ISSN (Print): 0972-2467
ISSN (Online): 0976-2477
SRELS Journal of Information Management, Vol 56(4), August 2019, p.159-170 DOI: 10.17821/srels/2019/v56i4/146594

Digital Libraries: Mission Accomplished?

David Bainbridge*

Department of Computer Science, University of Waikato Hamilton, New Zealand; davidb@waikato.ac.nz

Abstract

This article is the first in a series of three publications that reflects on the boundaries and perceptions of what we think of when we hear the term Digital Libraries, as substantiated through our mainstream digital library systems. In this first article we lay the groundwork, concentrating on detailing the capabilities of established digital libraries, as a point of reference. We present a range of examples and highlight the commonalities that occur despite dealing with diverse topics and subject matter. We then go on to detail generalized forms of digital library that occur and discuss the software architecture behind them.

Keywords: DL Architecture, Digital Libraries, Digital Library Software

1. Introduction

A quarter of a century on and the fledgling idea of Digital Libraries (DLs) have grown to become a mainstay for how we access information online—both for work and leisure. Today, it is easy to access a mind-boggling array of authoritative resources. For instance:

- A farmer, concerned about a problem occurring in one of their crops, can consult a compiled repository of technical reports for potential sources of the problem, along with details of countermeasures to take.
- A scholar of antiquity, seeking to clarify an issue that has arisen in an interpretation of a translated text, can quickly track down the original, and compare it with the translated version in question.
- A genealogist, from the comfort of the own living room, can scour historic court records looking for traces of the ancestors being researched.
- Literature buffs, preparing for a trip to a city they have often read about in their favourite books series but never visited, locate relevant audio books and transfer them to their tablet for listening to as they walk around the city, sightseeing.

And so, the examples could go on. But cast the eye of a digital library researcher over these online resources and it reveals that the technical capabilities of these systems have hardly changed at all from their predecessors developed some 20 years earlier. Are we really saying that digital library software has peaked, and achieves everything it needs to? Is it really a case of Digital Libraries: Mission Accomplished?

This is *not* the point of view taken here quite the opposite. In fact, it has been a concern about this status quo—one might even say quiescence—that has provided the impetus for writing this article, and the two that follow on from this. Collectively, these articles seek to raise awareness about the situation and, furthermore, to give examples that demonstrate practical ways in which digital library software can deliver more than what is typically provided. In doing so it is hoped that these articles help alter people's perception of what digital libraries can be, which in turn motivates a newer breed of DL software architecture that goes beyond the boundaries that seem to have established themselves.

This first article lays the groundwork and concentrates on detailing established digital library capabilities as a point of reference. We do this by showcasing four real-world different digital libraries that deliver on the scenarios this article started with. While there are many differences to do with the subject matter to which they each relate, put together in this sequence it is the uniformity of how the systems operate that stands out. We go on to frame these examples in the context of a widely used definition for what constitutes a digital library, and further, discuss more general forms of DL with respect to this definition. We

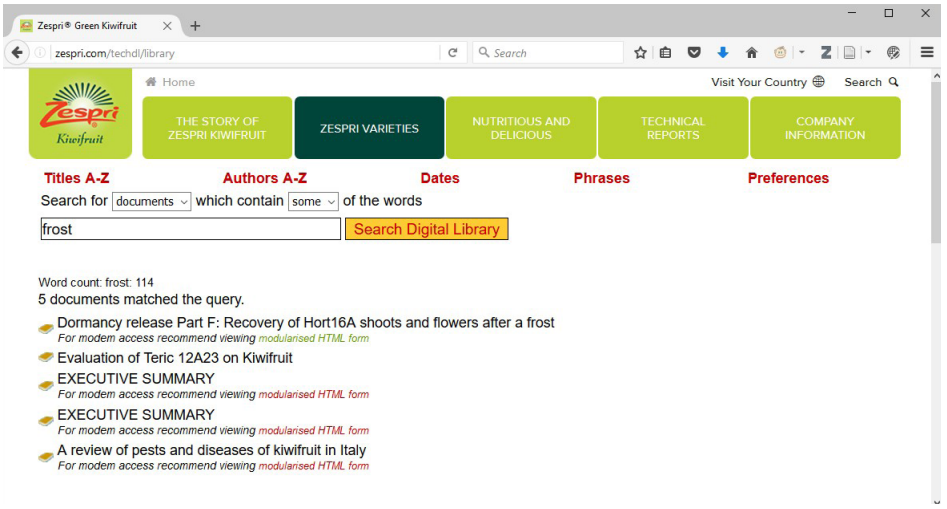
*Author for correspondence

round out the discussion in this article by describing how this has manifested itself in term of software architecture for digital libraries.

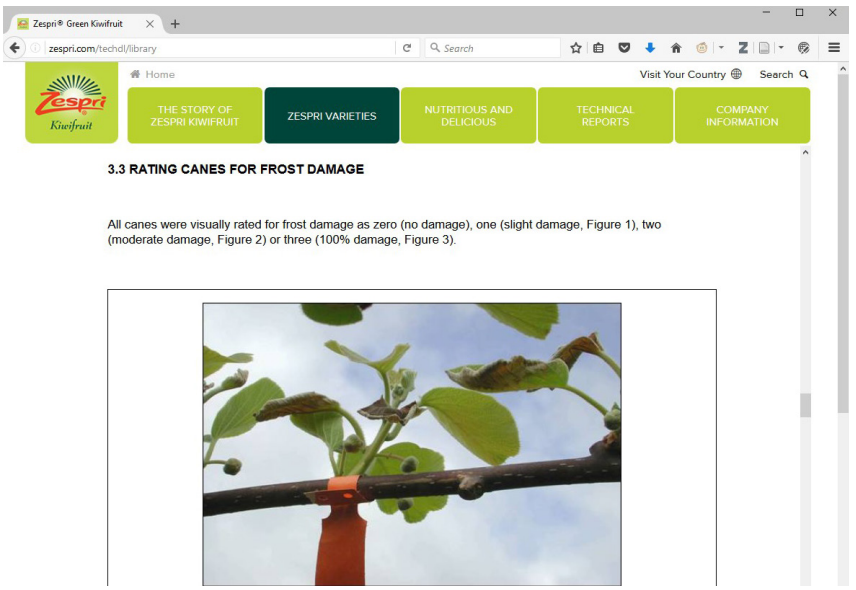
The second article, *Disassembling the Software Architecture of Digital Libraries: Getting More out of the Building Blocks* will review DL interoperability as its topic and focus on the *de facto* Digital Library software architecture, and show that it can in fact be more versatile than how it is typically deployed. It just takes a bit of additional thinking—and will—to make it so. In the third article, *Redefining Digital Library Boundaries: Being More than an Information Silo* we take music digital libraries as a motivating example to show the sorts of features a DL can have that go significantly beyond the status quo this first article shows, yet still turn out to fit within that existing, widely used definition of a digital library.

2. The Established World of Digital Libraries

Through Figures 1–4 we elaborate on the specific scenarios given above, drawing upon realworld Digital Libraries to give more detail.



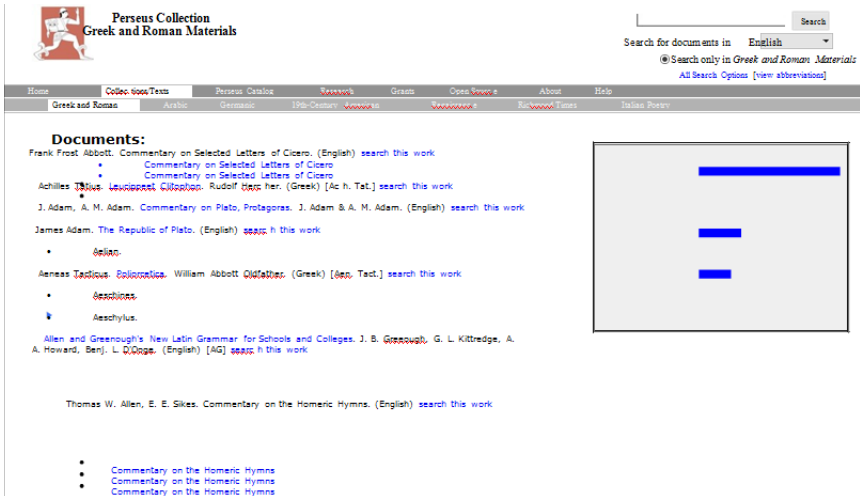
(a)



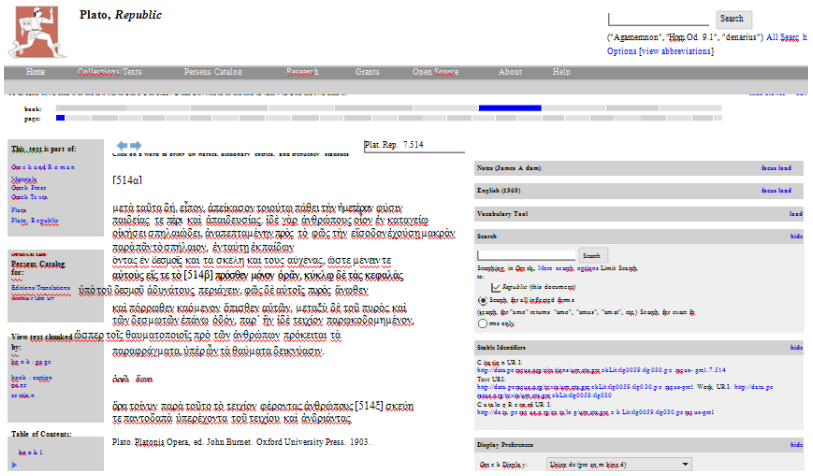
(b)

Figure 1. Locating information about frost damage in kiwifruit: **(a)** the result of searching for the term 'frost'; and **(b)** viewing one of the matching documents.

In the context of farming, consider the commercial growing of kiwifruit in New Zealand, where a farmer has noticed dried and browning leaves on the vines, which they suspect is related to a recent series of frosts. Still in the kiwifruit orchard and using their smart phone, they access the technical reports digital library provided by Zespri, the country's government-formed company for the growing and export of the produce. In Figure 1(a) the results are shown from a full-text search for the term 'frost' that the farmer has initiated.



(a)



(b)

Figure 2. Checking an interpretation of Plato's *Allegory of the Cave*: **(a)** locating the relevant text by author listing in the Perseus Digital Library; and **(b)** viewing the original Greek text.

Focusing in on the top hit, *Recovery of Hort16A Shoots and Flowers after a Frost*, they then study the contents of this document. Figure 1(b) is a snapshot taken from roughly halfway through the document, where photos of frost damage at various stages is shown. From here they can determine whether the damage they are seeing matches with the report's details, and if so, what to do next.

Our second example demonstrates a digital library used in an academic setting. Taking Græco-Roman culture as the focus, suppose that the body of text in dispute is the English version of Plato's *The Allegory of the Cave* translated by Paul Shorey (Cambridge, MA, Harvard University Press, 1969). Figure 2(a) shows our scholar accessing the Perseus Digital Library—a prominent source of on-line resources in the humanities, in particular classical studies—to locate the original Greek text. They have accessed the Greek and Roman.

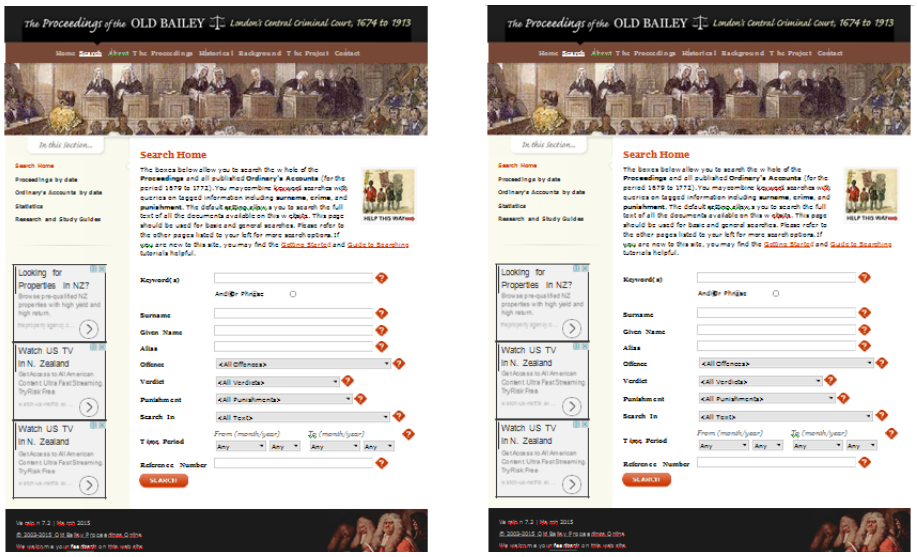


Figure 3. Looking for traces of ancestors in the Old Bailey Online digital library: **(a)** using the advanced search form; and **(b)** viewing a resulting document.

Collection within the digital library, and then used the alphabetical listing of authors to first access texts by Plato, and then, further, opened up the Greek version of *Republic*, the book containing the sought-after section. In Figure 2(b) the scholar has used the '*with-in book*' navigation features of the library to access Section 514, which is the starting point for *The Allegory of the Cave*, to commence their checking. If we were to extend the example further, the Perseus Digital Library even has a copy of the 1969 translation in it, and so the next logical step for the scholar would be to open up this version too—say by using the DLs full-text search feature—and position the new window side-by-side with the old to aid the comparison.

Moving on to our genealogist example, the Old Bailey Online is a digital library of the proceedings of court cases, spanning over 230 years (1674–1913), conducted at England's Central Criminal Court. It's fully searchable 190,000+ trial records is an exceptionally useful resource in which to locate authoritative information about people from this period of history who would otherwise not normally be officially documented, notwithstanding births, deaths and marriage records. For this example, we join an ancestral sleuth on the trail of a relative they have recently learnt about, Mr. L. Lyons, who by 1841 was living in Tasmania, Australia. This information came from sighting the marriage registry entry, but frustratingly the handwriting for his first name was hard to reliably decipher ('L????s'). Furthermore, they have hit a dead-end with trying to locate information in Australia prior to this year. Given Tasmania's reputation as a convict settlement, following a hunch, they visit the Old Bailey Online. They access the advanced search page, shown in Figure 3(a), and use the fielded form to enter the surname, and restrict the date range to 1820–1840.

Three matching documents are returned from this query (not shown). Accessing the trial details in turn, they find the last of these entries, shown in Figure 3(b), correlates well with known details about their relative.

His full name was Lewis Lyons—thankfully the trial records were typed—lived in London, and on the 7th April 1831 at the age of 23, he was found guilty of simple larceny with the sentence "Transported for Seven Years." The crime was theft of a 21 lb drum of cheese, worth 10 shillings, from a cart. Three witnesses spoke at the trial, one of them a police constable. He had seen 4–5 men lingering near the cart. He saw one of them, not Lyons, "go behind the waggon [sic], take something out, and give it to the prisoner." Lyons spoke in his own defence: "I was coming along; a man stopped me, and asked me to be so kind as to hold the cheese while he laced his boots – I held it for him about three minutes, when the officer took me." While in today's terms it is hard to comprehend the severity of the sentence, with what we now know about the widespread levels of poverty in Victorian London, having survived the journey to Tasmania and period of imprisonment Lewis Lyons would generally be considered as having had a better quality of life overall than if he had remained living where he was.

In the last of our opening examples we make audio the focus, to demonstrate that digital libraries are not just about locating and accessing text documents, and we do so using Literary Tourism as the motivator. This is a recreational pastime whereby people visit geographical locations related to works of fiction they have read. While not possible for all books, due to writing style decisions the author has made, prominent examples of Literary Tourism also tend to tally with being well established classics: James Joyce's novel *Ulysses* set in Dublin; the works of Dickens that are set in London, and so on. While the pastime predates digital libraries, such software is certainly something that can support someone in this pursuit.

In our illustrative scenario, we have a literary tourist who is a fan of Jane Austen, and has an up-coming trip to Bath, England, which is an historic spa town that Austen both lived in and features in some of her novels. For travelling to the town our fan has decided that a pleasant activity would be listening to audio book versions of *Persuasion* and *Northanger Abbey*. A quick bit of web searching helps them identify the LibriVox DL collection, hosted by the Internet Archive, as a fruitful source for what they are looking for. Figure 4(a) shows the collection's home page, where the tick-box filters down the left-hand side have been used to restrict the items shown to those authored by Jane Austen. Using their tablet, in Figure 4(b) our literary tourist has clicked on the first matching item for *Persuasion* that came up, which happens to be labelled Version 5, as there are in fact many contributed readings of the book in the collection. From here they can play the audio book, *in situ* in the web page, or else choose to download the audio file, which is available in various audio formats through links lower down the page. They choose to first listen to part of the audio in the browser to determine the quality of recording, and satisfied with what they hear, they then download it to their tablet, and return to the home page to locate the second of the Austen books they are interested in, *Northanger Abbey*.

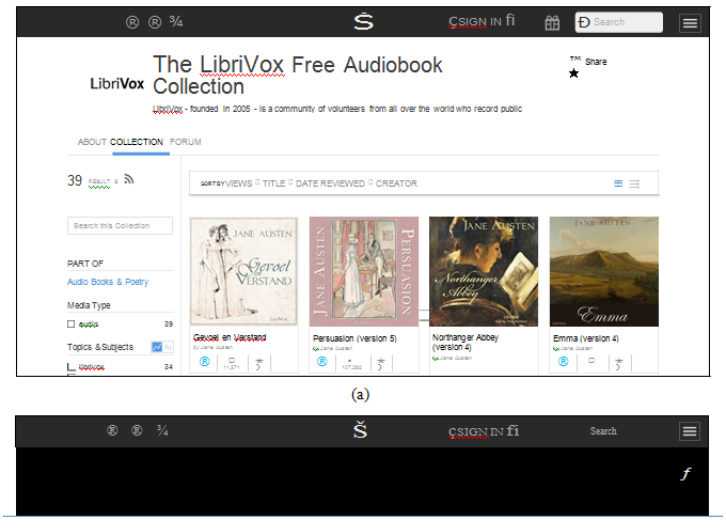
Extrapolating further along this trajectory, we would most likely find our traveller choosing to continue using the audio recordings once they reached the town of Bath and started their sightseeing; we could even imagine them savouring listening to the books once more on the return journey, or else topping up their tablet with pertinent additional audio books to listen to, now they have a clearer picture in their mind as to how the town looks.

3. Digital Library Definition

Despite the opening examples being intentionally diverse in the domains they draw upon, there is clearly much commonality in how they function. This is more formally brought out through the following widely used definition of a digital library, which is:

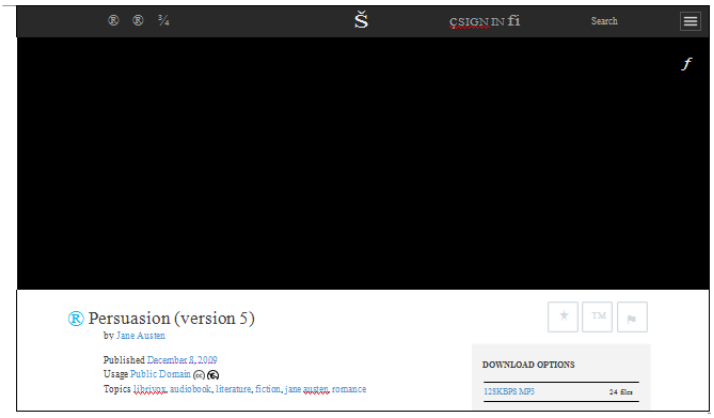
> *a focused collection of digital objects—including text, video, and audio—along with methods for access and retrieval, and for selection, organization, and maintenance of the collection.*

The given examples concentrate on how end-users interact with the online resources, which corresponds to the first part of the definition: access to and retrieval of a collection of digital objects. (We deal with the latter).



(a)

(a)



(b)

Figure 4. Locating novels by Jane Austen in audio book format within the Internet Archive's LibriVox collection: **(a)** filtering audio books to only those authored by Jane Austen; and **(b)** viewing the audio book version of *Persuasion*.

Part of the definition broadly classified as management, in the second article in this series.) The initial point of contact for the user with a collection is usually its home page. This is a good place to present details that help the user understand what the collection contains: in library science parlance, its policy statement. Also, unlike its physical counterpart which can be visually inspected to gauge how comprehensive it is, displaying how many entries the collection contains further helps a user determine the level of coverage the collection provides. Home pages exist in all four of our examples, however for brevity we only chose to show the ones provided by the LibriVox audio books and Perseus Græco-Roman collections.

Searching and browsing are the key activities that allow our users to access and retrieve items in a collection. In the cases of the kiwifruit farmer, and the genealogist we saw examples of searching. In the former, the digital library provides a quick search box available in the header of all pages in the collection that performs a general full-text query. The farmer used this to locate documents that contained the word 'frost'. Given that the collection is exclusively about kiwifruit, no other query terms were needed. In the latter, the Old Bailey Online digital library's advance search page was used to frame and submit a more precise query to pinpoint potentially relevant documents. The LibriVox and Perseus Digital Libraries also support searching it is just that this capability was not needed in the chosen scenarios.

Alphabetized lists based on metadata about the documents in a collection—such as ordering by document title, author, subject or keywords—is the most commonly encountered form of browsing. We see this in the Græco-Roman collection where the list of works ordered by author is utilized by our scholar to access *The Republic*. Again, the other digital libraries also provide this.

The filtering activity we see our literary tourist using to restrict items shown to only those authored by Jane Austen is a hybrid of browsing and searching, known variously as "faceted browsing" or "faceted search". Even though it logically has the same effect as using an advanced search form where *Jane Austen* is entered as the search term for Author, a key difference in the faceted approach is that valid items are displayed to the user up front, rather than the user having to figuratively pluck items from thin air and hope they correspond to an item in the index. The faceted approach, therefore, is advantageous in that it can shorten the time taken to find items, and reduces potential user frustration from the alternative advanced-search approach when queries return no matches; the flip side to this is that, for the technique to work well, the total number of possible items needs to remain small. Displaying the author facets on the home page to the Library of Congress' on-line catalogue would not work well, for example!

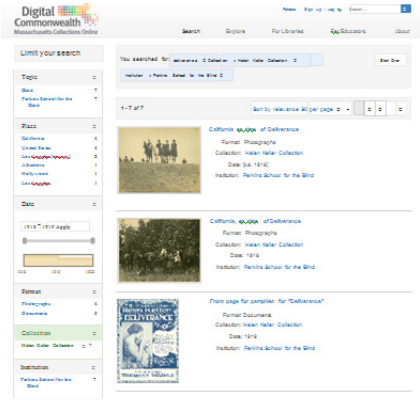
4. Forms of Digital Library

When working with certain types of content, distinctive forms of digital library occur— subclasses of DL, if you will - all of which still adhere to the given definition. In this section we present some additional examples of digital libraries. This time, rather than being motiving examples, they are chosen as exemplars of the DL form they represent. We start with DLs formed from content that needs to be digitized first, and then move on to content that is born digital.

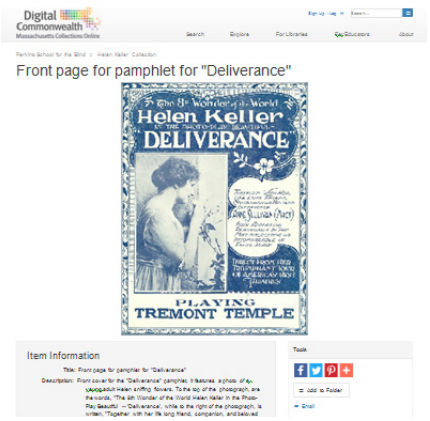
4.1 Digitized with Supporting Metadata

A natural extension to the electronic catalogue systems that libraries have long made use of is to combine it with a digitized form of the content it represents. Using principally the same mechanisms to locate items as before with minimal changes to the system's overall functionality, it can instantly serve up an online version of a located item, rather than merely report if the library holds that item or not. With a bit more effort, the system can be made available over the web, ET Voila` —a digital library!

It is a compelling notion that has led to the development of many digital libraries of this form, although two critical factors that impact its viability are copyright and digitization costs. For this reason, it is common to find libraries with special collections developing DLs along these lines, where the conditions are right. Special collections are typically modest in size, which helps keep the cost of digitization in check, and makes the task of determining any rights that might apply—and resolving where necessary—a manageable task. The key to forming this type of digital library is to base the search and browsing tasks around the metadata held in the catalogue. Figure 5 shows two snapshots taken from accessing the Helen Keller collection, produced by the Perkins School for the Blind to commemorate her life. In Figure 5a we see the result of a user searching the collection having entered the query term *Deliverance*, the name of the 1919 silent movie about her life. The main section of the returned web page shows a vertical series of thumbnail image. Horizontally adjacent to each of these, the digital library displays some textual metadata about the items: title, format, date.



(a)



(b)

Figure 5. The Digital Commonwealth's Helen Keller collection: **(a)** search for hetrogeneous items by meta- data; and **(b)** viewing a matching item.

This collection does not contain a digital version of the movie, but does include digitized items about Helen Keller's life, such as newspaper clippings, and personal photos. The first two items in the search results are photographs taken when Helen Keller travelled to California to participate in the filming of the movie; the third item is a pamphlet advertising the related theatre "photo play" Helen Keller undertook after the movie was made. It is this last item that our user has clicked on, shown in Figure 5b, to get more detail: displayed is a larger version of the pamphlet, and a detailed information box under it.

A key advantage to the approach of a digital library formed from digitized items supported by metadata is that it can be applied to a wide range of object types, including audio and video recording. In fact, the LibriVox audio book collection presented in Section 1 is another example of this form of digital library. A significant disadvantage to this form of collection, however, is that the metadata used is a surrogate for the items and consequently is only a summary of the item it describes. Even though a digitized version of the actual item resides in the digital library, if the terms our user chooses to search or browse by do not exactly match those used in the metadata fields, then the item will not be located. This could have happened in the Helen Keller example, for instance, if the pamphlet's metadata did not specifically record the movie name— despite it being visually prominent in the item itself. This deficiency is something the next form of digital library we look at seeks to address.

4.2 Digitized Textual Documents

The next form of DL collection we consider is the case of digitized text content such as books, journals, and newspapers. We shall collectively refer to this form of content as 'textual documents' to emphasize the point that, while these documents contain mainly text, they can also include figures, graphs, and tables . . . like this article, for instance!

When the source material being digitized is textual, an extension to the metadata approach, described in the previous section, is to apply Optical Character Recognition (OCR) to the scanned pages to acquire the full text, and allow the user to search this in addition to the metadata. This has the advantage of alleviating, to some extent, the metadata surrogate issue of the previous approach.

To describe this change as an extension, however, is really a misnomer because this conceptual addition— so simply described—introduces a plethora of factors that need addressing? Beyond the logistics of having multiple page textual documents scanned, beyond the issues of errors introduced through OCR, working in this way with documents of this form triggers numerous changes in the resulting digital library, giving rise to a distinctive form of digital library.

Unlike the previous approach, which effectively treats the digitized item as a black box purely described by its metadata, the consequence of OCR a document means we metaphorically open that box. Inside we find a myriad of logical structures that the digital library needs to contend with, including:

- pages, sections, and chapters for textbooks;
- multiple articles written by different authors in the case of journals, also broken up into pages and sections; and
- all of the above for newspapers with the addition of multi-columned items that can abruptly stop on one part of a page, only to be continued on a different page, further on in the newspaper.

Perhaps we should describe it as opening Pandora's Box!

When searching a digital library collection of textual documents, say for books on the cage oilseed press with the query *cage* and *press* it would be frustrating if all books that included the words 'cage' and 'press' *somewhere* in the text, were returned. One can imagine textbooks on many other subjects where the use of these two words would occur and what if the textbook used the plural form 'presses' or it chose to capitalize the words, and should they count as match as or not? A digital library that gives users the option of limiting query terms to appearing in the same chapter (section, or page) would help limit the scope of texts returned. For the subsequent complications mentioned, stemming words is a full-text indexing technique that would assist in finding plurals, and case-insensitive matching for capitalized words. Supporting these options in the digital library would further assist users in their information seeking tasks.

Implementing a digital library system that provides such an array of features is a significant undertaking. PapersPast, by the New Zealand National Library, is one such example. It contains over 3 million digitized newspaper pages spanning the 19th and 20th centuries. Figure 6 shows an example of a user searching for contemporary articles reporting on the New Zealand parliament passing legislation, in 1893, that gave women the vote—the first country in the world to do so. Figure 6a shows the result of a user searching for *suffrage* and *bill*, having restricted the date range to be 1890–1900, the approximate period our user recalls, from history class, the bill was passed. Clicking on the top matching item *Female Suffrage, Sir John Hall's Bill Carried*, leads to Figure 6b, where details of the speeches made in parliament prior to the vote are described. In this view of the newspaper, the digital library shows the matching article; from here the user can click on a change-view icon to see the article *in situ* on the page.

(a)

(b)

Figure 6. Searching the New Zealand National Libraries newspaper collection, PapersPast: **(a)** search results; and **(b)** viewing a matching article.

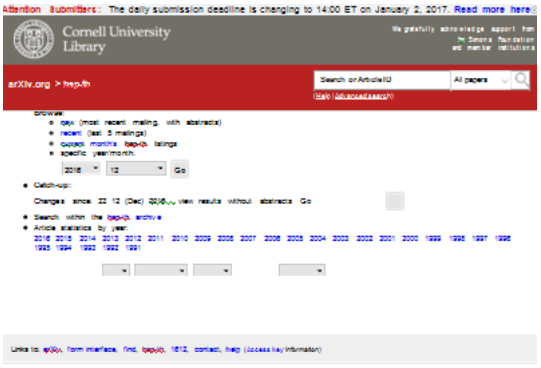
The Old Bailey digital library we encountered earlier is also an example of this form of DL. In that case we viewed the OCR'd text (Figure 3b), however the option to view the scanned original as well—which helps account for discrepancies between the original and the recognized text—is also there. This is accessed through a hyperlink located on right-hand side, although we did not include that snapshot in the given example.

4.3 Born Digital Content

Increasingly in our world, documents are born digital, meaning they are created from their inception on a computer, rather than having to undergo a digitization process is it scanning a textual document, or converting an analogue audio or video recording to digital form. A document is born digital when we write a report using a word processor, take a photo with our smartphone, or record the audio to an interview with our laptop.

When developing digital libraries from such content, this means metadata and document structure (in the case of textual documents) can be derived directly from parsing the file. In the case of text, rather than relying on error-prone heuristics that OCR processes use to differentiate section headings and article titles from the main text, this information can be determined unambiguously. In the case of photos, it is a simple matter through the phone's settings to allow the GPS location where the photo was taken, for instance, to be encoded into the photo file as metadata, and so on.

The previously encountered kiwifruit digital library of technical reports is an example of this (Figure 7).



(a)

(b)

Figure 7. The arXiv High Energy Physics Theory collection: **(a)** the about page; and **(b)** accessing a document in the collection.

Searching in the digital library can be performed at either the document or section level, the latter meaning that if multiple query terms are entered for the search, then they must all appear in the same section to be returned as a match. When viewing a document, top-right in the displayed webpage a table of contents is shown, which can be used for navigation purposes. This is out of sight in the figure as it was taken after the user had honed in on a particular part of the document: Section 3.3, Rating Canes for Frost Damage.

Figure 7 shows a further example of this born digital form of DL. It is taken from the arXiv (pronounced archive) web project (*www.arxiv.org*), not to be confused with the Internet Archive *(www.archive.org)*. Its origins pre-date the web, where scientists in the field of high energy physics started collating electronic preprint articles of their work using techniques such as a shared mailbox and anonymous FTP server. With the advent of the web, they formed a site for accessing the documents that we would recognize today as a digital library. The original focus of high energy physics has expanded to span astronomy, computer science, general physics, mathematics, quantitative biology, quantitative finance, and statistics, which are organized as separate collections. At the time of writing this paper, the digital library contained over 1.2 million articles.

Figure 7a shows the home page of the High Energy Physics – Theory collection, the original topic that motivated the digital library. There are various ways to access its content, with emphasis in the interface not surprisingly placed on viewing recently posted articles, given the intended audience. As a point of interest, let us use the digital library's features to determine what the first ever article included in it was. Using the collection's browse by date capability we determined 1991 was the furthest back the collection went. Clicking on that year produced a list of five items, the oldest being in August. And clicking on that brought up summary information about the document (Figure 7b), *Exact Black String Solutions in Three Dimensions*, authored by James Horne and Gary Horowitz. The article is 17 pages long, and through links on this page can be viewed in either Postscript or PDF format.

5. Building Blocks

The examples presented in this article help demonstrate the value of digital libraries. But in terms of the underlying software, how is all this achieved? Just how exactly does such content get into a digital library? What technically needs to be done so a collection can be searched at the page-level? Or browsed by different attributes such as date or subject? How is the presentation of information in a digital library controlled so that thumbnails of photographic content can be shown in one place, and in another place a PDF icon is used along with the title and author(s) of the document?

Figure 8 gives a general overview of the software architecture used to deliver these forms of DLs. The key components are: a database in which the metadata is stored; an indexer to perform the full-text searching; and a web server to allow users to access and interact with the digital library. Implicit in the diagram is the role of the server's file system, which is used to store the documents themselves, such as the PDF file of a report, or the MP3 file of an audio book reading.

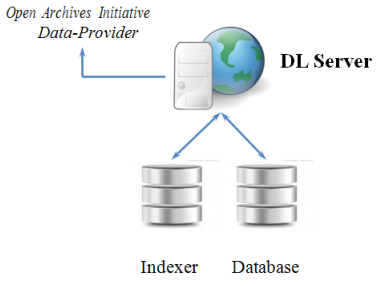


Figure 8. The key software components to a digital library.

There has been significant uptake in digital library systems of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH); so much so that it is considered these days, by many, a standard component. Complementing the human accessible interface provided through the web server, the OAI Data-Provider in Figure 8 provides a machine-readable access point allowing external software programs to interact with the digital library. Emphasis is placed on the metadata the DL contains along with timestamps that mark when it was added/changed, allowing external software to selectively harvest this information—hence the protocol's name.

Some variations to the general design exist. Advanced forms of databases now offer full text searching capabilities, effectively allowing this component to also perform the indexer function. Comparably, developments in indexing technology allows text and metadata to be stored inside the indexing tool, allowing, as an alternative, the database component in the DL design to be dispensed with. In the case of our first form of digital library discussed (Section 4.1, Digitized with supporting metadata), there is no requirement for full-text indexing and so that component can be omitted.

Documents and metadata are entered into the digital library through an ingest process, which in most situations is manually controlled. This results in updates to the indexer and the database. Batch processing of files is one approach; uploading items through a web form, typically one document at a time, is another. A comprehensive digital library solution provides both forms of ingest.

In terms of providing the web-browser user interface and controlling the presentation of content from the digital library, beyond the fact that the pages are ultimately rendered as HTML, there is wide variety of competing approaches. Notwithstanding the specifics of a given syntax, the commonality to these approaches is that they utilize a HTML templating mechanism in which server-side scripting elements that access the digital library's API are embedded. The templating allows for the design of pages for the DL such as the search, browse, help and so forth. Server-side scripting includes functionality such as *for-each* and *if-statements* allowing the programmatic control of the page produced: *if* the search returns no matches *then* display the message *No Results Found*; when there are matching results to display, *for-each* item display its title and author hyperlinked so clicking on it takes the user to the relevant document, retrieved from the digital library server's file system.

There are many different digital library systems to choose from. DSpace, ePrints, Fedora and Greenstone are examples of open source systems that have seen considerable uptake internationally.

6. Conclusion and Future Bearing

The examples given in this article help demonstrate the usefulness of digital libraries, but they also help show that they are somewhat conservative in the functional capabilities they provide. Even with the widely used digital library definition, there is nothing in it that says digital libraries are just about supporting alphabetized browsing and full-text searching for access; they can be browsed and searched in other ways too. Nor do digital libraries need to be solely organized by catalogued metadata; when documents are ingested, their content can be processed, and algorithmically derived metadata can be produced that enhances organizational structures. And where does it say that the digital library environment should stop when an item of interest has been located by the user? There is no reason why the digital library environment should not continue to support the user in their activities as they access and use the item. In the articles that follow, you will find details of these different, empowering ways to constitute a digital library—and more—all backed up with demonstrated use in practical situations.

In the articles that follow we put forward the argument that significant strands of digital library research are being overlooked, to the detriment of both the digital library, and wider information systems, we use. More should—and can—be done to tap into this potential. We will demonstrate this through worked examples which show practical ways in which to enrich mainstream DL systems and beyond—respectively redefining boundaries and perceptions.

Taking Greenstone as exemplar software architecture, the focus of the second article in this series will be on the core building blocks of the Greenstone software architecture, to show how easily they can be reconfigured to achieve interoperability with other DL systems. In the final article, we will explore the role of crowd sourcing in a digital library using music video content as a motivating example, and again working with Greenstone to show in a practical way that more advanced capabilities can be supported by existing general purpose software architecture.

7. Notes and Sources

The genealogy example given is based on a true example and relates to an ancestor of my wife. A little poetic license was taken in the narrative presented here in describing how the information was unearthed. The literary tourism scenario was inspired by project work with Annika Hinze, a department colleague [HB12, HB16].

The presented definition of a digital library dates back to 1998 and is widely used in DL literature. The definition is often incorrectly attributed as coming from *How To Build a Digital Library*, a book co-authored by Ian Witten, Dave Nichols and myself. There is a grain of truth in this, in that one of the authors (Witten) was behind the original formulation of the definition, but while *How To* uses the definition it is not the first place it appeared. The original definition appears in [WAS98]; and this in its turn was the result of a literature review at the time, seeking to find some shared/common definition of the term, digital library.

There are countless bespoke examples of digital libraries that embody the given definition. To highlight just a few: the Internet Archive [Kah97], which we met in the Jane Austen example, is one and contains a wide array of content, not just audio books; the HathiTrust Digital Library [Chr11] example of a DL formed from digitized textual documents—its size is truly vast containing over 4 billion pages . . . and growing.

Further to bespoke solutions, open source digital library toolkits such as DSpace [SBB+03], ePrints [TH00], Fedora [SWP03], and Greenstone [BDB+04] effectively provide "meta digital libraries" that embody this definition. These software programs do not deliver a specific digital library *per se*, rather they provide a software environment from which nascent digital libraries are born.

For the record it should be declared that Greenstone is produced by the digital library research group I direct, although for the main part this is incidental. The ideas expressed in the subsequent articles could just have equally been applied to other DL architectures, for instance DSpace or Fedora. With such detailed knowledge of the code, by choosing Greenstone it meant that the development of these research projects could advance at a faster rate than if DL architecture had been chosen.

8. References

1. Annika Hinze & David Bainbridge (2012). Listen to Tipple: Creating a mobile digital library with location-triggered audio books. In: Proceedings of the Second International Conference on Theory and Practice of Digital Libraries, TPDL'12, Berlin: Heidelberg. Springer-Verlag; p. 51-56,
2. Annika Hinze & David Bainbridge. (2016). Location-triggered mobile access to a digital library of audio books using Tipple. *Int. J. Digit. Libr., 17*(4), 339-65, November 2016. https://doi.org/10.1007/s00799-015-0165-z.
3. Bainbridge D., Katherine J Don, George R Buchanan, Ian H Witten, Steven Jones, Matt Jones, & Malcom I. Barr. (2004). Dynamic digital library construction and configuration. *Research and Advanced Technology for Digital Libraries, pages 1-13.* https://doi.org/10.1007/978-3-540-30230-8_1.

4. Brewster Kahle (1997). Preserving the Internet. *Scientific American, 276*(3), 82-83. https://doi.org/10.1038/scientificamerican0397-82.
5. Heather Christenson (2011). Hathitrust. *Library Resources & Technical Services, 55*(2), 93-102. https://doi.org/10.5860/lrts.55n2.93.
6. MacKenzie Smith, Mary R. Barton, Margret Branschofsky, Greg McClellan, Julie Harford Walker, Michael J. Bass, David Stuve & Robert Tansley. Dspace: An open source dynamic digital repository. *DLib Magazine*, *9*, 2003.
7. Robert Tansley and Stevan Harnad. Eprints.org software for creating institutional and individual open archives. D-Lib Magazine, *6*(10), 2000.
8. Thornton Staples, Ross Wayland, and Sandra Payette. The Fedora project-an open-source digital object repository management system. D-Lib Magazine, *9*(4), 2003. https://doi.org/10.1045/april2003-staples.
9. Witten I. H., Akscyn R. and Shipman F.M., editors. Proceedings of Digital Libraries '98, Pitts- burgh, PA, 1998. ACM Press.