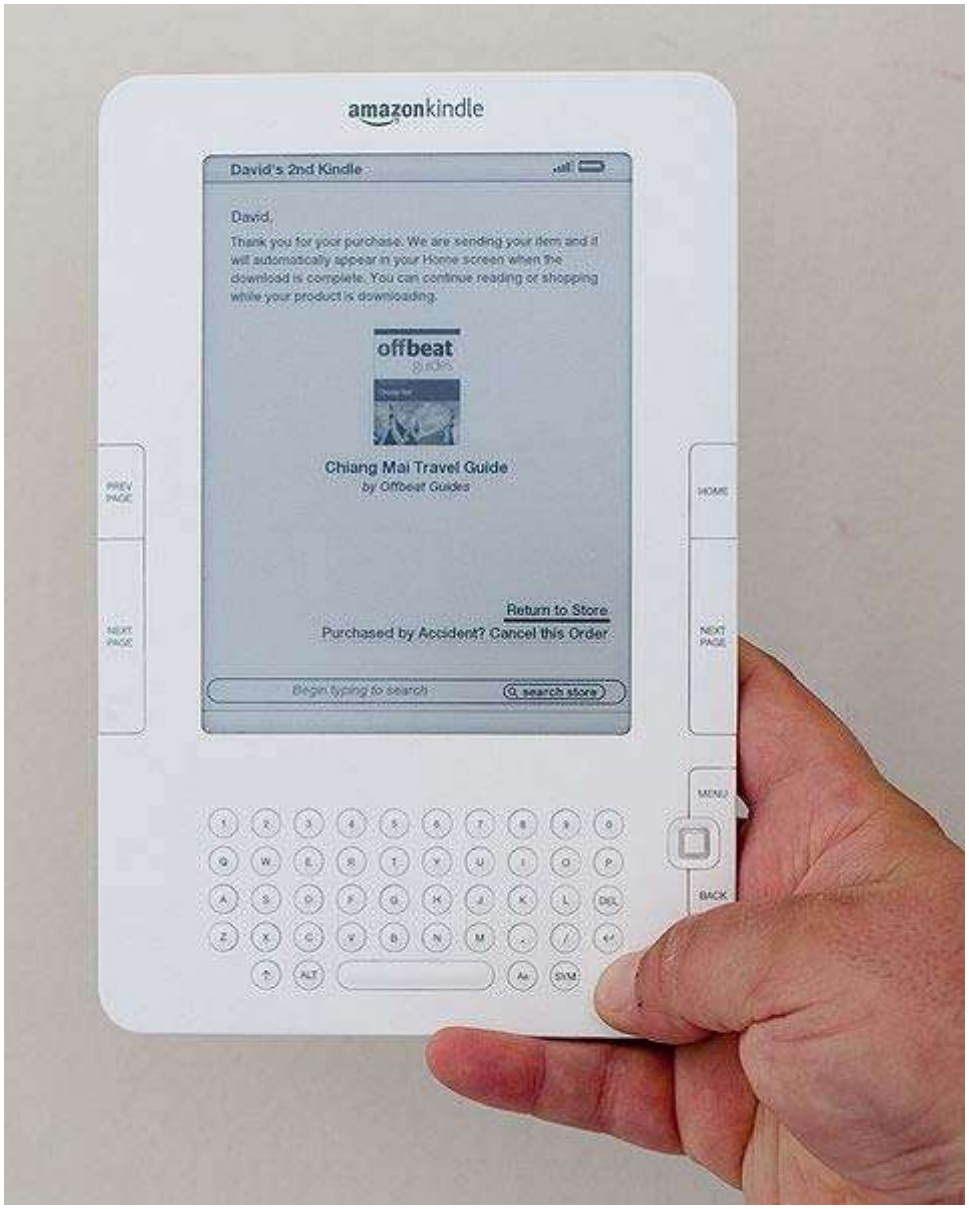
Published by:
College Publishing House
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

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.

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 V 3	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

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.

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.

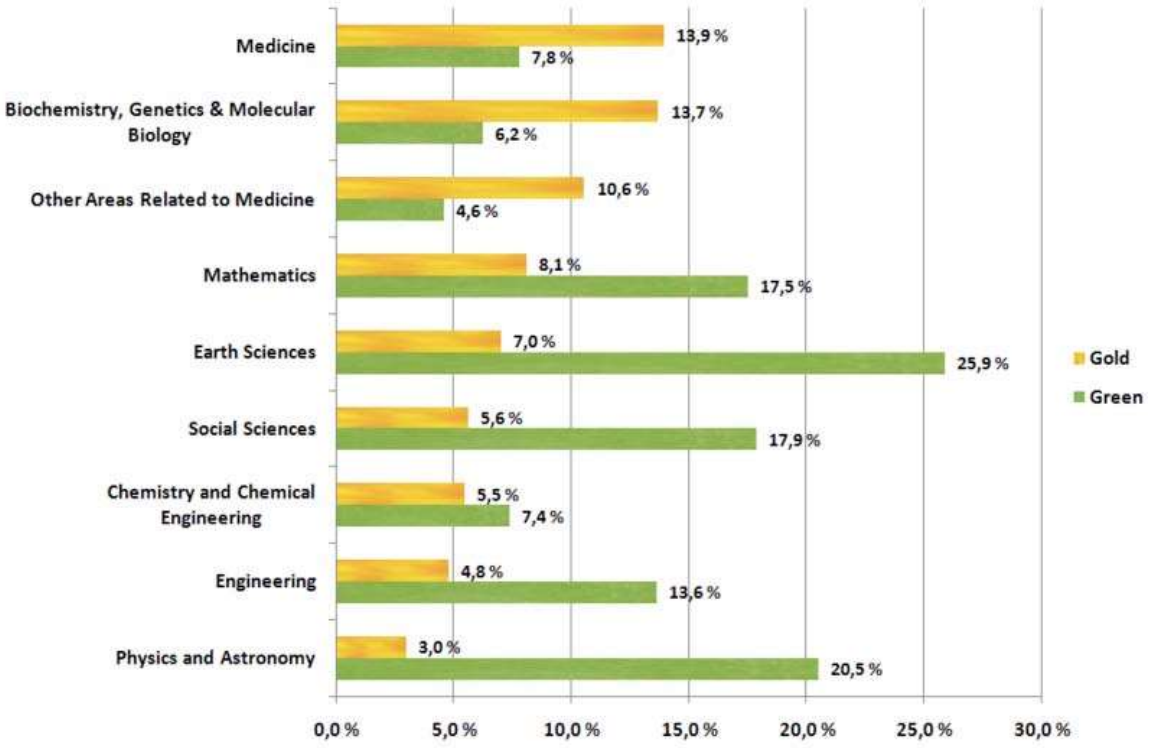
OA can be provided in two ways:

- "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

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 8

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.

Chapter 9

Introduction to 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.
- **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 10

Universal Library & University of Florida Digital Collections

Universal library

A **universal library** is a library with universal collections. This may be expressed in terms of it containing all existing information, useful information, all books, all works (regardless of format) or even all possible works. This ideal, although unrealizable, has influenced and continues to influence librarians and others and be a goal which is aspired to. Universal libraries are often assumed to have a complete set of useful features (such as finding aids, translation tools, alternative formats, etc).

History

The Library of Alexandria is generally regarded as the first library approaching universality, although this idea may be more mythical than real. It is estimated that at one time, this library contained between 30 and 70 percent of all works in existence. The re-founded modern library has a non-universal collections policy

As a phrase, the "universal library" can be traced back to the naturalist Conrad Gessner's *Bibliotheca universalis* of 1545.

In the 17th century, the ideal of universality continued to be attractive. The French librarian Gabriel Naudé wrote:

And therefore I shall ever think it extremely necessary, to collect for this purpose all sorts of books, (under such precautions, yet, as I shall establish) seeing a Library which is erected for the public benefit, ought to be universal; but which it can never be, unlesse it comprehend all the principal authors, that have written upon the great diversity of particular subjects, and chiefly upon all the arts and sciences; [...] For certainly there is nothing which renders a Library more recommendable, then when every man findes in it that which he is in search of ...

The universal library in fiction

Science fiction has used the device of a library which is universal in the sense that it not only contains all existing written works, but all possible written works. This idea appeared in Kurd Lasswitz's 1901 story "The Universal Library" and Borges's essay "The Total Library" before its more famous expression in Borges's story "The Library of Babel". Such a library, however, would be as useless as it would be complete. A similar idea was a planet called Memory Alpha, (from the Star Trek episode "The Lights of Zetar") which was the Federation's "storehouse of computer databases containing all cultural history and scientific data it has acquired.". It has been commented that the Internet already approaches this state.

In Discworld, Terry Pratchett's fantasy world, all libraries in the multiverse being connected in "L-space", effectively creating a single, semi-universal, library.

Modern times

With the advent of cheap, widely available digital storage, the ideal of universality, although still impossible to attain, has become closer to the feasible. Many projects are now attempting to collect a section of human knowledge into one database. These projects vary in breadth and scope, and none are complete. Examples include digitization projects such as Project Gutenberg and Carnegie-Mellon's Universal library, digital libraries which are using book scanning to collect public domain works.

Current barriers

Current barriers to the construction of a universal digital library include:

- Books have been lost. While the best-known lost-library may be the Library at Alexandria, wars, civil strife and natural disasters destroy libraries and archives on a regular basis. Further losses are due to neglect.
- Copyright. Many books are under copyright and current widespread business models require scarcity of books to remunerate authors.
- Censorship. Most jurisdictions have banned at least some banned books.
- Unpublished manuscripts. If unpublished manuscripts are included in the definition of *book,* catching newly-written manuscripts is likely to be a challenge.
- Current digitisation efforts are largely library-based and so materials deemed outside the scope of libraries are very poorly

University of Florida Digital Collections

The University of Florida Digital Collections (UFDC) are supported by the University of Florida Digital Library Center in the George A. Smathers Libraries at the University of Florida. The University of Florida Digital Collections (UFDC) comprise a constantly growing collection of digital resources from the University of Florida's library collections as well as partner institutions. Opening in April 2006, UFDC has added over 260,000 items - books, newspapers, oral histories, videos, photos, and more - with over 6 million pages.

Preservation and Access

All materials are freely and openly accessible (Open Access) and full text searchable. In UFDC, all items can be text searched simultaneously or certain collections can be selected for a faceted search. Because UFDC grew out of the efforts of the University of Florida Libraries' Preservation Department, all items are scanned at preservation quality and all are digitally preserved through the Florida Digital Archive. The page images are particularly important for the preservation of artifactually significant materials such as maps, artifacts, illustrated children's literature from the Baldwin Library of Historical Children's Literature, and other materials.

Interface and Usability

Because of the highly visual nature of so many items, the pages are displayed as zoomable images (through a JPG2000 server) and all can be browsed as thumbnails at the item and the collection level. Artifacts with multiple photos from multiple angles can be seen in motion, rotating in an Adobe Flash video view, and items can be searched by their geographic information (city, county, state, latitude and longitude) or viewed on a map through UFDC's use of the Google Maps API.

Statistics

UFDC includes books, articles, newspapers, photos, videos, audio, and more. As of November 2010, the collections had grown to over 6 million pages. This is an accelerated growth rate compared to the earlier million-page milestones (UFDC began in April 2006, reached 1 million pages in August 2007, and reached 2 million pages reached in July 2008).

Findability

Along with loading new items regularly, UFDC was optimized for search engine findability using static pages and adding RSS feeds in 2008. This process included creating static pages for all items on the mirror site UFDC2 (http://www.uflib.ufl.edu/ufdc2 instead of http://www.uflib.ufl.edu/ufdc) and then in November 2010 re-optimizing the site with the single URL (http://ufdc.ufl.edu/), creating

RSS feeds automatically for new items loaded and for all items, and optimizing all code for faster loading. These changes were necessary because of the deep web structure of UFDC which, like so many digital library collections, has many directory levels and dynamic URLs that cause difficulty for search engines even with properly implemented sitemaps.

Chapter 11

Discipline-Oriented Digital Libraries

Analytical sciences digital library

The **Analytical Sciences Digital Library** (ASDL) was founded in 2001 as one of several digital libraries in the National Science Digital Library, funded by the National Science Foundation. The library is a collection of peer-reviewed electronic resources on chemical measurements and instrumentation. The collection also contains materials on active learning and its use for effective instruction in the analytical sciences. The resources in ASDL are freely available and widely used by students, teachers and practitioners of analytical chemistry and its application areas. The site includes a collection of annotated electronic resources catalogued with the Open Archive Initiative and Dublin Core Metadata Initiative, making the collection searchable by any other group that uses these definitions.

Since 2004, the Journal of the Analytical Sciences Digital Library, JASDL, has published peer-reviewed online articles in the categories of courseware, labware, educational practices, undergraduate research, and poster sessions. The site is an open source site, and therefore publication is under the Creative Commons license. As a result authors retain copyright privileges and are free to publish their work elsewhere. This allows for a wider variety of published works to be available freely to the scientific community.

The ASDL community of users can participate in activities that promote analytical chemistry and help advance the education and training of future members of the analytical chemistry community by submitting and viewing posters for the ASDL online poster session, posting your information in the Analytical Sciences Professional Directory, contributing a url for consideration for the web collection, writing a JASDL article on an innovative aspect of your teaching or research with undergraduates or by volunteering to review new ASDL materials.

In 2007 ASDL partnered with the Analytical Division of the American Chemical Society to broaden their ability to serve as a connection place online for the analytical sciences community.

Anemi, The Digital Library of Modern Greek Studies

ANEMI is a digital library that aims to provide simple and quick access to a rich collection of digitized material related to Modern Greek Studies. Apart from finding bibliographic information, the researcher can also browse the documents themselves in electronic form. They can find a great number of old and rare documents, as well as recent publications for which their creators allowed the digitization and free distribution over the Internet.

Collections

- **Neoellinistis** the digital library of bibliographies, dictionaries and handbooks for the Greek Modern Studies

This collection provides free access to bibliographies, dictionaries, encyclopaedias, handbooks, chronologies and other tools related with Greek Modern Studies. It also provides the users with the possibility of locating relevant alternative information where the digitization is prohibited by the Greek law. The material that it is included in Neoellinistis is organised according to the work of Politis Alexis,*THE HANDBOOK OF MODERN GREEK STUDIES , Crete University Press, 2005*.

- **Greek Digital Bibliography 15th - 20th century**

By using the digital technology, the Greek Digital Library regenerates the national bibliographic landscape of the period 1476-1900. Entries that concern it, are catalogued electronically and, where feasible, are linked with the corresponding digital item. Our vision is, Anemi to become a union agent, which will be used in order to map out the registration and display of all digitization efforts, concerning the period 1476-1900, avoiding thus, redundant digitization efforts. Since December 2006, 8,000 bibliographic records are available in Anemi's data base as well as a vast amount of corresponding digitized pages.

- **Anacharsis**

Rare collections from the Library of University of Crete, with travel literature, have been catalogued. The bibliographical records are linked with the corresponding digital items which are hosted either in the local library or other bibliographic agents elsewhere.

- **Markos Mousouros**

It is a digital collection with books and archival materials about Crete. The main part of the items that are available in the collection come from the Library of the University of Crete. Among them is, the incunabulum: Etymologikon Mega, which was printed by the

Cretans Zacharias Kalliergis and Nikolaos Vlastos in Venice in 1499. It is the library's pride.

History

Anemi was founded in 2006 by the University of Crete Library. It embodies the final result of the Programme "Digital Library of Modern Greek Studies" which was funded by the Operational Programme "Information Society" (3rd CSF 2000-2006).

AnimalBase

Logo of the AnimalBase project

AnimalBase is a project brought to life in 2004 and is maintained by the University of Göttingen, Germany. The goal of the AnimalBase project is to digitize early zoological literature, provide copyright-free open access to zoological works, and provide manually verified lists of names of zoological genera and species as a free resource for the public. AnimalBase contributed to opening up the classical taxonomic literature, which is considered as useful because access to early literature (especially for the late 18th century) can be difficult for researchers who need the old sources for their taxonomic research.

AnimalBase data are public domain. The public use of AnimalBase data is not restricted or conditioned. AnimalBase covers all zoological disciplines. In the field of biodiversity informatics AnimalBase is unique in providing links between the names of generic and specific taxa and their digitized original descriptions, with a special focus on literature and names published prior to 1800.

History

The project was initiated in 2003 with funding from the German Research Foundation (DFG). The database and its web interface began development in 2004 and was launched online in 2005. The web interface was designed to be accessible to older generations of computers and web browsers. Between 2003 and 2005 approximately 400 zoological publications from the 1550s to 1770 were digitized while 10,000 linked zoological names were incorporated into the database. A second effort (2008–2011) has included additional

digitized literature giving access to several ten thousand more zoological names that were extracted from the original sources and linked with their digitized original descriptions.

Digitization and linking to digitized literature from other sources

Early zoological publications are digitized under the highest quality standards by the Center for Retrospective Digitization in Göttingen (Göttinger Digitalisierungszentrum, GDZ) of the Göttingen State and University Library (Staats- und Universitätsbibliothek Göttingen, SUB), which is one of Germany's largest libraries, including over 4.5 million volumes. If available, AnimalBase also provides links to the digitized public domain content of other providers, such as the Biodiversity Heritage Library (BHL) or Gallica.

Extracting taxonomic names from the literature

Zoological names of generic and specific taxa that were established in early zoological publications have been entered manually into the database, including original and corrected spellings of names, type localities and page numbers where these names were originally established in the publications. Furthermore, the names have been compared to the entries in Sherborn's Index Animalium (1902, 1922–1933) and the Nomenclator Zoologicus by Neave (1939/1940, updated). The nomenclatural status of names has been verified under the current edition of the International Code of Zoological Nomenclature (ICZN). Possible discrepancies between older databases and the findings from AnimalBase name research (i.e. nomenclatural priority or incorrect spellings) have occasionally been discussed in the comments provided for each taxon. The entire process has followed the established AnimalBase standard.

AnimalBase standard

Specific and generic names of taxa are entered manually with reference to the original description. A basic entry of a taxon name follows the AnimalBase standard, which is composed of the following guidelines:

- The correct original spelling of the generic and specific name is checked according to the nomenclatural rules (ICZN Code, current edition).
- Specific names are strictly entered in combination with the genus to which they were attributed in the original description. In cases where the generic name was spelled incorrectly by the author who established the specific name, the AnimalBase Team combines the specific name with the correct spelling of that genus.
- Contingently incorrect original spellings of the taxon name (i.e. being different from the correct spelling) are entered in addition, retaining the original use of diacritics, ligations, upper-case letters for specific names, hyphens, spaces between words, incorrect subsequent spellings of the generic name in a genus-species combination used by the author.
- The name of the author is provided, spelled according to the name given on the title page of the original source (examples: Linnæus 1758, Linné 1766).

- The year or date of publication is provided, as determined by the nomenclatural rules (true date of publication).
- The gender treatment is provided (changeable or unchangeable specific names, according to post-classical Latin grammar rules). It is often difficult to determine whether a Latin name is an adjective or not, so resulting AnimalBase entries of this data may not be 100% reliable.
- A link to the digitized publication containing the original description is provided.
- The page of the original description of the taxon name is provided (which is the page where the name was first mentioned and made available). If the described animal is shown on figure plates and without a name mentioned on the plates, this is not necessarily indicated.
- The type locality is provided as given in the original description if it is easily recognizable or can be inferred implicitly from the work. This procedure may be inaccurate (because the type locality is the locality where the name-bearing types came from, which are not researched by the AnimalBase team), but certainly provides useful information.
- The higher animal group in which the taxon is classified is provided (phylum, class, order or likewise clade).
- Entries are cross-checked with Index Animalium and (in case of genera) with Nomenclator Zoologicus. Incorrect entries (subsequent uses of previously established names, which Sherborn often listed incorrectly as new or nomina nuda) and incorrect spellings of names in these databases are not copied into AnimalBase, if detected as such.
- Incorrect subsequent spellings are not generally considered as new taxa, but may be listed as not available and discussed if the name was mentioned as a new name in the Index Animalium or is otherwise important.
- All animal groups are treated consistently by following the ICZN rules (fish names follow the same rules as insect names).

The AnimalBase standard allows for easy access to the primary data of the original description. The AnimalBase team does not align the spellings or authorship of each originally established name with those contained in various different zoological databases, specialized by discipline or region. The primarily scope of AnimalBase is to provide links to the digitized original descriptions, but it can also be consulted for correct spellings and authorships of zoological names. In this regard, AnimalBase may potentially provide a useful update of Sherborn's Index Animalium.

AnimalBase also provides the option to combine the original names with their current allocations (current genus-species combinations). For example, detailed biological information and pictures are available for 2,500 species of European non-marine molluscs, which includes more than 6,000 photographs. AnimalBase is a collaborative and open resource project, all registered collaborators are able to correct or enter data. This includes uploading pictures (copyright-free) of animal species, biological data including measurements and diagnostic characters, distributional data, the current conservation status, etc. The pictures and data provided by AnimalBase are expressly permitted for use on other websites (copied or linked).

Astrophysics Data System

Logo of the ADS

The **Astrophysics Data System** (usually referred to as **ADS**), developed by the National Aeronautics and Space Administration (NASA), is an online database of over eight million astronomy and physics papers from both peer reviewed and non-peer reviewed sources. Abstracts are available free online for almost all articles, and full scanned articles are available in Graphics Interchange Format (GIF) and Portable Document Format (PDF) for older articles. New articles have links to electronic versions hosted at the journal's webpage, but these are typically available only by subscription (which most astronomy research facilities have). It is managed by the Harvard–Smithsonian Center for Astrophysics.

ADS is a powerful research tool and has had a significant impact on the efficiency of astronomical research since it was launched in 1992. Literature searches that previously would have taken days or weeks can now be carried out in seconds via the ADS search engine, custom-built for astronomical needs. Studies have found that the benefit to astronomy of the ADS is equivalent to several hundred million US dollars annually, and the system is estimated to have tripled the readership of astronomical journals.

Use of ADS is almost universal among astronomers worldwide, and therefore ADS usage statistics can be used to analyze global trends in astronomical research. These studies have revealed that the amount of research an astronomer carries out is related to the per capita gross domestic product (GDP) of the country in which he/she is based, and that the number of astronomers in a country is proportional to the GDP of that country, so the total amount of research done in a country is proportional to the square of its GDP divided by its population.

History

For many years, a growing problem in astronomical research (as in other academic disciplines) was that the number of papers published in the major astronomical journals was increasing steadily, meaning astronomers were able to read less and less of the latest research findings. During the 1980s, astronomers saw that the nascent technologies which

formed the basis of the Internet could eventually be used to build an electronic indexing system of astronomical research papers which would allow astronomers to keep abreast of a much greater range of research.

The first suggestion of a database of journal paper abstracts was made at a conference on *Astronomy from Large Data-bases* held in Garching bei München in 1987. Initial development of an electronic system for accessing astrophysical abstracts took place during the following two years; in 1991 discussions took place on how to integrate ADS with the SIMBAD database, containing all available catalog designations for objects outside the solar system, to create a system where astronomers could search for all the papers written about a given object.

An initial version of ADS, with a database consisting of 40 papers, was created as a proof of concept in 1988, and the ADS database was successfully connected with the SIMBAD database in the summer of 1993. The creators believed this was the first use of the Internet to allow simultaneous querying of transatlantic scientific databases. Until 1994, the service was available via proprietary network software, but it was transferred to the nascent World Wide Web early that year. The number of users of the service quadrupled in the five weeks following the introduction of the ADS web-based service.

At first, the journal articles available via ADS were scanned bitmaps created from the paper journals, but from 1995 onwards, the *Astrophysical Journal* began to publish an on-line edition, soon followed by the other main journals such as *Astronomy and Astrophysics* and the *Monthly Notices of the Royal Astronomical Society*. ADS provided links to these electronic editions from their first appearance. Since about 1995, the number of ADS users has doubled roughly every two years. ADS now has agreements with almost all astronomical journals, who supply abstracts. Scanned articles from as far back as the early 19th century are available via the service, which now contains over eight million documents. The service is distributed worldwide, with twelve mirror sites in twelve countries on five continents, with the database synchronized by means of weekly updates using rsync, a mirroring utility which allows updates to only the portions of the database which have changed. All updates are triggered centrally, but they initiate scripts at the mirror sites which "pull" updated data from the main ADS servers.

Data in the system

1284 papers about M101 are available through ADS, from as long ago as 1850

Papers are indexed within the database by their bibliographic record, containing the details of the journal they were published in and various associated metadata, such as author lists, references and citations. Originally this data was stored in ASCII format, but eventually the limitations of this encouraged the database maintainers to migrate all records to an XML (Extensible Markup Language) format in 2000. Bibliographic records are now stored as an XML element, with sub-elements for the various metadata.

Since the advent of online editions of journals, abstracts are loaded into the ADS on or before the publication date of articles, with the full journal text available to subscribers. Older articles have been scanned, and an abstract is created using optical character recognition software. Scanned articles from before about 1995 are usually available free, by agreement with the journal publishers.

Scanned articles are stored in TIFF format, at both medium and high resolution. The TIFF files are converted on demand into GIF files for on-screen viewing, and PDF or PostScript files for printing. The generated files are then cached to eliminate needlessly frequent regenerations for popular articles. As of 2000, ADS contained 250 GB of scans, which consisted of 1,128,955 article pages comprising 138,789 articles. By 2005 this had

grown to 650 GB, and is expected to grow further, to about 900 GB by 2007. No further information has been published.

The database initially contained only astronomical references, but has now grown to incorporate three databases, covering astronomy (including planetary sciences and solar physics) references, physics (including instrumentation and geosciences) references, as well as preprints of scientific papers from arXiv. The astronomy database is by far the most advanced and its use accounts for about 85% of the total ADS usage. Articles are assigned to the different databases according to the subject rather than the journal they are published in, so that articles from any one journal might appear in all three subject databases. The separation of the databases allows searching in each discipline to be tailored, so that words can automatically be given different weight functions in different database searches, depending on how common they are in the relevant field.

Data in the preprint archive is updated daily from the arXiv, the main repository of physics and astronomy preprints. The advent of preprint servers has, like ADS, had a significant impact on the rate of astronomical research, as papers are often made available from preprint servers weeks or months before they are published in the journals. The incorporation of preprints from the arXiv into ADS means that the search engine can return the most current research available, with the caveat that preprints may not have been peer reviewed or proofread to the required standard for publication in the main journals. ADS's database links preprints with subsequently published articles wherever possible, so that citation and reference searches will return links to the journal article where the preprint was cited.

Software and hardware

The software runs on a system that was written specifically for it, allowing for extensive customization for astronomical needs that would not have been possible with general purpose database software. The scripts are designed to be as platform independent as possible, given the need to facilitate mirroring on different systems around the world, although the growing use of Linux as the operating system of choice within astronomy has led to increasing optimization of the scripts for installation on that platform.

The main ADS server is located at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and is a dual 64-bit X86 Intel server with two quad-core 3.0 GHz CPUs and 32 GB of RAM, running the CentOS 5.4 Linux distribution. Mirrors are located in Brazil, China, Chile, France, Germany, India, Indonesia, Japan, Russia, South Korea, United Kingdom, and the Ukraine.

Indexing

ADS currently receives abstracts or tables of contents from almost two hundred journal sources. The service may receive data referring to the same article from multiple sources, and creates one bibliographic reference based on the most accurate data from each source. The common use of TeX and LaTeX by almost all scientific journals greatly

facilitates the incorporation of bibliographic data into the system in a standardized format, and importing HTML-coded web-based articles is also simple. ADS utilizes Perl scripts for importing, processing and standardizing bibliographic data.

The apparently mundane task of converting author names into a standard *Surname, Initial* format is actually one of the more difficult to automate, due to the wide variety of naming conventions around the world and the possibility that a given name such as Davis could be a first name, middle name or surname. The accurate conversion of names requires a detailed knowledge of the names of authors active in astronomy, and ADS maintains an extensive database of author names, which is also used in searching the database (see below).

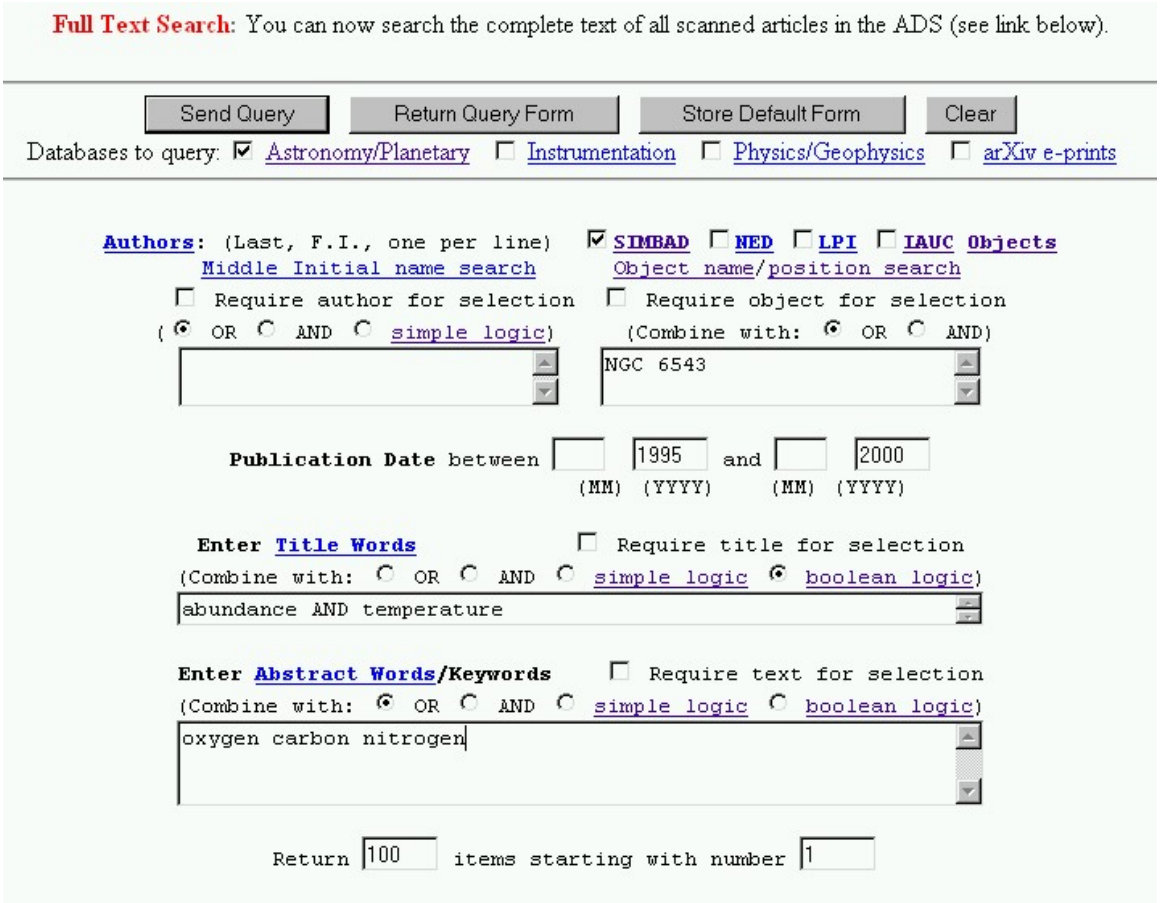
For electronic articles, a list of the references given at the end of the article is easily extracted. For scanned articles, reference extraction relies on OCR. The reference database can then be "inverted" to list the citations for each paper in the database. Citation lists have been used in the past to identify popular articles missing from the database; mostly these were from before 1975 and have now been added to the system.

Coverage

The database now contains over eight million articles. In the cases of the major journals of astronomy (*Astrophysical Journal*, *Astronomical Journal*, *Astronomy and Astrophysics*, *Publications of the Astronomical Society of the Pacific* and the *Monthly Notices of the Royal Astronomical Society*), coverage is complete, with all issues indexed from number 1 to the present. These journals account for about two-thirds of the papers in the database, with the rest consisting of papers published in over 100 other journals from around the world.

While the database contains the complete contents of all the major journals and many minor ones as well, its coverage of references and citations is much less complete. References in and citations of articles in the major journals are fairly complete, but references such as "private communication", "in press" or "in preparation" cannot be matched, and author errors in reference listings also introduce potential errors. Astronomical papers may cite and be cited by articles in journals which fall outside the scope of ADS, such as chemistry, mathematics or biology journals.

Search engine



An example of a complex search combining object, title and abstract queries with a date filter

Since its inception, the ADS has developed a highly complex search engine to query the abstract and object databases. The search engine is tailor-made for searching astronomical abstracts, and the engine and its user interface assume that the user is well-versed in astronomy and able to interpret search results which are designed to return more than just the most relevant papers. The database can be queried for author names, astronomical object names, title words, and words in the abstract text, and results can be filtered according to a number of criteria. It works by first gathering synonyms and simplifying search terms as described above, and then generating an "inverted file", which is a list of all the documents matching each search term. The user-selected logic and filters are then applied to this inverted list to generate the final search results.

Author name queries

The system indexes author names by surname and initials, and accounts for the possible variations in spelling of names using a list of variations. This is common in the case of names including accents such as umlauts and transliterations from Arabic or Cyrillic script. An example of an entry in the author synonym list is:

AFANASJEV, V
AFANAS'EV, V
AFANAS'IEV, V
AFANASEV, V
AFANASYEV, V
AFANS'IEV, V
AFANSEV, V

Object name searches

The capability to search for papers on specific astronomical objects is one of ADS's most powerful tools. The system uses data from the SIMBAD, the NASA/IPAC Extragalactic Database, the International Astronomical Union Circulars and the Lunar and Planetary Institute to identify papers referring to a given object, and can also search by object position, listing papers which concern objects within a 10 arcminute radius of a given Right Ascension and Declination. These databases combine the many catalogue designations an object might have, so that a search for the Pleiades will also find papers which list the famous open cluster in Taurus under any of its other catalog designations or popular names, such as M45, the Seven Sisters or Melotte 22.

Title and abstract searches

The search engine first filters search terms in several ways. An M followed by a space or hyphen has the space or hyphen removed, so that searching for Messier catalogue objects is simplified and a user input of M45, M 45 or M-45 all result in the same query being executed; similarly, NGC designations and common search terms such as Shoemaker Levy and T Tauri are stripped of spaces. Unimportant words such as AT, OR and TO are stripped out, although in some cases case sensitivity is maintained, so that while **a**nd is ignored, **A**nd is converted to "Andromedae", and **H**er is converted to "Herculis", but **h**er is ignored.

Synonym replacement

Once search terms have been pre-processed, the database is queried with the revised search term, as well as synonyms for it. As well as simple synonym replacement such as searching for both plural and singular forms, ADS also searches for a large number of specifically astronomical synonyms. For example, spectrograph and spectroscope have basically the same meaning, and in an astronomical context metallicity and abundance are also synonymous. ADS's synonym list was created manually, by grouping the list of words in the database according to similar meanings.

As well as English language synonyms, ADS also searches for English translations of foreign search terms and vice versa, so that a search for the French word *soleil* retrieves references to Sun, and papers in languages other than English can be returned by English search terms.

Synonym replacement can be disabled if required, so that a rare term which is a synonym of a much more common term (such as 'dateline' rather than 'date') can be searched for specifically.

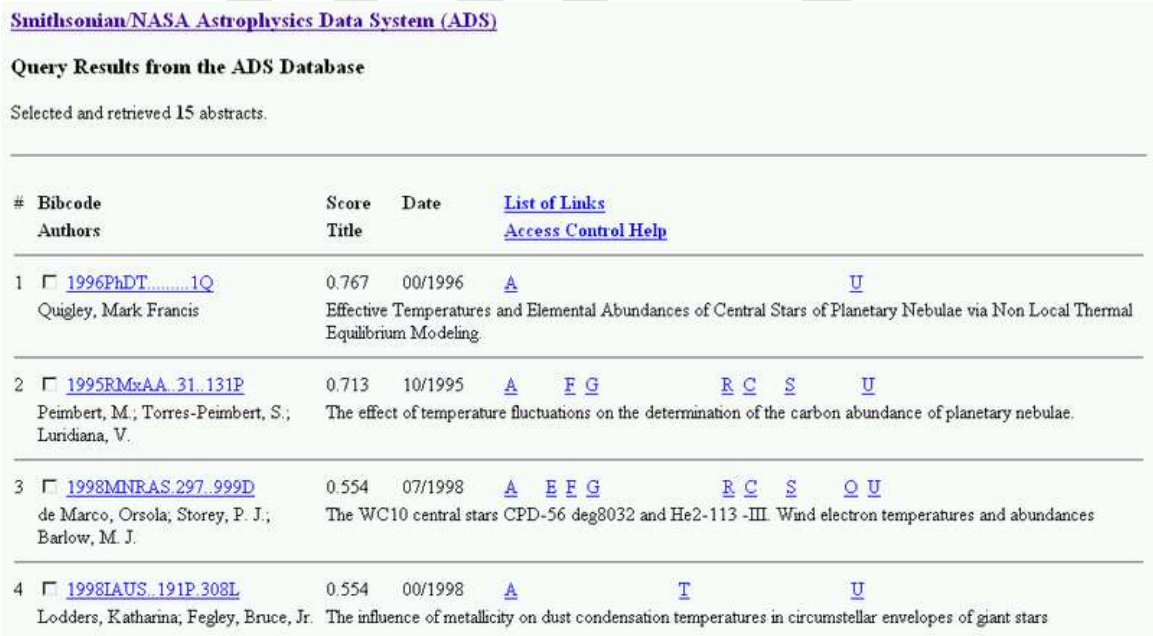
Selection logic

The search engine allows selection logic both within fields and between fields. Search terms in each field can be combined with OR, AND, simple logic or Boolean logic, and the user can specify which fields must be matched in the search results. This allows complex searches to be built; for example, the user could search for papers concerning NGC 6543 OR NGC 7009, with the paper titles containing (radius OR velocity) AND NOT (abundance OR temperature).

Result filtering

Search results can be filtered according to a number of criteria, including specifying a range of years such as '1945 to 1975', '2000 to the present day' or 'before 1900', and what type of journal the article appears in – non-peer reviewed articles such as conference proceedings can be excluded or specifically searched for, or specific journals can be included in or excluded from the search.

Search results



Smithsonian/NASA Astrophysics Data System (ADS)

Query Results from the ADS Database

Selected and retrieved 15 abstracts.

#	Bibcode Authors	Score Title	Date	List of Links Access Control Help
1	1996PhDT.....1Q Quigley, Mark Francis	0.767 Effective Temperatures and Elemental Abundances of Central Stars of Planetary Nebulae via Non Local Thermal Equilibrium Modeling.	00/1996	A U
2	1995RMxAA..31..131P Peimbert, M.; Torres-Peimbert, S.; Luridiana, V.	0.713 The effect of temperature fluctuations on the determination of the carbon abundance of planetary nebulae.	10/1995	A F G R C S U
3	1998MNRAS.297..999D de Marco, Orsola; Storey, P. J.; Barlow, M. J.	0.554 The WC10 central stars CPD-56 deg8032 and He2-113 -III. Wind electron temperatures and abundances	07/1998	A E F G R C S O U
4	1998IAUS..191P.308L Lodders, Katharina; Fegley, Bruce, Jr.	0.554 The influence of metallicity on dust condensation temperatures in circumstellar envelopes of giant stars	00/1998	A T U

Search results page from ADS – A, F, G, C, R etc. are links to associated data for each abstract such as full-text article, citations, also-read papers and so on.

Although it was conceived as a means of accessing abstracts and papers, ADS provides a substantial amount of ancillary information along with search results. For each abstract returned, links are provided to other papers in the database which are referenced, and

which cite the paper, and a link is provided to a preprint, where one exists. The system also generates a link to 'also-read' articles – that is, those which have been most commonly accessed by those reading the article. In this way, an ADS user can determine which papers are of most interest to astronomers who are interested in the subject of a given paper.

Also returned are links to the SIMBAD and/or NASA Extragalactic Database object name databases, via which a user can quickly find out basic observational data about the objects analyzed in a paper, and find further papers on those objects.

Impact on astronomy

ADS is almost universally used as a research tool among astronomers, and there are several studies that have estimated quantitatively how much more efficient ADS has made astronomy; one estimated that ADS increased the efficiency of astronomical research by 333 full-time equivalent research years per year, and another found that in 2002 its effect was equivalent to 736 full-time researchers, or all the astronomical research done in France. ADS has allowed literature searches that would previously have taken days or weeks to carry out to be completed in seconds, and it is estimated that ADS has increased the readership and use of the astronomical literature by a factor of about three since its inception.

In monetary terms, this increase in efficiency represents a considerable amount. There are about 12,000 active astronomical researchers worldwide, so ADS is the equivalent of about 5% of the working population of astronomers. The global astronomical research budget is estimated at between 4,000 and 5,000 million USD, so the value of ADS to astronomy would be about 200–250 million USD annually. Its operating budget is a small fraction of this amount.

The great importance of ADS to astronomers has been recognized by the United Nations, the General Assembly of which has commended ADS on its work and success, particularly noting its importance to astronomers in the developing world, in reports of the United Nations Committee on the Peaceful Uses of Outer Space. A 2002 report by a visiting committee to the Center for Astrophysics, meanwhile, said that the service had "revolutionized the use of the astronomical literature", and was "probably the most valuable single contribution to astronomy research that the CfA has made in its lifetime".

Sociological studies using ADS

Because it is used almost universally by astronomers, ADS can reveal much about how astronomical research is distributed around the world. Most users access the system from institutes of higher education, whose IP address can easily be used to determine the user's geographical location. Studies reveal that the highest per-capita users of ADS are France and Netherlands-based astronomers, and while more developed countries (measured by GDP per capita) use the system more than less developed countries; the relationship between GDP per capita and ADS use is not linear. The range of ADS usage per capita

far exceeds the range of GDPs per capita, and basic research carried out in a country, as measured by ADS usage, has been found to be proportional to the square of the country's GDP divided by its population.

ADS usage statistics also suggest that astronomers in more developed countries tend to be more productive than those in less developed countries. The amount of basic research carried out is proportional to the number of astronomers in a country multiplied by the GDP per capita. Statistics also imply that astronomers in European cultures carry out about three times as much research as those in Asian cultures, perhaps suggesting cultural differences in the importance attached to astronomical research.

ADS has also been used to show that the fraction of single-author astronomy papers has decreased substantially since 1975 and that astronomical papers with more than 50 authors have become more common since 1990.

Comparison to what the National Library of Medicine has done

The National Library of Medicine also has a free to users, supported by tax dollars, online database of scientific publications, PubMed.

Avalon Project

The **Avalon Project** is a digital library of documents relating to law, history and diplomacy. The project is part of the Yale Law School Lillian Goldman Law Library.

The project contains online electronic copies of documents dating back to the beginning of history, making it possible to study the original text of not only very famous documents such as the Magna Carta, the English Bill of Rights, and the United States Bill of Rights, but also the text of less well known but significant documents which mark turning points in the history of law and rights.

The site has full search facilities and a facility to electronically compare the text of two documents.

At the same website is *Project Diana: An Online Human Rights Archive*.

International Children's Digital Library

The **International Children's Digital Library Foundation** (ICDL) is a free online library of digitized children's books in many languages from various countries. Designed specifically for use by children ages 3 to 13, the Library is a housed by the International Children's Digital Library Foundation and was originally developed in the College of Information Studies and the Human-Computer Interaction Laboratory at the University of Maryland, College Park.

Children can search for books by location, color, length, intended age group, content type, and emotional quality, among other qualifiers. An advanced search option is also provided for more experienced or older users, and all users can register to save search preferences and favorite books.

Books are selected based on quality and appropriateness and are presented in their original language with copyright permission from publishers or authors. The Library's ultimate goal is to foster a love of reading, a readiness to learn, and a response to the challenges of world literacy.

History

The International Children's Digital Library was initially launched in November 2002 under the direction of University of Maryland Computer Science professor Dr. Allison Druin and in collaboration with researchers from other fields, such as information studies, art, psychology, and education, in order to better understand children's online habits and to encourage a love of reading and increased literacy. Children from Bowie, Maryland, tested the original Java prototype from 1999–2002, and since then children from five locations around the world have also contributed to the design process as the library's "Kidsteam Program".

The 2002-2005 phase of development saw a working model of the Library accessed by over one million users around the world and home to 1,000 books. Funding initially came from the National Science Foundation and the Institute of Museum and Library Services.

In April 2006, the International Children's Digital Library became part of the newly formed International Children's Digital Library Foundation, a non-profit corporation under the leadership of Tim Browne as Executive Director and original project leaders Dr. Allison Druin, Dr. Ben Bederson, and Dr. Ann Weeks as Directors. The Library's principal support comes from the Library of Congress, National Science Foundation, the Institute of Museum and Library Services, and Microsoft Research.

Christian Classics Ethereal Library

Christian Classics Ethereal Library

URL	http://www.ccel.org
Commercial?	No
Type of site	Digital library
Registration	None
Owner	Calvin College
Created by	Harry Plantinga
Launched	1993

The **Christian Classics Ethereal Library** (CCEL) is a digital library that provides free electronic copies of Christian scripture and literature texts.

CCEL is a volunteer-based project founded and directed by Harry Plantinga, a professor of computer science at Calvin College. It was initiated at Wheaton college in 1993. and currently supported by Calvin College.

The purpose of the CCEL is simply "to build up Christ's church and to address fundamental questions of the faith." The documents in the library express a variety of theological views, sometimes conflicting with those of Calvin College.

CCEL stores texts in Theological Markup Language (ThML) format and automatically converts them into other formats such as HTML or Portable Document Format (PDF). Although they use mainly Public Domain texts, they claim copyright on all their formatting. Users must login to their website to download all formatted versions of the text.

CCEL is funded by advertisements, sales of cd-roms (available since 1997), sales of some books not freely downloadable, and individual gifts. Calvin College has also provided them with space, network access, and significant financial support.

As of 2006, the library was recording about 200,000 page views per day and providing about 2 TB of information (equivalent to over a million books) in a month.

A 2002 reviewer acknowledged that while the site is "intended to be a basic online theological library," it was actually much more valuable than that: it is "a treasure of primary sources for anyone teaching Western Civilization or more specialized courses in medieval or Reformation history." They also specifically noted that the ability to search the music "for specific note patterns" was valuable to musicologists.

As of 2005, the primary users of the library fell into three main categories. These are university professors and their students using texts from the library as required reading without running up the students' bill for textbooks, people preparing sermons and Bible studies, and those reading for individual edification.

Collection of Computer Science Bibliographies

The **Collection of Computer Science Bibliographies** is one of the oldest (if not the oldest) bibliography collections freely accessible on the Internet. It is a collection of bibliographies of scientific literature in computer science and (computational) mathematics from various sources, covering most aspects of computer science. The bibliographies are updated weekly from their original locations.

As of 2009 the collection contains more than 2.8 million unique references (mostly to journal articles, conference papers and technical reports), clustered in about 1700 bibliographies, and consists of more than 4.4 Gb (950 Mb gzipped) of BibTeX entries. More than 600,000 references contain cross-references to citing or cited publications.

More than 1 million references contain URLs to an online version of the paper. Abstracts are available for more than 1 million entries. There are more than 2,000 links to other sites carrying bibliographic information.

Duplicates and links

As the Collection of Computer Science Bibliographies consists of many subcollections there is a substantial overlap (roughly 1/3). At the end of 2008 there were more than 4.2 million records which represent about 2.8 million unique (in terms of normalized title and authors' last names) bibliographic entries.

The number of duplicates may be seen as a feature, because there is a greater chance for finding a freely available full text PDF of a searched publication. Publications are clustered by title and last names of authors, so it is possible to find an extended version (e.g. Technical Report or Thesis) of an article.

There are also generated links to Google Scholar and IEEE Xplore in the case no full text link was available directly. Almost every bibliographic query may be served in RSS format.

Major subcollections

- arXiv
- BibNet
- CiteSeer
- DBLP
- LEABib
- NCSTRL

History

The collection was started in 1993 by Alf-Christian Achilles with a simple email-based interface and limited number of entries. One year later the first web interface has been made available. Since then the Collection was maintained by Achilles in his spare time. At the end of 2002 the maintenance has been handed over to Paul Ortyl.

Judaic Digital Library

The **Judaic Digital Library** is a specialized collection of Judaica titles designed mostly for educators, clergy, as well as advanced students of Hebrew Bible and Jewish Studies. Prepared in Secure Searchable Image Format, it allows its publisher, Varda Books, to deliver authoritative electronic editions of previously published by other publishers, typographically-complex books, with advanced online functionality for computer-assisted reading and research.

Most books of commentaries on Hebrew Bible feature thousands and tens of thousands of "live" biblical references: clicking on any of these references produces The JPS Hebrew-English Tanakh in a separate window with indicated biblical Hebrew verse side-by-side with its English translation!

JDL includes electronic editions of two major encyclopedias -- the classic 13-volume Jewish Encyclopedia and Hastings' Encyclopedia of Religion and Ethics -- along with more than 200 of important, many award-winning, volumes covering virtually all aspects of Jewish experience.

Chapter 12

Geographic Region-Oriented Digital Libraries

African Journals OnLine

African Journals OnLine (**AJOL**) is a non-profit organisation dedicated to improving the online visibility of and access to the published scholarly research of African-based academics. By using the internet as a gateway, AJOL aims to enhance conditions for African learning to be translated into African development.

Information inequality

Of the 50 countries throughout the world classified as Least Developed Countries (LDCs) by the United Nations, 33 are in Africa. There is widespread awareness of the importance of education in addressing poverty in the long term, usually with an emphasis on primary and secondary education. A concurrent focus on higher education in the continent is also needed for African countries to sustainably develop their capacity and economies and lift the region out of under-development.

Primarily due to difficulties accessing them, African research papers have been under-utilised, under-valued and under-cited in the international and African research arenas. To date, the main information resources, published journals and journal articles available to and used by researchers, librarians and students in Africa are the same as those used in Europe and America. This is because information from the developed world is usually more readily available than that of developing countries. However, it does not adequately reflect the research output of Africa and is not always relevant or appropriate for higher education in Africa. Although access to global information resources is essential; equally important and essential is access to the local research output from the continent.

Despite the wide range of capacity and resources within and between African countries, a legitimate generalization is that strengthening research and research-publishing are crucial priorities for improving higher education in Africa. At the same time as information sources from the developed world are currently made available for free *to* Africa (such as HINARI, AGORA, OARE, JSTOR African Access Initiative, and Aluka), there needs to be a corresponding focus on the online availability of information *from*

Africa if increased local capacity in research and dissemination is to be attained. To this end, in a high-tech, information hungry and rapidly globalising world, higher education in Africa needs technological tools to share and build on its own research output with neighbouring countries and the rest of the world.

Scholarly journals remain a vital and entrenched means of academic communication. In the information age, providing electronic access to journals is becoming the norm if that research is to reach the international audience who need to be aware of it. Many worthy peer-reviewed scholarly journals publishing from Africa lack the means to host their content online in isolation. Others do have sufficient resources but cannot attain the online visibility necessary to increase awareness of the valuable research contained within. There is a need to support the ongoing functioning and sustainability of journals publishing research from Africa.

Increasing access to African information

The mission of AJOL is to support African research and counter the "North-South" and "West-East" inequality of information flow by facilitating awareness of and access to research published in Africa. Information from developed countries is not necessarily as relevant or appropriate for Africa as that from within the continent. AJOL provides an online system for the aggregation of African-published scholarly journals and offers global access to and visibility of the research output of the continent. As such, AJOL's primary beneficiaries are scholarly, peer-reviewed, African-published journals, and secondary (also direct) beneficiaries are African and international members of the scholarly community needing to access African-published research.

AJOL hosts African-published, peer-reviewed scholarly journals for free – and includes both open access and subscription-based journals. The meta-data of all participating journals is open access on the AJOL website. AJOL also provides an article download service for researchers to access full text of individual articles.

AJOL hosts over 350 peer-reviewed journals from 27 African countries covering a variety of disciplines including health, education, agriculture, science and technology, the environment, and arts and culture. The number of participating journals and researchers using the service is growing continuously. AJOL hopes to eventually include all quality, peer-reviewed journals on the continent.

Countries with journals on AJOL

The AJOL website receives over 100,000 visits per month from over 190 countries around the world.

Background

The AJOL project was initiated in 1997 by the International Network for the Availability of Scientific Publications (INASP), a charitable organisation based in Oxford, in the United Kingdom. After a positive evaluation of the pilot in early 2000, AJOL was re-launched and expanded. Through INASP, AJOL formed a partnership with the Public Knowledge Project (PKP) relating to the open source software that underpins AJOL's online services. Following the proven need for the AJOL model in developing countries, INASP is currently establishing similar fledgling "JOL"s in Bangladesh, Vietnam and Nepal.

Partner organisations

- Over 350 participating African scholarly journals. List of participating journals
- International Network for the Availability of Scientific Publications; Collaborating with a wide network of partners in sister organisations, development agencies and publishers, INASP has implemented programmes in more than 40 countries worldwide. These programmes are designed for stakeholders engaged in all stages of the research communication cycle, with activities targeted to the needs of researchers, editors, national publishers, and librarians as well as ICT professionals. Following on from successfully initiating and establishing AJOL, INASP has established similar online journal projects in other regions, particularly in South and South East Asia. INASP also runs the Programme for the Enhancement of Research Information (PERI), which provides support to researchers around the world through access to information and training and support for the use of information.
- NISC SA; NISC SA (www.nisc.co.za) is an electronic publishing company specialising in bibliographic database products and African academic literature.
- The Public Knowledge Project; The Public Knowledge Project is a federally funded research initiative at the University of British Columbia and Simon Fraser University on the west coast of Canada. It seeks to improve the scholarly and public quality of academic research through the development of innovative online environments. PKP has developed free, open source software for the management, publishing, and indexing of journals and conferences. Open Journal Systems and Open Conference Systems increase access to knowledge, improve management, and reduce publishing costs. The AJOL database was developed using the open-source journal management software called Open Journal Systems (OJS). Working collaboratively with this organisation, AJOL has been able to create a high quality website with greatly enhanced functionality.
- The Association of African Universities

Donor partners

AJOL is currently supported by the Ford Foundation and through INASP's Programme for the Enhancement of Research Information, by the Royal Danish Ministry of Foreign Affairs (RDMFA), Sida, the UK Department for International Development (DFID) and the Norwegian Agency for Development Cooperation (Norad).

African Journals Online participates in the WorldWideScience global science gateway.

Hosted journals

Active journals

Some of the journals hosted by AJOL are:

- *Acta Theologica* (ISSN 1015-8757))

- *African Environment* (ISSN 0850-8518)
- *African Journal of Cross-Cultural Psychology and Sport Facilitation* (ISSN 1119-7056)

Discontinued journals

AJOL also hosts the archives of several discontinued journals:

- *African Journal of Political Economy*
- *African Journal of Political Science* (1997-2003, ISSN 1027-0353; from 1986-1990 published as *African Journal of Political Economy* ISSN 1017-4974)
- *African Journal of Applied Zoology*, *African Journal of Applied Zoology and Environmental Biology*
- *African Journal of Applied Zoology and Environmental Biology* (1999-2006, ISSN 1119-023X; original title *African Journal of Applied Zoology*)
- *African Studies Monograph* (2001-2007, ISSN 1119-7196

Aluka

Aluka

URL	http://www.aluka.org
Commercial?	not-for-profit
Type of site	Digital library
Owner	Aluka
Created by	Aluka

Aluka is an online digital library focusing on materials about Africa. Aluka's mission is to connect scholars from around the world by building a common platform that allows online collaboration and knowledge sharing. Aluka's audience is higher education and research communities worldwide.

Aluka has been an initiative of Ithaka, which is a non profit organization that has a mission of incubating promising new projects that support the use of technology for the benefit of higher education. An assumption of the incubation process is that successful projects will eventually become independent or join larger, existing organizations serving the academic community. In June 2008, the Ithaka and JSTOR Trustees approved a recommendation that the Aluka initiative be integrated into JSTOR.

Founded in 2003, Aluka was an initiative of Ithaka Harbors, Inc., a non-profit organization based in New York City and Princeton, New Jersey. The initial funding was provided by theMellon Foundation, the William and Flora Hewlett Foundation, and the Stavros S. Niarchos Foundation.

The first release of Aluka took place in early February 2007 with preview access to JSTOR subscribers. In Africa, Aluka is free to all academic and other not-for-profit institutions.

The name 'Aluka' is derived from a Zulu word meaning 'to weave'.

Content

Initial focus of Aluka digital library is in three major areas:

- **African Plants**: Collection of African plants specimens and related materials contributed by the African Plants Initiative.
- **Cultural Heritage**: Collection of images, documents and 3D models documenting African heritage sites, including Timbuktu, Djenne, Lalibela, Kilwa Kisiwani, Lamu, and Elmina. This content area also includes a large collection of African Rock Art from many African nations.
- **Struggles for Freedom**: Documents, images and other materials documenting the liberation struggles in Southern Africa, including those from Angola, Botswana, Mozambique, Namibia, South Africa and Zimbabwe.

Aluka seeks to attract other collections of scholarly interest from institutions and individuals worldwide. By bringing materials together, Aluka creates new opportunities for research and collaboration. Documents and materials that were previously hard or impossible to access are now available for researchers around the world.

California Digital Library

The **California Digital Library,** or CDL, is the University of California's 11th University Library. The CDL was founded to assist the ten University of California libraries in sharing their resources and holdings more effectively, in part through negotiating and acquiring consortial licenses on behalf of the entire University of California libraries system. Its current mission is to support the assembly and creative use of the world's scholarship and knowledge for the University of California libraries and the communities they serve.

Among its programs and services are the Online Archive of California, Calisphere, Counting California, the union catalog of the UC libraries, Melvyl, and the eScholarship Publishing Program, which provides open-access and alternative publication services to the University of California.

Central and Eastern European Online Library

The **Central and Eastern European Online Library (C.E.E.O.L.)** is an online archive providing access to full text articles from humanities and social science scholarly journals on Central, Eastern and South Eastern European topics.

The subject areas include: anthropology, culture and society, economy, gender studies, history, Judaic studies, fine arts, literature, linguistics, political sciences and social sciences, philosophy, religion, reviews, etc.

C.E.E.O.L. is initiated and maintained by Questa.Soft GmbH in Frankfurt am Main, which is a software development company, founded in 1998 by two partners, Wolfgang Klotz and Aurelian Urzica. From the very beginning, the company started to build up the technical and logistical basis for the future operation of C.E.E.O.L. In 2000 the company completed the development of a content management pilot project and offered publishers from Central and Eastern Europe a platform for the online distribution of their publications—C.E.E.O.L.

C.E.E.O.L. may be seen as an heir of the East/West European Cultural Centre “Palais Jalta” in Frankfurt am Main which, for the twelve year period from 1991 to 2003, attempted primarily to balance the flow of information from West to East with a significant counter current in the opposite direction. C.E.E.O.L., as the “virtual successor” of “Palais Jalta”, shares this aim and continues this work on a global scale via the Internet.

As a logical enhancement of C.E.E.O.L. and following numerous requests from C.E.E.O.L.'s partner publishers, Questa.Soft GmbH decided to provide the Eastern European publishers with an additional online platform DiBiDo dedicated to e-books from and about Central, Eastern and South Eastern Europe, covering the same subject, language and geographic areas as the periodical database. The DiBiDo platform allows the participating publishers to distribute their books in digital format, to readers from across the world.

Digital Himalaya

The **Digital Himalaya** project was conceived of by Professor Alan Macfarlane and Dr Mark Turin as a strategy for archiving and making available valuable ethnographic materials from the Himalayan region. Based at the Department of Social Anthropology at the University of Cambridge, the project was established in December 2000. From 2002 to 2005, the project moved to the Department of Anthropology at Cornell University and began its collaboration with the University of Virginia. As of 2009, Digital Himalaya is back in Cambridge.

Primary Objectives

When established in 2000, Digital Himalaya project had three primary objectives:

- To preserve in a digital medium archival anthropological materials from the Himalayan region that are quickly degenerating in their current forms, including films in various formats, still photographs, sound recordings, field notes, maps and rare journals.

- To make these resources available over broadband internet connections, coupled with an accurate search and retrieval system useful to contemporary researchers and students.

- To make these resources available on DVD to the descendants of the people from whom the materials were collected by making them both easily transportable and viewable in a digital medium.

First Phase

Five ethnographic collections representing a broad range of regions, ethnic groups, time periods, and themes were selected for digitisation in the first phase of the project, along with a set of maps of Nepal and important journals on Himalayan studies.

Domínio Público

Domínio Público is a Digital library created by the Brazilian government, under the *Secretaria de Educação à Distância do Ministério da Educação* (the Secretariat for Distance Education of the Ministry of Education), with the goal of harnessing the diffusion of cultural works under public domain. It contains more than 10,000 works in text format and another 4,000 in other formats (music, video, images etc.), the majority in Portuguese. literary works are in PDF format, and include contributions from different Brazilian universities (and their respective virtual libraries), international organisms as UNESCO, and the work of volunteers and similar organizations (it contains many works in English contributed from Project Gutenberg, for example).

Although it focuses on works by Brazilian authors and in Portuguese, it accepts collaborations in all languages, provided that they are in the public domain. In order to facilitate the work of volunteers and prospective contributors, the *Domínio Público* web site maintains a list of Brazilian authors with works under public domain, prepared by the National Library of Brazil.

European NAvigator

European NAvigator

ENA on 2 March 2010

URL	http://www.ena.lu/
Commercial?	No
Type of site	Multimedia encyclopedia
Created by	CVCE

European NAvigator (**ENA**) is a web-site about the history of European integration and related institutions since 1945 focusing on the development of a united Europe.

Using the site is free, although the documents are protected by copyright.

ENA is developed by the CVCE (Centre Virtuel de la Connaissance sur l'Europe - Virtual Resource Centre for Knowledge about Europe), a Luxembourg-based public undertaking that is actively supported by the Ministry of Culture, Higher Education and Research.

ENA is available in English, French, German and Spanish, though some documents are available in other languages.

Multimedia resources

ENA is a large multimedia knowledge base:

- original texts (treaties, etc.)
- video clips
- audio clips
- press articles
- photos
- interactive maps
- cartoons
- tables

Parts

The content of the ENA is divided into five parts:

- 'Historical Events' contains material on all the events that have contributed to the European integration process;
- 'European Organizations' looks at the operation of all the institutions of the European Union (e.g. European Parliament, European Commission) and the various other European institutions;
- 'Special Files' are devoted to specific subjects;
- 'Interviews' contains exclusive interviews with people who have played a part in the European integration process (Jacques Santer, Otto von Habsburg, etc.);
- 'ENA & Education' provides resources for teachers to enable their pupils to learn about European integration.

Thesaurus

A thesaurus organises the material according to subject area.

Glossary

A glossary explains terms relating to the European institutions and the history of Europe (e.g. ECSC, Marshall Plan,...) that appear in ENA.

Everglades Digital Library

The **Everglades Digital Library** is hosted and supported by the Florida International University Libraries, in collaboration with Everglades National Park, the University of Florida Libraries, and numerous other agencies and research organizations. The Everglades Digital Library is a library with multiple large and growing collections that regularly add new materials, including scientific and technical reports, natural history writings, educational resources, maps, photographs, and additional contextual materials on and relating to the greater Everglades.

Florida Digital Newspaper Library

The **Florida Digital Newspaper Library** provides access to the news and history of Florida through local Florida newspapers. The Florida Digital Newspaper Library is supported by the University of Florida's George A. Smathers Libraries and hosted in the University of Florida Digital Collections funded partially by grants and sources, including Florida's Library Services and Technology Act (LSTA) Grants Program, the National Endowment for the Humanities' National Digital Newspaper Program, the Institute for Museum and Library Services, the University of Florida, by Florida Heritage Project funds from the University of North Florida and the [University of South Florida], and with the assistance of digital library endowment from the Estate of the late Governor and Mrs. C. Farris Bryant (whose papers are within the Bryant Collection).

In addition to multiple funding sources, these newspapers are indexed and included in multiple collections, including Chronicling America, and being indexed by the Florida Electronic Library and Google News Archive.

Collection Themes and Titles

The **Florida Digital Newspaper Library** includes newspapers by Cuban exiles, like *El Avance Criollo* which was published in Cuba and then continued in Miami. Other titles include:

- *Citrus County Chronicle* (Inverness and Crystal River, Fla.) 1890-current (LCCN sn87070035): The Citrus County (FL) Chronicle was begun by Albert M. Williamson in July 1890 as a means of announcing goods for sale to people in the Inverness (FL) and Floral City (FL) areas to the north of Tampa (FL). Walter Warnock, the county clerk in Inverness, assumed ownership of the paper in the late 1890s and added news items and features. George Butler, then 70 years old, became owner and editor of the Citrus County Chronicle in 1914, and later that year Albert W. Butler became editor and owner. Joseph J. Wilson, of Clearwater, took over the paper in April 1929 and was an activist editor promoting goals to return the area to prosperity.
  Taylor Dawson became the editor in 1935, the same year the Scofield Publishing Company, owner of the Citrus County Chronicle, also acquired the Dunnellon (FL) Sun (LCCN: sn98026435). Scofield sold the paper in May 1946 to N.A. Perry of Bradenton, who sold shortly afterward to J.R. Hough. A year later, Col. George H. Johnson bought both the Citrus County Chronicle and the Dunnellon Sun from Hough. Then in 1948, the Citrus County Chronicle was sold to Paul W. Ramsey who had been the city editor at the Chicago Sun (LCCN: sn87082566). Ramsey sold the paper to the Bennett-Hahn Company in June 1959, which then sold it to Frances and Carl turner of Wisconsin.
  At Turner's death in 1962, his wife sold the paper to a St. Petersburg (FL) group that included former St. Petersburg mayor Herman Goldner. It was then bought by David S. Arthurs and in September 1980 merged with Landmark Community Newspapers, Inc. The Citrus County Chronicle became a daily in 1986, serving all of Citrus County. The newspaper is currently (ca. 2007) owned by Landmark Communications, Inc. In July 1990, the paper moved from Inverness to Crystal River. Citrus Publishing also publishes weekly editions of Sumter County Times (LCCN: sn95072059), Riverland News (LCCN: sn96027433), South Marion Citizen, the Visitor, Chiefland Citizen (LCCN: sn96027361) and Williston Sun News Pioneer.

Hungarian Electronic Library

The **Hungarian Electronic Library** (Hungarian: *Magyar Elektronikus Könyvtár*) is one of the most significant text-archives of the Hungarian Web space showcasing a variety of primary and secondary sources. Contains thousands of full-text works in the humanities and social sciences. Topics covered include science, math, technology, arts, and literature. Most texts are in Hungarian, though some have been translated into English.

Kujawsko-Pomorska Digital Library

Within the framework of Scientific Libraries Consortium of Kujawsko-Pomorski Region, Nicolaus Copernicus University Library in Toruń has started a long-term enterprise of building a digital library called **Kujawsko-Pomorska Digital Library**. The project implementation was financed by EU Structural Funds and first collections are to be created in the years 2005-2006. At the end of 2006 the collections were accessed.

The main aim of the project is to create a regional digital library to support the development of an intellectual and innovative potential of the society, to make a quick access to information and knowledge content possible, and to protect valuable documents of the region and national literature works. The project was innovative and experimental, as concerns Polish libraries (one of the first ones in Poland).

Project participants in expenses and its implementation:

- Nicolaus Copernicus University Library in Toruń – the coordinator of the project;
- NCU Collegium Medicum Library in Bydgoszcz - partner;
- Kazimierz Wielki University Library in Bydgoszcz – partner.

As well as remaining libraries of Scientific Libraries Consortium of Kujawsko-Pomorski Region, which will use and develop the project.

Following groups of collections:

- Science Collection – including digital copies of selected handbooks, monographs and scientific articles from the region.
- Cultural Heritage Collection – including digital copies of the most valuable and used items: incunabula, old prints, manuscripts, iconographical collections, cartographical items, emigrational collections, etc.
- Regional Collection – collecting digital copies of: regional publications, leaflets, posters, playbills, photographs, invitations, exhibition catalogues and trade fairs of the region.
- Music Collection – digital copies of scores.
- Maps Collection – digital copies of selected maps of the region.

Kujawsko-Pomorska Digital Library is to serve scientists, students, pupils and all the citizens of the region. Its implementation will depend on all librarians engaged in the project and the authors, who would want to place their works in a digital library. In the future it will depend also on institutions, which would wish to enlarge its holdings of other collections important for the region. Nicolaus Copernicus University Library is opened for each good proposal towards enriching digital base of this digital library.

Panjab Digital Library

Panjab Digital Library	
Revealing the invisible heritage of Panjab	
Country	Panjab
Type	Digital library
Established	2009
Location	SAS Nagar, Chandigarh
Collection	
Items collected	manuscripts, book, photographs, newspapers, magazines, Sound recordings
Size	10,000 books/manuscripts
Access and use	
Access requirements	Open to anyone with a genuine need to use the collection
Other information	
Director	Davinder Pal Singh
Website	http://www.panjabdigilib.org/

The **Panjab Digital Library** is a NGO that is digitizing and preserving cultural heritage of Panjab. There are many historically significant documents stored and made available online. Its scope covers Sikh and Punjabi culture. The library funded by The Nanakshahi Trust was finally launched online in August 2009. It is located at Chandigarh.

Coverage

Scanned collections include manuscripts held by the Panjab Languages Department, items from the Government Museum and Art Gallery Chandigarh, Chief Khalsa Diwan, SGPC, DSGMC and manuscripts in the Jawahr Lal Nehru Library of Kurukshetra University.

Digital Library of India

Digital Library of India, part of the online services of the Indian Institute of Science, Bangalore and partner in the Million Book Project, provides free access to many books in English and Indian languages. The scanning of Indian language books has created an opportunity for developing Indian language optical character recognition (OCR) software. The publications are mainly in PDF or QuickTime format.

Because of copyright laws, the texts are all out of copyright and therefore not sources for current information, but rather useful for history and background.

As of November 10, 2006, DLI had scanned 84,895 titles.

Representative titles include:

- *Ancient India*, McCrindle J. W.. 1885.
- *Ancient Indian Polity*, Aiyangar K. V. Rangaswami. 1935.
- *History Of The Parsis Vol-I*, Karaka Dosabhai Framji. 1884.
- *A Treatise On Kala-Azar*, Brahmachari Upendranath. 1928.

Digital Library of Slovenia

The Digital Library of Slovenia is a web portal providing ready access to digitised knowledge and cultural treasures. It offers a free search through sources and free access to digitised contents, such as periodicals, books, manuscripts, map, photographs, music and manuals.

History

The Digital Library of Slovenia is a part of The National and University Library of Slovenia (NUK). Strategic origin of the project can be found in The Strategy of the Republic of Slovenia in the Information Society si2010, the resolution on the National Programme for Culture to cover the 2008-11 period and many other European Union documents and initiatives which define the meaning and guidelines of further development towards an open, developed and, first of all, a universally accessible information society.

As a response to these challenges, the NUK started to develop a web portal that would guarantee an equal and free of charge access to relevant information, irrespective of user's location. The portal would be also a support to learning, scientific and research work, cultural development, progress of national identity and also to a virtual

environment. The Digital Library of Slovenia was presented to public for the first time in November 2005 and is also a part of The European Library, a joint project of European national libraries.

Technical aspects

The Digital Library of Slovenia portal acquires publications metadata directly by using Z39.50 protocol which provides access to a library catalogue, or the information is a result of a digitisation of materials. Basis for metadata description is the Dublin Core (DC) standard which suggests enlistment of 15 basic elements for describing digital objects that may be extended at will by using expandable DC scheme.

Data are taken through the HTTP protocol by placing demand on OAI-PMH server that enables influx and transmission of data from practically all existing metadata schemes. Server provides a result in one of the XML format, defined in advance. As data collections can be huge, because transferred through the web, and formation and transfer of huge XML files might take too long, transfer of collection is split in optional number of series.

National libraries, archives, depositary units and other owners of electronic contents who receive data are often faced with question referring to permanent tracking of published electronic documents. Accessibility of articles (as for instance as basis of reference) should be ensured irrespective of the fact that Uniform Resource Identifier (URI) of a document can be changed (even often). One of the most prevalent international approaches enabling tracking of digital contents irrespective of variability of their web location is use of DOI (Digital Object Identifier) service. Other solution that can be implemented on local level is called URN (Uniform Resource Name).

While Digital Object Identifier is administered by the International DOI Foundation (IDF) organisation and the usage of a DOI identifier is a payable service, the URN is a solution that can be developed and used free of charge. National and University Library provided realisation of the Digital Library of Slovenia. Principle of the operation and usage of a URN reference number to a great extent resembles to the DOI indicator.There are two major system functions: the URN generator – service defining unique URN record for every object and the URN resolver – interpreter that connects a given URN record with a digital object.

Collections

The Digital Library of Slovenia offers access to publications, published on the web and to digitised contents, published on classic carriers. Through a single access and a cross-search feature it unifies collections of a different kind in one place. Searching in collections is possible according to bibliographic data (author, title, publishing date, content, type of material …). Full-text retrieval technology, supported by OCR, is used to allow you to search directly in the text.

Currently available collections:

- ***Scientific and scholarly articles***: there are more than 3,000 articles published in many internationally recognised professional and scientific journals.
- ***Articles***: the collection of articles provides tools for browsing of older newspapers that gave character of the period (Kmetijske in rokodelske novice, Ljubljanski zvon, Dom in svet, Novi akordi, Sodobnost, Štajerc, Nova muzika).
- ***Photographs***: photographs of the poets Anton Aškerc, Karel Destovnik Kajuh, the bishop Anton Jeglič, senators of the Drava Banovina and many others; there are also postcards with Jurij Vega, France Prešeren, and old Ljubljana; caricatures by Maksim Gaspari and Hinko Smrekar; reproductions of Ivan Cankar's drawings and manuscripts…together there are more than 10,500 items.
- ***Books***: the database includes books from 1830 until today. It provides free access to some of the most important works of the Slovenian authors, such as Ivan Cankar's Erotika, Nina, Hlapci; Dragotin Kette's Poezije; Josip Murn's Pesmi in romance. For the admirers of poetic realism, the collection includes Janko Kersnik and his work Kmetske slike. A treasure of special kind is the first translation of the libretto of Richard Wagner's opera Lohengrin, published as a separate edition.
- ***Posters***: the collection includes advertising, promotional, film and war posters and also political posters by well-known authors, such as Maksim Gaspari and Ive Šubic. All together, there are almost 4,000 posters available.
- ***Sheet music***: in more than 1,000 records, there are works of the Ipavec family, Danilo Fajgelj, Risto Savin, Stanko Premrl, Emil Adamič, Slavko Osterc, Marij Kogoj. The arrangements of the Slovenian folk songs and the songs from the turning points in the Slovene history and one of the first arrangements of Prešeren's Zdravica (Slovenian national anthem) are also included in this database.
- ***Higher education publications***: offers free access to complete scientific research works of various domains, ranging from natural, social to library sciences.
- ***Maps***: the selection of old maps provides access to the famous Peter Kozler's map *Zemljovid slovenske dežele in pokrajin*, Valvasor's maps of Carniola, and all the others, including Sebastian Münster, who pictured the Slovenian ethnical territory as it was in the 16th century. In addition, the Maps database also includes old plans of the capital of Ljubljana (from 1820 to 1920).
- ***Sound recordings***: it consists of the recordings of solo singers and ensembles dating from the beginning of the 20th century, including a wide range of folk songs.

Georgia Library Learning Online

Georgia Library Learning Online, more commonly known as **GALILEO**, is a virtual library operated by the University System of Georgia. There are numerous databases available, including abstracts and full-text. The Digital Library of Georgia is also part of the GALILEO system.

The full system is only available to those with a password, or those accessing it from computers with authorized IP address ranges. These include computer networks within the university system (USG) and The Georgia Department of Technical and Adult Education (DTAE), as well as public schools and grade schools within the state. Public libraries and some private schools also have access. Passwords change four times each year and are different for each institution.

Homeland Security Digital Library

The **Homeland Security Digital Library** (HSDL) is the nation's premier collection of documents related to homeland security policy, strategy, and organizational management. Library access is offered to U.S. citizens who are federal, state, tribal, and local U.S. government officials; members of the U.S. military; homeland security researchers and academics; or security staff protecting organizations vital to U.S. infrastructure. The mission of the Homeland Security Digital Library (HSDL) is to strengthen the national security of the United States by supporting federal, state, local, and tribal analysis, debate, and decision-making needs and to assist academics of all disciplines in homeland defense and security related research.

History

Following the events of September 11, 2001 and the government response to address the security of the country, there was a recognized need for a formal higher education program that would discuss and debate strategic solutions for the new critical issues and threats facing our country. In response, the Naval Postgraduate School (NPS), with the sponsorship of the Department of Justice, launched a graduate degree program in 2002. The program's initial requirements called for development of the Homeland Security Digital Library (HSDL), an online resource designed to support NPS' new homeland security master's students, and to capture and archive the current homeland security debate. Since then, the HSDL has become the major digital library for homeland security scholars and professionals; it serves almost 15,000 individual account holders and gives general access to over 500 different government agencies and educational or research institutions. The library is currently in partnership with the Government Printing Office and the I3P Institute. The HSDL can be easily summed up by its tagline, "Securing the Homeland Through the Power of Information."

Content

General Collection The HSDL's General Collection contains over 65,000 documents covering a range of historical and contemporary issues in homeland security and its related fields. The collection includes open source material from a wide variety of sources including federal, state and local government; international governments and institutions; nonprofit organizations; and private entities. Special collections include a large number of homeland security related Congressional Research Reports (CRS), Government Accountability Office (GAO) hearings and testimony, national strategy and policy documents, and state and local government plans and policies. In addition to text-based material, the digital collection contains other media formats such as maps, images, video, audio and websites. Within the General Collection, HSDL content team members have designated special smaller collections of key documents that provide easy, direct access to major legislation, significant policy, presidential directives, executive orders Homeland Security theses and the latest documents of topical interest or potential importance.

Restricted Collection The Restricted Collection contains sensitive materials collected from national, state and local fusion centers, threat analysis centers, and law enforcement organizations. (Access to this section is restricted to U.S. government officials and requires prior approval by HSDL staff.)

News Digest Collection The News Digest Collection includes documents, both current and archival, covering intelligence, infrastructure, terrorism and other homeland security concerns.

I3P Cyber Infrastructure Collection The cyber infrastructure collection is a partnership of the Institute for Information Infrastructure Protection (I3P) and the HSDL. It is focused on identification of information assets in the broad area of information infrastructure protection and cyber security. The I3P's effort in this partnership was sponsored by the National Institute of Standards and Technology (NIST).

Public Resources

HSDL Blog: *On the Homefront* – The HSDL blog is an open listing of homeland security related documents that are rapidly posted as they become available to the public. Each blog entry includes a link and a short summary plus any additional helpful information. Homeland security professionals can also find listings of current homeland security events (e.g., conferences and workshops). Blog posts are available via RSS and on Twitter. Other public resources on the HSDL site include a customized search of selected homeland security blogs, and resource pages for grants, homeland security books and journal sources.

Technology

The HSDL technical infrastructure relies on a number of open-source web applications. The library's interface to its patrons is custom-developed web code built upon the SOLR

search engine. The library content team currently uses Scout Portal Toolkit (SPT) to input and manage the collections' metadata. On the Homefront, the library's blog and the events and conferences calendar use the Drupal CMS). These three applications, in addition to the custom code developed around them, are all open-source LAMP environments (linux, apache, mysql and php). The Homeland Security Blog search is maintained with a Google Custom Search appliance.